UNIVERSITY OF CALIFORNIA SAN DIEGO

Universal Informatics:
Building Cyberinfrastructure, Interoperating the Geosciences

A Dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Sociology (Science Studies)

by

David Ribes

Committee in charge:

Professor Steven Epstein, Chair
Professor Geoffrey C. Bowker, Co-Chair
Professor Andrew Lakoff
Professor Chandra Mukerji
Professor Susan Leigh Star

2006

The Dissertation of David Ribes is approved, and it is acceptable in quality and form for publication on microfilm:

____

____

____

____

Co-Chair

____

Chair

University of California, San Diego

2006

DEDICATION

For Jimmy Jones

EPIGRAPH

The world is One just so far as its parts hang together by any definite connexion. It is many just so far as any definite connexion fails to obtain. And finally it is growing more and more unified by those systems of connexion at least which human energy keeps framing as time goes on.

– William James

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation could not have been researched and written without the assistance, support and advising of many people. I would like to thank my chair, Steven Epstein, and my co-chair, Geoffrey Bowker. In distinct and complementary styles each have contributed to this research in innumerable ways.

I would also like to thank my other committee members – Susan Leigh Star, Andrew Lakoff and Chandra Mukerji – for their valuable contributions and many interesting conversations over the years. Each have redirected my research trajectory to its benefit. Naomi Oreskes also provided support in the early phases of the data collection.

The science studies program at UCSD, including the staff, students and faculty, have provided thorough training and exposure to a thriving and diverse intellectual milieu. More recently my work with the Comparative Interoperability Project (CIP) has provided insight across individual projects; I would like to thank Karen Baker and Florence Millerand.

I would also like to thank the participants of GEON, in particular Chaitan Baru, Bertram Ludaescher , Dogan Seber, and Margaret Banton who have each facilitated a very generous access to the field.

Financial and material support for this research has come from several institutions. A portion of the early part of my data collection was supported by GEON and the SDSC. Later research was supported by an National Science Foundation (NSF) grant for the CIP (#04-33369) at the Scripps Institution for Oceanography, and an NSF dissertation improvement fellowship (#0525985).

Finally, my research owes a great debt to my family and friends.

CURRICULUM VITAE

___ ACADEMIC PREPARATION __________

Ph.D. Sociology (Science Studies) – 2006 – University of California, San Diego

Dissertation Title – Universal Informatics: Building Cyberinfrastructure, Interoperating the Earth Sciences

Co-Chairs: Steven Epstein and Geoffrey C. Bowker
Committee members: Susan Leigh Star, Andrew Lakoff, Naomi Oreskes

M.A. in Sociology – McGill University, Montreal, Canada (2000)
Thesis: "The Unreflexive Scientist? The Case of Visual Representation."
Advisors: Axel van den Berg and Alberto Cambrosio

B.A. in Sociology and Mass Communications – York University, Toronto, Canada (1998) – Graduated Magna cum Laude

___AREAS OF SPECIALIZATION________

Information Systems & Infrastructures – Cyberinfrastructure	Methodology – Actor-Network Theory, Ethnomethodology, Grounded Theory, Enactment Theory
Qualitative Field Work & Analysis – Archiving, Collaborative Research, Digital Tools to Support	History Of Computing and Supercomputing – Policy and Funding aspects
Organizational and Institutional Studies	Interoperability, Standards, Ontologies, Metadata, Workflow – Process and Production

___ SINGLE OR FIRST AUTHOR PUBLICATIONS __________

"Elements of Social Science Engagement in Information Infrastructure Design" *ACM: The Proceedings of the* 7th Annual *International Conference on Digital Government Research, San Diego, CA, May 21-24, 2006 (co-author Karen Baker).*

"GEON: Organizing For Interdisciplinary Collaboration," *Doing Science on the Net* ed. Olson, Olson, Zimmerman, MIT Press (forthcoming).

"Comparative Interoperability Project: Configurations of Community, Technology, Organization." *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (2005). (co-authors, Baker, Bowker, Millerand).

"Studying Digital Library Users in the Wild" Conference Report from Joint Conference on Digital Libraries (JCDL) in *D-Lib Magazine* Vol. 11 Num. 7/8, July/August, 2005. (co-author, Michael Khoo).

- see also position paper "The Positions of the Social Scientist: Social and Technical Acts of Intervention." *D-Lib Magazine* Vol. 11 Num. 7/8, July/August, 2005.

"The Matter of Users." Review of *How Users Matter: The Co-Construction of users and technologies*, Nelly Oudshoorn and Trevor Pinch (eds.). Metascience (2005).

___ UNDER REVIEW ______

"Mining the Arts for Digital Visualization," *Social Studies of Science* (SSS).

"A Learning Trajectory for Ontologies," *Journal of the Association of Information Systems* (JAIS) Special Edition on Ontologies

___ OTHER PUBLICATIONS _________

"ASIST05 Interoperability Strategies for Scientific Cyberinfrastructure: Research and Practice." *American Society for Information Systems and Technology ASIST* (2005). (Author list: Baker, Ribes, Millerand, Bowker).

"Continuing an Ethnographic Approach - Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study." *LTER Network Newsletter* (Spring 2005). http://interoperability.ucsd.edu/documents/05LTER_NetworkNews_Spring. (Author list: Baker, Bowker, Millerand, Ribes).

"Book-Length Scholarship on Lesbian, Gay, Bisexual, and Transgender Issues As Reflected in *Contemporary Sociology,* 1972-2000" Report prepared for the for the American Sociological Association's Committee on the Status of Gay, Lesbian, Bisexual, and Transgender Persons in Sociology, 12 December 2001. (with Steve Epstein).

___ AWARDS, FELLOWSHIPS AND RESEARCH GRANTS __________

NSF Doctoral Dissertation Improvement Grant, National Science Foundation (~$8000) 2005-6

"The Standardized Revolution of Science: Building Cyberinfrastructure for the Geosciences"

Values in the Design of Information System Design. Two Week Graduate Student Workshop. Center For Science Technology and Society, Santa Clara University. Organizers: Helen Nissenbaum and Geoffrey C. Bowker.

Internship IBM Research – Extreme Blue Program

Internship Grant, Science Studies, UCSD ($1000) 2002

Entrance Fellowship UCSD ($15,000) 2000

Samuel Lapitsky Fellowship, McGill ($3000) 2000

Entrance Scholarship McGill Majors ($1500) 1998

Faculty of Arts Sessonial Academic Achievement List York University (1995-8)

Dean's Discretionary Scholarship York University (1995-8)

___ SELECTED RESEARCH EXPERIENCE ______

Research Assistant, Scripps Institute for Oceanography (SIO) (2004-present)
Comparative Interoperability Project – PIs: Geof Bowker, Karen Baker
http://interoperability.ucsd.edu

This research involves the comparative historical and ethnographic study of three information infrastructures: GEON, the geosciences network; Ocean Informatics (OI) at Scripps and; LTER – Long Term Ecological Research. The project goal is to understand how technical interoperability solutions (ontologies, metadata, standards) are tied to social and organizational strategies and configurations. The interdisciplinary research team includes a historian, an information manager and a human-computer interaction expert. This is a three year NSF funded project under 'Human and Social Dynamics' (2004-7).

Research Assistant, Geoffrey C. Bowker (2003-4)
Continuing research on the GEON cyberinfrastructure. This work includes working closely with information scientists and geo-scientists in the GEON project, but also more generally with IT professionals at the San Diego Supercomputer Center. Current research is focused on the development of ontologies, workflows and visualization programs, and to a lesser extent the implementation of the GRID. This work is participant observation, and thus includes providing regular organizational feedback to GEON developers.

Ethnographer, San Diego Supercomputer Center, La Jolla, California (2003)
Conducted ethnographic and infrastructural research in collaboration with information scientists and geo-scientists in the GEON cyberinfrastructure project

(www.geongrid.net), with particular emphasis on building robust and effective cross-disciplinary information systems for the study in the geosciences.

Ethnographer, GEON – the geosciences network, various locations (2002-3)
Conducted ethnographic and infrastructural research in collaboration with information scientists and geo-scientists in the GEON cyberinfrastructure project (www.geongrid.net), with particular emphasis on building robust and effective cross-disciplinary information systems for the study in the geosciences.

Researcher, American Sociological Association (2001)
Funded by the ASA, a quantitative survey of the representation of gay and lesbian book reviews was researched and presented for use in ASA evaluation.

Research Assistant, Steven Epstein (2000-2)
Researching the emergence of 'race' as a salient category in clinical trials and subsequent statistical analysis

SELECTED SYNERGISTIC ACTIVITIES

IBM -- Extreme Blue Intern – Almaden Research Labs (Summer 2004)
-Jeanette Blomberg, Paul Maglio – Rick Perrett (Business Mentor)
- http://www-913.ibm.com/employment/us/extremeblue/

Worked in a team including technical programmers and a business manager; acted as an ethnographer and field researcher at an IBM client site investigating organizational performance. Research included interviews, shadowing, focus groups and document collection. In turn, data gathered during this research was used to inform the development of an IT tool for facilitating project management co-ordination within the interorganizational boundaries of IBM as well as across the divide between the IBM delivery team and client.

Informal Advisory Panel, "SDSC Innovation: Communication for Cyberinfrastructure Collaborations, C3" (2004)
Researcher and consultant for communication practices in the development and deployment of cyberinfrastructure. This project involves research on new publishing models and technological collaboration environment.

Intern, UCSD-LIBRARY Special Collections (2002)
Research and management of display for the 50^{th} anniversary of DNA; final product "When did we *know* the structure of DNA?" Using Francis Crick's personal documentation (located as UCSD Special Collections) a history of challenges to the structure of DNA was collected and placed on exhibition.

Recruiter, Trayectos, UC San Francisco (2002)

Under Professor Hector Carillo, UCSF, recruited and pre-interviewed subjects for research in the National Institute of Health research project Trayectos: Social Context and HIV Risk among Mexican Gay Immigrants.

Conference Reporter (2002)
Various conferences at the Banff Center for the Arts. Included ethnographic and interview work, summary write-ups, and publications.

___ SELECTED TEACHING EXPERIENCE __________

Teaching Assistant, Program in Urban Studies, UCSD
The City in History (twice)
The City in the World
Urban Studies Theory

Teaching Assistant, Department of Sociology, UCSD
Sociology of Medicine – Prof. Epstein

Teaching Assistant, Department of Sociology, McGill
Sociology of Technology – Prof. Smith
Sociological Theory – Prof. Nelson
Television and Society (twice) – Prof. Nelson

___ CONFERENCE PRESENTATIONS &c. __________

"Infrastructuring Movements: Infrastructure as an Actor's Category" *Society for the Social Studies of Science Annual Conference*, Vancouver, BC, Can., 2006.

"Elements of Social Science Engagement in Information Infrastructure Design" 7th Annual *International Conference on Digital Government Research, San Diego, CA, May 21-24, 2006 (co-presenter Karen Baker).*

"What is Social Science to Cyberinfrastructure? Intervening on the Inheritors of Another Social," *Society for the Social Studies of Science Annual Conference*, Pasadena, CA, October, 2005.

"Metadata Deployment." NIEeS Workshop *Activating Metadata: The role of metadata in effective spatial data exploitation*, Cambridge, UK, July 6-7, 2005 – Organizers: Lex Comber, Mark Gahegan, Francis Harvey, Peter Fisher, Richard Wadsworth.

"Comparative Interoperability Project: Configurations of Community, Technology, Organization." *Joint Conference on Digital Libraries*, Denver, CO, June 7-10, 2005.

"A Learning Trajectory for Ontologies," co-presented with Geof Bowker at 4th *Annual Knowledge and Organizations Conference*, Long Beach, CA, May 6-7, 2005.

"An Institutionalist Engine for DCP: A Case Study," presented at *DCP Workshop at CSCW* (Distributed Collaborative Practice at *Computer Supported Collaborative Work*), Chicago Il., Nov. 6-11, 2004.

"Interoperability and the Machinery of Difference," presented at the *Society for the Social Studies of Science Annual Conference*, Paris, France, August 25-6, 2004.

"Social Integration in GEON" presented to the *NSF Site-Review team for GEON (year 2)*, San Diego Supercomputer Center, August 12, 2004.

"Pervasive Issues in Building Cyberinfrastructure" presented to the Director's *Management Team (DMC) at the San Diego Supercomputer Center*, February 9, 2004.

"Making Automation: Digital Visualization Technologies in Science," presented at the *Society for the Social Studies of Science Annual Conference*, Atlanta, October 15-18, 2003.

"Social Informatics: GEON," presented at the *GEON All-Hand Meeting, San Diego Supercomputer Center*, San Diego, May 23-27, 2003.

"In\Formalization: Mining the Arts for Digital Visualization and Design," presented at the *Infrastructures of Digital Design Conference*, San Diego, January 31st-Feb 2, 2003.

"Mining the Arts for Scientific Visualization," presented at the *UCSD Science Studies Colloquium Series*, San Diego, November 25, 2002.

"The Standardized Interface: An Exploration of the Object", presented at the Canadian Communications Association, Edmonton, May 15-17, 2001.

___ CONFERENCE ORGANIZATION _________

Workshop Organizer, Studying Digital Library Users in the Wild, workshop at *Joint Conference for Digital Libraries* (JCDL), Denver, CO, July 2005. Published: *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (2005) – (with Mick Khoo).

Panel Organizer, Interoperability: Beyond Standards (Two Part): i- Analyses of Interoperability ii- Interventions, panel at *Society for the Social Studies of Science Annual Conference*, Pasadena, CA, October, 2005 (with Geof Bowker)

Panel Organizer Cyberinfrastructure and E-Science, panel at *Society for the Social Studies of Science Annual Conference*, Paris, France, August 25-6, 2004 (with Geof Bowker)

Conference Organizer, *The Infrastructures of Digital Design*, graduate conference held at the University of California, San Diego, January 31st - Feb 2, 2003. More information available at http://infrastructures.ucsd.edu.

LANGUAGE SKILLS

Fluent Spanish

Working Knowledge French

ABSTRACT OF THE DISSERTATION

Universal Informatics:
Building Cyberinfrastructure, Interoperating the Geosciences

by

David Ribes

Doctor of Philosophy in Sociology (Science Studies)

University of California, San Diego, 2006

Professor Steven Epstein, Chair

Professor Geoffrey C. Bowker, Co-Chair

The creation of cyberinfrastructure is an ambitious U.S. endeavour to build large-scale information infrastructure for the sciences. Dubbed 'revolutionary' by their advocates, cyberinfrastructure names the goal of building a unified information substrate to 'interoperate the sciences' and promote multidisciplinary research collaborations.

This dissertation is based on a three-year ethnography of one such *emergent infrastructure* project: GEON, the geosciences network. I identify, as a principal research object, the *logic of interoperability*: an emerging set of techniques and technologies

which seek to preserve the specificities of heterogeneous sciences while linking them. In principle *standardization* offers the benefit of making possible communication, data sharing and integrated computing systems; however, in practice such projects often fail or generate substantial opposition.

I argue that the logic of interoperability seeks to blunt the politics of standardization while retaining its enabling properties. Rather than erasing disciplinary difference interoperability calls for the sciences to be known and mapped in order to make possible an automated crossing. In this vision, the specificity of the sciences are preserved while domains are linked through relations of *mediation*.

Drawing from research in Science and Technology Studies and the methodologies of actor-network theory and ethnomethodology, I trace the enactment of the logic of interoperability in GEON at three *scales of action*: institutional, organizational, and technical. At each scale I sustain a focus on the material and organizing practices of members as they work to interoperate the earth sciences. At the institutional scale there is a growing impetus and increasingly sophisticated skill-set for the arrangement of multidisciplinary collaborations of domain and computer science. At the organizational scale new methods for constructing large-scale umbrella infrastructures are being invented. At the technical scale a set of technologies of interoperability are under development which seek to automate translations of the data, language, concepts, and knowledge of science itself. Together these point to a mounting confluence of efforts at interoperability seeking a ‘revolution’ of science at all scales of action and positing a new model of governance for science.

– Introduction –
–

'A Revolution for Science and Engineering'

Today the rhetoric of revolution is strong within the sciences. The effects of this revolutionary vision are percolating from the highest institutional arrangements of science's organization to the details of practice and scientific knowledge. These changes are centered around 'cyberinfrastructure': a cross-institutional and multidisciplinary endeavour to transform the information infrastructures of the sciences and engineering. Cyberinfrastructure came to prominence with the publication of the influential blue-ribbon National Science Foundation (NSF) advisory panel's report 'Revolutionizing Science and Engineering Through Cyberinfrastructure', commonly dubbed the Atkins Report.

Initially an evaluation of the NSF's Partnerships for Advanced Computational Infrastructure (PACI) program, the Atkins Report has come to mean much more. With the creation of new Office for Cyberinfrastructure directly under the supervision of the NSF director, this vision has already spearheaded large scale re-organization[1]. It has led to the allocation of substantial funds for the design of cyberinfrastructure, and it is leading to the production of new models for infrastructure development in large-scale collaborations of computer and domain scientists.

[1] As of June 1st, 2006, Daniel Atkins, chair of the Atkins Report, assumed a position as the head of this new Office of Cyberinfrastructure (OCI).

The Atkins Report is a visionary call for transforming the organization of science: the informational arrangements of communication and collaboration, of the archives of the sciences, and research activity itself. The goal is to provide a unified platform for the everyday work of scientific activity, now framed as data collection, processing, representation, collaboration and communication.

'Cyberinfrastructure' is today's name for the explicit investment across departments, disciplines, universities and institutions for the development of a single informational base in science and engineering. From the perspective of cyberinfrastructure the disciplines –geology, biology, ecology and economics alike – emerge as crucial islands in a networked medium of information exchange. Acting as substantive and topical harbours, the sciences are sites for the systematic investigation of the natural world, for the production of new knowledge, or in the case of engineering, for nature's engaged rearrangement.

But from cyberinfrastructure these disciplines are today discontinuous: 'siloed,' 'stovepiped,' 'islands,' 'balkanized'. Knowledge is shared sluggishly, distance hampers collaboration, labours are wasted recreating local solutions, and raw data moves not at all while time eventually ridicules the meticulous and expensive efforts of its collection. This is the *problematic of cyberinfrastructure*. Here the enactment of discontinuous disciplines is coupled with a goal for their reintegration, for the creation of new inter-disciplines, and for the provision of a common means for communication and coordination:

> We envision an environment in which raw data and recent results are easily shared, not just within a research group or institution but also between scientific disciplines and

> locations. There is an exciting opportunity to share insights, software, and knowledge, to reduce wasteful re-creation and repetition. Key applications and software that are used to analyze and simulate phenomena in one field can be utilized broadly. This will only take place if all share standards and underlying technical infrastructures. (Atkins 2003:256)

This problematization is matched with a vehicle for its resolution: the production of a shared information infrastructure. To the extent that the border of information technology can be made contiguous with all science and engineering it is also a line along which to stage a revolution in the organization and practice of the sciences. A key term, and a shorthand for the larger goals of cyberinfrastructure, is *interoperability*: the functioning of software across platforms, data across media or hardware across design. Interoperability is the unproblematic movement across the lines, borders and boundaries of the sciences. But what are the lines across the sciences? And what is the boundary between a science and its information infrastructure?

--

This dissertation is an empirical investigation of enacting the boundaries and their crossings in the cyberinfrastructure project: GEON. GEON is the geosciences network, 'cyberinfrastructure for the earth-sciences'. It is a five year project, funded by the NSF with the goal of providing an umbrella infrastructure for the broader earth-sciences:

> To integrate, analyze, and model 4D data poses fundamental IT research challenges due to the extreme heterogeneity of geoscience data formats, storage and computing systems and, most importantly, the ubiquity of hidden semantics and differing conventions, terminologies, and ontological frameworks across disciplines. (GEON Proposal:1)

GEON a distributed project including participants from over a dozen institutions in the U.S. It is conceived as multidisciplinary, with participants from geosciences as heterogeneous as paleobotany to geo-physics. Doubly multidisciplinary, half of its principal investigator (PI) team are information technologists and computer scientists housed at the San Diego Supercomputer Center (SDSC). Triply multidisciplinary, it has also had an ethnographer on board since its formal inception: the 2002 'kick-off meeting'. Far more than a narrowly conceived technical project, GEON has come to be seen as a common platform for a future geoscience, connecting the institutions, data and practice of its users; it is to be a repository for earth science's data and knowledge, and; it is a community building project.

Building GEON to stretch across geoscience data, communities and institutions is an exercise in extension, in acting across the scales. The work of enacting cyberinfrastructure has quickly overflowed the borders which formally circumscribe GEON. The various GEON participants have built an organization which pushes its emerging infrastructure *up* to the institutions of earth and computer science. GEON is funded as a computer science research project, as basic earth science research, *and* as a prototype umbrella with the goal building infrastructure. These goals are crafted to enrol both computer and geoscience researchers. The results are to be displayed as applications of computer science accomplishment, contributions to earth science and long term infrastructure.

The various GEON participants have built an organization which pushes its infrastructure *down* to the conduct of science, to data and knowledge of the earth sciences and to the pressing questions of contemporary research. Coupled with a goal to apply high-end computing research, GEON is intended to stretch information technology into the very fabric of knowledge of the earth sciences. Knowledge integration and management technologies known as 'ontologies' are planned which will bring together heterogeneous data, knowledge and experts. Creating ontologies requires inventing novel forms of collaboration between computer and geoscience to translate knowledge into a computable form, all while maintaining a recognizable appearance to those whose knowledge is represented.

--

Stretching to institutions, technologies and organization has been the everyday work of enacting GEON. It is the work this researcher has traced through ethnography, that is, through methods not that dissimilar from those of everyday action in GEON. This is not the modeling of a 'sociology from above' but rather a situated view from the everyday activities of GEON participants. The range of movements in this study echo those of the actor's themselves. It is the broad scales of these member's activities which speak to the ambitions and scope of the project.

GEON is an attempt to tie together particular technologies, institutionalized goals and scientific knowledges to emerging organizational forms. Its 'push' has come 'from above' with visionary calls and funding incentives in the NSF. Its 'pull' has come 'from below' with proposals emerging from practicing earth and computer scientists. GEON's long term goal is nothing short of a revolution in the everyday research of geoscientists, and the production of a unified informational base for geoscience. This study of GEON will follow these flows as boundaries of scale and discipline are enacted and then crossed. I will dovetail the practical and theoretical questions: 'What is the work of building an umbrella information infrastructure?' and 'How are information technologies to be made the universal base for the sciences and engineering?'

--

In science and technology studies (STS) we have named *infrastructure* as a phenomena and as an analytic category (Hughes 1983; Hughes 1989). We have done this partially to bring attention to unexplored facets of scientific activity – both by scientific actors and within STS. Infrastructure is backgrounded, it is invisible, and it is frequently taken for granted (Star and Ruhleder. 1994). The work of infrastructure and its maintenance is itself often that of undervalued or invisible workers (Shapin 1989). In this marginalized state its consequences become difficult to trace and politics easily buried in technical encodings (Hanseth, Monteiro et al. 1996; Monteiro and Hanseth 1997). The design of infrastructure itself can make its effects more or less visible (Bowker, Timmermans et al. 1997). Calls to study infrastructure in STS have engendered methods for making it visible (Edwards 2003): practical methods such as observing during moments of breakdown (Star 1999) or conceptual methods such as 'infrastructural inversion' (Bowker 1994).

With cyberinfrastructure we have an actor's infrastructural inversion. GEON has set itself the task of becoming increasingly visible to geo-scientists:

- GEON's participants include well established research scientists, computer and geo alike.
- CI has become institutionalized, now one of three 'offices' within the NSF.
- The Atkins Report called for one billion dollars of funding.

- There are well over a score of CI projects currently planned or in progress, funded by various state institutions.
- Failures to effectively develop CI have been tied to wasted funds and lost data in the sciences; to slow-downs in the production of knowledge or technology, and; to the decline of the nation in a competitive global economy.
- A sociologist was requested to participate: GEON is taken to be of sufficient import and complexity to merit social research. The NSF has supported this social research, along with comparative studies of GEON to other CI's.

The call to infrastructure has become a powerful actor's category; it has moved from silence to phenomena. This shift, from margin to center, from invisible to observed, requires a transformation in our own method of study. Susan Leigh Star reminds us that *infrastructure is relative*. In this study *relation* (Strathern 1995) will become the object: cyberinfrastructure is '*for* the sciences and engineering' and GEON is 'cyberinfrastructure *for* the earth sciences'. Cyberinfrastructure is a particular computerization movement (Kling and Iacono 1988; Kling 1991; Kling and Iacono 1995). It is an *infrastructure movement* built upon the boundary 'IT/domain' for the resolution of disciplinary difference 'domain/domain'. In its everyday work it is these boundaries and their crossings which must be enacted. For GEON this has meant an accountability to i- the computer sciences and contemporary research; ii- to the geosciences and contributions in the study of the earth, and; iii- to a long-term vision of cyberinfrastructure and its goals for a universal informatics.

A Universal Informatics at the Boundaries of Science

> [O]ne should attempt to abandon the usual practice of conceiving of power as a function of a so-called society, and, conversely, attempt to construct sociology from the chip's architectures
>
> --Joseph Kittler

Paraphrasing Michel Foucault, Joseph Kittler has called for an investigation not of society in the computer chip, but of the architecture of society in the chip. This reversal shifts our attention from an extant society manifested in its technologies to the enactment of that society along with its technologies. The Atkins Report is an outline for a new practice of science, an arrangement of information technologies and a role for the social sciences in the production of this informational revolution. In no formulation is cyberinfrastructure defined solely as a technology; the architecture of cyberinfrastructure is tied to a model for its enactment. In this architecture the production of a common informational base calls for bringing together resources from the computational, social and domain sciences. This vision of technical development simultaneously enacts the divided sciences and poses a common informational substrate to bind them together. This is what I call the *logic of interoperability*. In this chapter I will outline the architecture for cyberinfrastructure; in the rest of the dissertation I follow its enactment as work at the institutional, organizational and technical scales.

--

This exposition of the logic of interoperability relies heavily on the notion of boundary work. Sociologist of science Thomas Gieryn defines this as the “attribution of selected characteristics to the institution of science (i.e., to its practitioners, methods, stock of knowledge, values and work organization) for purposes of constructing a social boundary ”(Gieryn 1983:782). Gieryn contrasts his approach with efforts at *demarcation* in which the analyst attempts to define an ultimate criteria for science, for example, by reference to a unified scientific method. With a focus on boundary work attention is directed away from generating a definitive set of criteria for science and to the performative work of actors as they define or negotiate ‘what is and is not science’. Rather than outlining an idealized typology, the attribution of science becomes an empirical phenomenon, situated in action, and with observable consequences.

Boundary work in this text will function both more narrowly and more broadly than Gieryn’s formulation. First, for Gieryn the primary boundary of interest is between science and either ‘the public’ or more simply ‘non-science’[2]. He describes an extensive ‘cartography’ of enacted disciplinary differences and the power relations along these(Gieryn 1999). For Gieryn boundary work itself is primarily that of those building an ‘inside’ for science, attempting to secure resources, authority or legitimacy through exclusion. In this sense Gieryn is in dialogue with traditions that advocate demarcation, such as positivism, functionalist sociology, logical empiricism and falsificationism (Merton 1942; Popper 1965; Compte 1975[1853]).

[2] In the philosophy of science the *generalized problem of demarcation* also includes boundaries for pseudo-science and religion

In this text boundary work will be any actor's work, scientist or otherwise, in enacting boundary, whether it be science or technical infrastructure. We will see that technologists dedicate as much work as scientists to build an autonomous inside for science and a distinct role for information infrastructure. As with Gieryn each enactment of a boundary is particular, it is a matter of situated talk and action. However, I follow two sets of boundary work more specifically; I will call these domain/domain and domain/IT. While formulated and enacted in situ, laminated (Boden 1994) uses have emerged in GEON, and CI more generally.

Second, with Gieryn boundary work is primarily a rhetorical activity: institutions, knowledge and practices are defined as either science or not. In this text boundary work will refer to any kind of activity in shoring up a boundary. This includes rhetorical distinctions in discourse, but also institutional or organizational arrangements and technical design.

Third, rather than securing an authoritative impermeability for science, as with Gieryn's work, here assembling boundaries is coupled with the goal of establishing means for commerce across those boundaries: domain to infrastructure to domain. In the model applied in this study the enactment of boundaries is always coupled to their *mediation.* The work of establishing boundaries contains within it a model and impetus for coordinated communication, collaboration and automated integration. Interoperability travels on the enactment of boundaries and their crossings.

"Cyberinfrastructure"

Cyberinfrastructure (CI) collects a broad constellation of visionaries, institutions, information technologies and trajectories for the future. Put briefly, the goal of the Advanced Cyberinfrastructure Program (ACP) is to gather and facilitate the technical, scientific and human expertise necessary for researching and producing an infrastructure of information technology for the sciences and engineering. The term cyberinfrastructure gained prominence in early 2003 with the publication of the Atkins Report. However, at the time the ideas, and to an extent the term, had already established a foothold in information technology (IT) circles. Projects such as GEON were funded explicitly *as* cyberinfrastructure; it is a vision of technology which is liberally scattered throughout its proposal.

Cyberinfrastructure is seen as transcending the disciplinary ties of the sciences but also of particular institutions. Projects such as the Bioinformatics Research Network (BIRN) funded by the National Institutes of Health (NIH), or the Open Science Grid (OSG) co-funded by the NSF and the Department of Energy, have also been dubbed CI. While GEON is primarily an NSF project, with cost-sharing between computer and geoscience directorates, a substantial portion of its funding comes from the United States Geological Survey (USGS). What connects between these diverse scientific disciplines and institutions is understood to be cyberinfrastructure itself: a base of shared information technology enabling coordination across science's range of expertise, topic, methodology, organization, technique or technology.

In the Atkins Report the natural sciences are usually the exemplar; however, CI makes no formal distinctions between natural and social sciences. For example, in March of 2005 a workshop was held entitled 'SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences.' Chaired by Francine Berman, director of the San Diego Supercomputer Center (SDSC), and political scientists Henry Brady, the explicit goal of this workshop was to begin formulating high-end goals and means to build CI for social sciences[3]. CI also explicitly includes engineering and makes occasional references to the humanities. The entire spectrum of scholarly and academic pursuits is encompassed. Cyberinfrastructure seeks to operate on research and investigative endeavours. The distinctions natural/social/engineering/humanities are not particularly significant, or, rather, they are no more significant than the myriad of boundaries that come to characterize the already divided and heterogeneous 'sciences'.

The Atkins Report describes its larger vision of cyberinfrastructure as:

> the creation of thousands of overlapping field and project specific collaboratories or grid communities, *customized at the application layer but extensively sharing common cyberinfrastructure*. The cyberinfrastructure should include grids of computational centers, some with computing power second to none; comprehensive libraries of digital objects including programs and literature; multidisciplinary, well-curated federated collections of scientific data; thousands of online instruments and vast sensor arrays; convenient software toolkits for resource discovery, modeling, and interactive visualization; and the ability to collaborate with physically distributed teams of people using all of these capabilities (Atkins 2003:7, emphasis added)

Thus CI includes heterogeneous but connected -- 'interoperable' -- resources for computing, visualization, archiving, and for communication: machine/machine links such as networking, machine/human such as interface and visualization, and human/human

[3] The report from this workshop is available at www.sdsc.edu/sbe/ (accessed June 1, 2006).

such as distanced collaboration. In this vision scientists working across the globe along with their laptops and software are be able to work together with ease.

--

Beginning with 'infrastructure' and then turning to 'cyber' it is worth decomposing the definition of cyberinfrastructure more thoroughly. Its problematization and goals appear at all scales of activity in GEON. Star and Ruhleder's explicit definitions of infrastructure resonate with visions of CI, in particular three characteristics:

> **(1) Long Term:** "infrastructure has reach beyond a single event"
>
> **(2) Useful to a Broader Community:** "infrastructure has reach beyond […] one-site of practice"
>
> **(3) Transparent to that Community:** "in the sense that it does not
>
> have to be reinvented each time or assembled for each task, but invisibly supports those tasks" (Star and Ruhleder. 1994)

Within CI these goals are more commonly articulated of as 'extensibility,' 'customizability' and 'modularity,' respectively. In each of these aspects cyberinfrastructure comes to be seen as an investment in a broader community, beyond that community and for support that will last across time.

Computer applications can be designed to answer a specific research question or conduct operations on particular datasets. These are often called 'one-off' or 'stovepipe' solutions; useful to and used by a single scientist or research team for distinct tasks. These applications are not designed to easily link across platforms or facilitate growth. For example the code in which the application is written may lack annotations and the coherent organization which permit future programmers to easily modify the program[4]. Contrariwise programs can be written with considerations such as

i) extensibility: a design which deliberates on future growth;

ii) customizability: the facility with which a program can be adapted for differing tasks; and;

iii) modularity: enabling the interchange of portions of code or subsets of programs as with generic tools, customizable interfaces, or platform independence.

Each of these concerns of information infrastructure continue to be sites of research within computer and information science. Within these fields there are several competing lines of technical solutions to questions of extensibility, customizability or modularity. These are the 'basic research' aspects of cyberinfrastructure.

Just as the final infrastructures themselves will play a role in new domain science research, the development and deployment of infrastructure is a contribution to computer science:

> An ACP [advance cyberinfrastructure program] could revolutionize computer science and engineering research itself because, for example, of its inherent complexity and requirements for systemic integration, the opportunity for synergy between creating and applying new knowledge, and the need for a more integrated understanding of the technical and social dimensions (Atkins 2003:20)

[4] MacKenzie describes the historical rise of an aesthetic or 'elegance' in code writing and an ethos amongst programmers for annotating which comes out of the team-based writing of large complex programs (MacKenzie 2001) .

Basic research remains a primary goal of the NSF. For example GEON itself was funded under the NSF Information Technology Research (ITR) program. ITR specifically demands new, experimental and high-risk research as defined by computer and information scientists themselves. In its proposal GEON defines its goals as the production of infrastructure, but also as contributing new knowledge to both computer and geo- science.

Extensibility, customizability, and modularity are some of the key terms in cyberinfrastructure applications today. They make up the ethos of infrastructure. The goal is the creation of applications which will serve more than individual scientists for a single set of investigations. The technologies themselves should facilitate modification for a broader set of 'community needs,' through considered design features. They have the dual *responsibility* of supporting specific research activity and linking across these in a more general fashion.

The Problematic of Interoperability

We are now in a position to understand more thoroughly the scope outlined by the quote from the Atkins Report above, specifically the vision that it will be "customized at the application layer but extensively sharing common cyberinfrastructure." This is the central goal of cyberinfrastructure, and it is articulated very similarly in GEON. Within GEON, as within CI, applications enable a scientist to address particular scientific questions while also permitting data and operational results to transfer to other applications or researchers.

This is also how the central problem of CI and GEON are framed. Rather than the philosophical tension between the general and the particular, in cyberinfrastructure we have its *problematization*. In other words we have here the goal of simultaneously producing sufficiently general resources to serve all the sciences and engineering while still meeting the requirements of particular disciplines, sub-disciplines and individual scientific users. *This is the problematic of interoperability*. A goal to push through this tension drives much of the technical – but also organizational – development of cyberinfrastructure**.** *The sustained goal of this dissertation is the analysis of the practical methods by which members enact interoperability.*

--

The unity, or universality, of science has had many advocates in the past, but none have had quite the same formulation as that found in cyberinfrastructure. A still common positivist formulation of universality has been at the level of a shared method[5]. Such positivist views were as quickly taken up by scientists as by philosophers:[6] as eminent biologist and statistician Karl Pearson once wrote "The unity of all science consists alone in its method, not in its material."[7] We have all learned of 'the scientific method'. In many high-school classrooms it remains a staple in the rhetoric of scientific education. However, in CI such a view is rarely heard in any sustained manner.

---

[5] The variations on the unity of science thesis are seemingly endless, many emerging from the logical positivist tradition including reductionism, linguistic, logical variations. There are also many previous lineages such as the 17th century universal languages (Slaughter 1982), and more recently international classifications (Bowker and Star 1999). Lest the reader see the history of science on a single trajectory it is clear that each of these universalist movements has contained its own proponents of discontent (Foucault 1970). In STS today views of a unified science are far less common and notions of disunity or studies of epistemic heterogeneity are often formulated (Galison and Stump 1996; Cartwright 1999; Knorr-Cetina 1999).

[6] This is a formulation from empirical or logical positivism. For analyses of its formulation in Comptian positivism see (Kolakowski 1972).

[7] Quoted in Porter (1995:21).

The logic interoperability poses a view of science in the reverse. Rather than beginning from a unified science, efforts for interoperability begin by the enactment of disciplinary difference. While scientists participating in GEON have occasionally made reference to a common ‘scientific method,’ most statements of vision, and all practices of interoperability, take for granted a fundamental heterogeneity in the sciences. This heterogeneity extends from the methods and means of data collection, to the conventions of representation and classification, and to the models of knowledge itself.

Below is a diagram from the Atkins Report. This is the ‘the architecture of CI’ (fig.1). This diagram summarizes the scope and structure of CI, and for us, captures the full logic of interoperability: boundary work, problematization and mediation. Architecture diagrams are a common visual trope within computer science. They represent interdependent sub-systems of hardware or software operations. However CI is not understood in such a narrow technical fashion. In the diagram below human activity, organizations, institutions, software and hardware technologies have been blended within each layer. In both the traditional and Atkins’ use of an architecture diagram, the system builds up from a foundation: upper portions of the architecture are supported by portions directly below and the relationships are of interdependent but discrete subsystems:

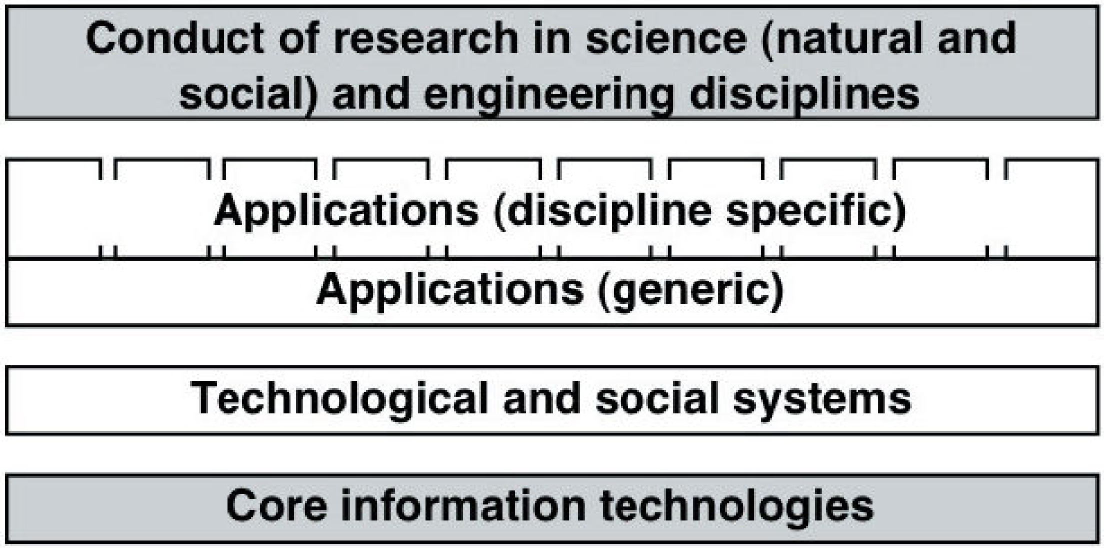

Figure 1: Architecture for CI, from Atkins (2003: 49).

At the top we have an abstracted 'conduct' of science and engineering. This is the practical everyday work and research, often generically called 'the domain' within IT circles. The legend for these diagrams notes that "the shaded boxes fall outside the scope of this report," (Atkins 2003:49) – that is, the details of scientific practice, or conduct, are not the concern of the Report.

The Atkins report is *about* cyberinfrastructure rather than science, per se. CI 'supports' the sciences, and is in turn supported by core information technologies (see below). The conduct of science is not unified, but particular. CI comes in contact with the variety of the sciences in 'tailored applications' (layer 2) – software which is designed to match the particularities of scientific conduct. A primary concern in the development of CI is this tailoring to the heterogeneities of science, as well as linking across these. In this sense the conduct of science is seminal to CI but because it is specific and heterogeneous, little more can be said about it in a general report. It is the task of specific projects, such as GEON, to tailor CI efforts to the heterogeneities of the sciences. Enacting the conduct of science as specific and heterogeneous is crucial to the logic of interoperability.

In this definition of the conduct of science we have the first set of boundaries in framing the problematic of interoperability: domain/domain. Conduct in the sciences is heterogeneous. This is articulated in the report as local practices and methods, specific vocabularies and concepts or journals read only in particular communities. These are what constitute domain/domain difference. But these differences at the level of conduct are mirrored by informational practices, and herein emerges the problematic of interoperability; data, standards and archives are incompatible, unmanaged or threatened:

> - Absent coordination, researchers in different fields and at different sites will adopt different formats and representations of key information, which will make it forever difficult or impossible to combine or reconcile.
> - Absent systematic archiving and curation of intermediate research results (as well as the polished and reduced publications) data gathered at great expense will be lost.
> - Effective use of cyberinfrastructure can break down artificial disciplinary boundaries, while incompatible tools and structures can isolate scientific communities for years
>
> (Atkins 2003:11)

Heterogeneity of databases, standards and information more generally are problematized as a concern at the application layer and include, for example, forms of data collection, curation and representation. Note that the conduct of science is not what is addressed as a problem. The heterogeneity of disciplinary ‘conduct,’ as scientific research, does not appear in this excerpt. Again, this falls outside the considerations of the Atkins Report.

Rather what is addressed is something different: the informational practices, or more specifically the relation of science to information. What is described here is the need for coordination at the level of data formats, schemas, database organization and for systematic archiving of data and results. In the excerpt above, each articulation of problem (loss of data) is tied with its solution (systematic archiving and curation). The problematic of interoperability is coupled to its resolution through mediation.

Interoperability as Mediation

The heterogeneities of method become the object of study and intervention in cyberinfrastructure, not with the goal of a methodological unification, per se, but with one of *mediation*. The preferred term varies – interoperability, integration, mediation – but in all cases the heterogeneous methods, classifications and epistemic criteria of *the divided sciences remain unchanged and become linked*. Again, disciplinary specificity is not problematized; in their research of the natural and social 'worlds' it is inherent in the domains to vary by methods, by concepts and more mundanely by data structure. It is these necessary heterogeneities that are taken to be the source of boundaries. And it is the boundaries, rather than the differentiated conduct, that are problematized. The goal is to bridge these by the construction of an underlying informational substrate. *This is interoperability as mediation*.

For example, the primary technical focus in this dissertation is GEON's efforts at "knowledge mediation" – software tools to assist in navigating across the disciplinary differences of the geosciences at the level of data, language and concepts. In the lingo of computer science these technologies are known as *ontologies*.

In its larger vision the Atkins Report casts domain difference at the level of the disciplines; in a somewhat more specific manner, GEON casts domain difference within the geosciences. The goal is to construct interoperated resources that are tailored for particular scientific domain applications while still operating on a layer of software and hardware common across particular instances of 'scientific conduct'.

Ontologies are understood to capture differences in language, concept and data structure and then translate across these. In a community of already collaborating scientists, with shared vocabularies, commensurate measurements and similar data formats, integration can be conducted as what in GEON is called a "simple one-world scenario": "several geoscientists doing similar experiments, observations, and/or simulations and, therefore, using overlapping and, ideally, standardized, data schemas," (GEON Proposal:8). Here integration, while requiring work, occurs at the levels which information theory takes to be tractable (Sheth 1999).

GEON's more ambitious goals require representing domain knowledge for integration of what are called "complex multiple-worlds scenarios": "where the mediated view has to span seemingly unrelated data, the source information is related only *indirectly*, and often in *highly complex ways* that are clear only to experts," (GEON Proposal: 9, emphasis in original). This is the realm of knowledge representation and ontologies. In a complex worlds knowledge representation the computing system has available the terminology, concepts and inter-relationships from multiple scientific domains. Coupled with a human expert the goal of knowledge representation is in facilitating a search for topical datasets or for data integration: the vision of ontology is to link theory, concepts, geographic or temporal information, a research institute or particular investigators, to name only a few. In this manner extant disciplinary differences (of language, concept or data format) do not need to be changed – the conduct of science remains heterogeneous – because such mediation technologies as ontologies can facilitate translations across these boundaries.

The boundaries to be crossed are by no means limited to the 'human' (of language or concept), but also of human/machine, and machine/machine. Cyberinfrastructure is an umbrella term covering a broad range of support for scientific activities, ranging from data collection to automating laboratory activities and for supporting distanced collaboration. Engineering centered projects such as the National Ecological Observatory Network (NEON) are intended to serve as platforms for heterogeneous and physically distributed instrumentation and remote sensor networks. Put briefly, this is large scale data production, and then its organization to ensure preservation and accessibility for environmental scientists. GEON is primarily a resource for data curation, access and manipulation. In other words, "GEON has no data", but instead seeks to take existing earth science datasets and computational tools, index them, provide methods of access through homogenous interfaces and services. In this sense GEON is data-centric. Interoperability is a concern which spans these services e.g. one must be able to compute distributed geoscience data and then send it to the visualization tool and the hope is to conduct all this through a single web portal.

Cyberinfrastructure seeks to support this range of practical activities: 'the conduct of science'. A common short-hand for this is 'end-to-end functionality', where an entire trajectory of scientific activity from data collection to knowledge dissemination is supported by an interlocked set of information technologies. And all of these are to be interoperable: data should flow smoothly from an instrumented fault-line in the earth, to the laptop of a scientist, to her tool for visualization tool and finally to its publication.

Information technologies will service *the sciences.* Scientific and disciplinary work is itself considered highly diverse, as are epistemic criteria. Science is not the isolated practitioner behind laboratory walls testing evidence against hypothesis, but includes a range of activities such as data collection, representation, curation and communication. Scientific work includes collaboration, locally and at a distance. It is messy, hands on, and a distinctly practical activity; it is technological, informational and communicational. This is the scientist 'in practice,' 'in a community,' and 'with instruments' rather than the scientist in thought.

--

In the history of computing there are predecessors to such ambitions of mediation. Both the enactment of a divided science and the solution of a unified mediation have come before. For example, Geoffrey Bowker has written of the 'imperialist' goals of cybernetics:

> Cybernetics, through its universal language, described what could in the broadest sense be called 'a new economy of the sciences'. By this I mean that it sought to order the sciences in a different manner from other universal disciplines by simultaneously offering new way in which they could cognitively interact with each other, and establishing new sources of funding to facilitate these interactions (Bowker 118-9)

Cybernetics would be a new 'universal discipline.' First it would be capable of translating all previous scientific knowledge into its own modalities. The most significant of these modalities has been the Claude Shannon's 1948 'mathematical theory of communication,' now more commonly referred to as 'information'. Second, cybernetics would provide a platform from which to view the activity of individual sciences and in turn could provide a vantage point for its meaningful (re)organization:

> Cybernetics could operate […] as a primary discipline, directing others on their search for truth […] At both the superstructural and infrastructural level, the rhetoric held that cybernetics was unavoidable if one wanted to do meaningful, efficient science. (Bowker 121)

But if cybernetics could be described as imperialist, then the strategy of cyberinfrastructure appears much less so. Cyberinfrastructure, as of yet, has not been conceived of as a discipline at all – or even as multidisciplinary. Rather, in the formal statements of cyberinfrastructure, but also the details of a technical imaginary and the everyday practice of enacting GEON there is enormous rhetorical and practical work dedicated to preserving the autonomy of science funding, scientific conduct and in particular scientific content. Building cyberinfrastructure will also be basic research for domain and computer science.

CI *begins and enacts* a vision of heterogeneous sciences, datasets, institutions and disciplinary arrangements with established trajectories. None of these are considered immobile, but they must be accounted for, known, and engaged in infrastructure development:

> Cyberinfrastructure should be produced and managed in a way that enables research communities/projects to tailor efficient and effective application-specific [...] knowledge environments for research and education.

> will also enable the federation of the necessary multidisciplinary, multi-institutional, and geographically dispersed human expertise, archival data, and computational models, (Atkins 2003:18)

With all this heterogeneity and acknowledged momentum of difference what is an "extensively common cyberinfrastructure"? The foundation for diverse science, local practice and heterogeneous data formats are the information technologies themselves.

The Logic of Interoperability and the General Technologies of the Particular

To recap, I have described the logic of interoperability as the simultaneous articulation of a problematic of interoperability and posing its resolution through mediation technology. The logic of interoperability requires the distinction of two sets of boundaries and then the establishment of relations across these. The CI architecture diagram above defines the boundary and relations 'IT/science' or more generally 'IT/domain'. Infrastructure supports 'the conduct of science,' or the systematic investigation of phenomena, primarily characterized as data collection, manipulation, representation and then findings. 'Conduct' itself remains relatively indefinite, although it is understood as science-specific and thus heterogeneous, it is roughly described as "a social activity, pursued by individuals, collaborations, and formal organizations" (Atkins 2003: 14).

In the CI architecture the conduct of science is supported by the application layer: "The conduct of science and engineering research is built (in part) on these applications, which are tailored to the specific needs of people, groups, organizations, and communities conducting that research" (Atkins 2003:48). Applications themselves can be as heterogeneous as the conduct of science. While the exemplars are quantitative data manipulation and representation in the natural sciences (such as mapping or visualization), examples also extend to other computational approaches such as knowledge management, social networks analysis or sophisticated search tools.

The heterogeneity of conduct in the sciences is endogenous and necessary to that activity; the investigation of the natural and social worlds requires methods tailored to their phenomena. These are the boundaries across domain science. However, these lead to difficulties in communication, coordination and in the exchange of data and knowledge. This is the problematic of interoperability. Finally, problematization of disciplinary difference is coupled with posing cyberinfrastructure as the solution: crossing balkanized domain/domain boundaries and transcending isolation. The means and methods of mediation are what I call the *general technologies of the particular*.

--

The image below is an excerpt from the architecture diagram (fig.2). It is the 'application layer'. On the top are the discipline specific applications – these serve the conduct of science. Applications are software or hardware (or combinations) that offer 'functionalities' to scientific users. For example, GEON has developed a software program that maps focal mechanisms as 'beach-balls' following the visual conventions of seismology. The tool visualizes a dataset of fault configuration as geographic information system (GIS) map. The application renders the intractable raw data of the earth sciences in a form accessible to the practicing scientist, supporting further research, or 'conduct'.

In the diagram, below the discipline specific applications are the generic. These are one step removed from the conduct of science and link across discipline specific applications. They mediate applications. For example, the ‘beach-ball’ visualization tool is in fact a compilation of existing tools in the earth sciences: geographic information systems (GIS) and the generic mapping tool (GMT). These are linked together by a software tool known as. The GEON workflow tool permits the geoscientist user to continue using both GIS and GMT tools, software which previous to the design of the workflow was not interoperable *workflows* (this example is followed more closely in Ch. 4.2). Workflows mediate multiple datasets and applications in configurations specified by the user. Similarly, in the example above of ‘multiple-worlds mediation’ ontologies were used to integrate across heterogeneous concepts, knowledge and data.

Workflows and ontologies are general technologies of the particular. They do not offer a direct ‘functionality’ to the conduct of science; rather they are one step removed from it. They mediate between particular knowledges, practices, software, data standards, and applications. In the words of the Atkins Report “Applications are a hybrid case with shared *responsibility* between technological and disciplinary programs,”(2003:52, emphasis added):

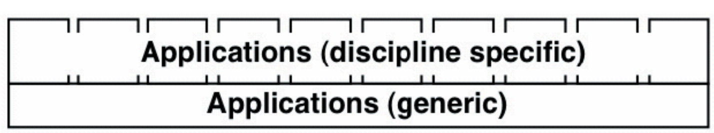

Figure 2: Selection from architecture for CI, from Atkins (2003: 49).

The ‘responsibility’ here is to resolve between specific applications and general infrastructure. They must link both to the conduct of science and to other applications. Again, it is not the domain scientist whose primary responsibility it is to mediate. Their responsibility is the conduct of science: basic research. Rather, the 'responsibility' is that of the information technologist tasked to design applications. This is a restatement of the ethos of infrastructure: as a responsibility applications should be designed with extensibility, customizability and modularity in mind.

I provide this simplified diagram (fig.3) which neatly represents the two boundaries and the relationships which together make up the logic of interoperability:

i- all domains are distinct (domain/domain) and require 'discipline specific' applications,

ii- all domains are in contact with information technology (IT/domain) which have the dual responsibility of being 'discipline specific' and 'generic.'

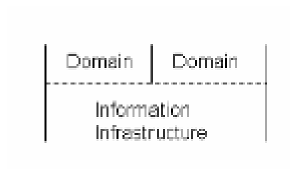

Figure 3: Relations of difference in the logic of interoperability.

As boundaries are established so is an order of relations between them, with infrastructure providing "coordination [...] and systematic archiving and curation."(Atkins 2003). The flow of data within resources rendered interoperable appears as follows: domain1=>infrastructure=>domain2. Domain1 and domain2 may be as 'distanced' as geology to biology, or as 'close' as geo-physics to seismology. What generates the connective string is the information infrastructure itself, which mediates differences in conduct, data, resources, language or knowledge.

Universal Informatics

To complete our tour of the architecture, below applications there remains 'middleware' and the 'core technologies'[8]. It is particularly indicative of the logic of interoperability. Within CI these are the informationally stabilized hardware and networking components: "The base technologies underlying cyberinfrastructure are the integrated electro-optical components of computation, storage, and communication that continue to advance in raw capacity at exponential rates," (Atkins 2003:6). While continuing to advance in terms of storage capacity, computational and access speed they are understood as having achieved a sufficient logical independence (see Appendix A) as to permit semi-autonomous development from software application writing.

Thus middleware and the core technologies mediate the application layer. Even generic applications such as ontologies and workflows must be linked amongst themselves. As we delve 'deeper' through the system architecture, away from the disciplines and towards the infrastructure, we approach a higher informational state, dividing content from form (Hayles 1999; Hayles 2002). While the sciences, and their conduct, is specific and thus divided, information technologies mediate this through generality and connectivity: interoperability. The diagram below (Fig.4) summarizes the boundaries and boundary relations which have been described above.

[8] Within this introduction I will treat these as one. In fact, 'middleware' is a significant object of research and development within CI circles, and at the SDSC. However, it is not generally a project within GEON. For the purposes of this argument, it is sufficient to understand that middleware links the application layer to the core technologies. The 'deeper' one traverses the architecture the closer one comes to a 'pure' informational state.

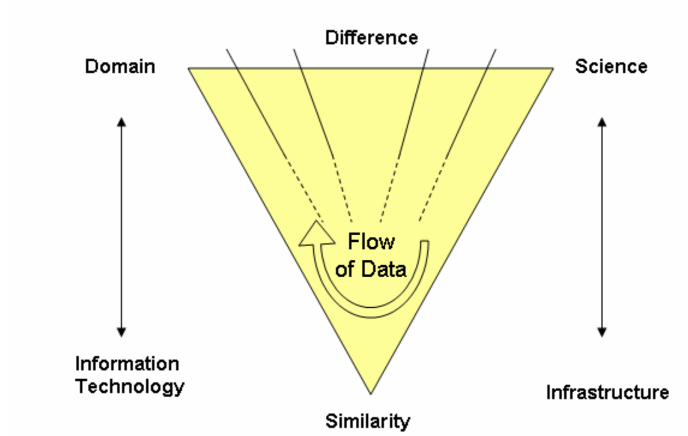

Figure 4: The logic of interoperability: at the top resides the specific conducts of the sciences, at the bottom information technology/infrastructure. Computing resources come in contact with all disciplines; information infrastructure is the point of mediation, 'interoperation', across differences at all scales of action: method, language, concepts, and institutions. The deeper into the infrastructure. the greater difference is rendered homogenous as information.

--

There are multiple methods involved in this *boundary work and mediation.* Thus far those identified have been primarily discursive: the writings in the Atkins Report and the GEON proposal. The importance of a shared vision cannot be overemphasized (Verran 2001) and we will return time and again to the programmatic statements of participants within GEON and CI more broadly. However discourse and vision alone do not cyberinfrastructure make.

The word interoperability itself flows freely within CI-circles, particularly amongst information technologists and the more technically oriented domain scientists. Close contact will be maintained with the common uses and understandings of the term; however, my goal is the respecification (Garfinkel 1991) of interoperability: to populate it with actors and work practices across the institutional, organizational and technical scales.

More important than the visionary statement ‘from above’ is the active uptake by everyday participants of CI: in this dissertation, the GEON participants. This dissertation is divided into three parts, what I describe as the scales of activity: institutional, organizational and technical.

At the scale of institutional activity, along with programmatic statements, are the funding arrangements and the institutional representatives. Here, as a practical achievement, the NSF must be made an institution for infrastructuring and the geosciences must come to see cyberinfrastructure as an avenue to further scientific research.

At the scale of organizational activity is the invention of a new form -- ‘emergent infrastructure’ -- for planning a long-term community resource. As a practical achievement, GEON participants must come to know their future user base and cross the boundaries of information technology and earth science.

Finally, at the scale of technical activity the boundaries must be bridged, a technical solution to disciplinary difference is enacted. Here, as a practical achievement, the knowledge of the earth sciences is rendered tractable in knowledge representations, and encoded into ontologies for automated travel across the interoperated geosciences.

Working across the scales is the coordination of all these diverse strategies and an alignment with the larger vision of cyberinfrastructure. In this section I have outlined the logic of interoperability. In the next I describe the rationale and sensibility in my approach to studying *member’s methods* for working across the scales. This is the practice of interoperability.

Across the Scales: The Practice of Interoperability

> One day the husband of the woman is called to the artist's studio. "What do you think?" asked Picasso, indicating the nearly finished picture. "Well…" said the husband, trying to be polite, "it isn't how she really looks." "Oh," said the artist, "and how does she really look?" The husband decided not to be intimidated. "Like this!" said he, producing a photograph from his wallet. Picasso studied the photograph. "Mmm…" he said, "small, isn't she?"
>
> -- Gareth Morgan

> Hobbes states that there is no difference between the actors which is *inherent in their nature*. We cannot distinguish between macro-actors (institutions, organizations, social classes, parties, states) and micro-actors (individuals, groups, families) on the basis of their dimensions, since they are all, we might say, the 'same size', or rather since size is what is primarily at stake in their struggles it is also, therefore, their most important result.
>
> -- Bruno Latour and Michel Callon

The scales are in their framing. They are made of material arrangements, discursive assertions and the routines of practice which link the two; the consequences overflow each. In the first epigraph above we have an encounter between Picasso and husband[9]. If we take Picasso as the authority, we can read the narrative as a modernist undermining of representationalism: the photograph has *no more* relationship to the wife than Picasso's (presumably) cubist abstraction and the husband is asserting a naïve realism. We could also assert that no one can be more authorized than a husband to speak on the appearance of his wife. Taking the side of the husband and his photograph we are

[9] I would like to thank Florence Millerand for pointing out further depths of nuance in this vignette; what initially appeared interesting by undermining representationalism revealed itself as fascinating by showing the enactment of a realism. The tale is attributed to Charles Hampden-Taylor.

then champions of the post-modern local against an abstract expert knowledge. However, what if we take the vignette as only an extract an ongoing conversation, then the excerpt becomes a local moment of negotiation in establishing criteria for evaluation: with familiarity of wife and photograph in hand the husband asserts a mechanical realism, and with painting and a wing of avant-garde supporters Picasso undermines a realism to replace it with his own. As Callon and Latour note, size is what is at stake in their struggles. If, as the pragmatists argue, only consequence determines outcome then there is nothing in the vignette which should lead us either way. Perhaps it is the silenced wife that will in the end appear and speak her size.

--

The very name GEON evokes an immediate image of enormity. All the institutions, technologies and even scientific objects closely associated with GEON also evokes scale: GEON is cyberinfrastructure for the geosciences; the aim of this National Science Foundation funded five year "large-ITR" is to build an information infrastructure for geo-scientists in order to provide a more holistic picture of the earth; to enable the next generation of science and education "GEON is developing advanced information technologies to support "intelligent" search and semantic data integration, visualization of multidisciplinary information spaces and 4D scientific datasets, and access to high performance computing platforms for data analysis and model execution," (eongrid.org, accessed September 05, 2005); GEON's member institutions and participants include multiple American universities, federal agencies, and commercial partners, as well as international collaborators -- it stretches from coast to coast in the US. The organizational and technical core of GEON is located at the "national leading edge" San Diego

Supercomputer Center (SDSC), and its own infrastructure rests upon the internet, the new academic internet2, computing nodes at Berkeley's Lawrence Livermore labs and it draws upon the experimental TeraGrid backbone. In a word*, and in words*, there is nothing small about GEON.

In a casual conversation with a colleague, a fellow social scientist who studies interactions across video-conferencing technologies, I was surprised to discover that he considered our projects and our methods wholly incommensurable. "I just study people across screens," he said to me. To this I responded that I was an ethnographer: "most of the time I study people without even any screens".

He clarified: "I only ever look at two people talking across a screen at the same time, maybe three or four if they are ambitious," with surprise I responded that I would have trouble following four people talking at the same time. I started wondering if *he* had some special techniques to study four people speaking simultaneously. But it seems he used the same Sony digital recorder as I did, and we shared the same transcriptionist. Our transcriptionist simply marks down "[kerfuffle]" when people speak simultaneously on her audio-files.

Attempting to bring rationality to the conversation my thoroughly microsociological friend countered "You study a whole infrastructure, for all of geoscience!" and I answered "But all of geoscience is never there!" Finally with baited finger he pointed at the glossy GEON fold-out flyer I had laid down, and declared "but they're going to revolutionize the geosciences!" [10]

---

[10] Langdon Winner has wholly disavowed the revolutionary rhetoric in IT circles as forms of ideology, or worse, as marketing hype (Winner 1984). We need not take the term 'revolution' at face value but we also

This surprise, this misunderstanding, is perhaps rooted in my view of activities in GEON. Participants have engaged in a series of tasks which in contrast to cyberinfrastructure, TeraGrids and revolution appear rather mundane. Seeing this is the privilege of the ethnographer.

The first GEON meeting, its 'Kickoff', was held in conference room not unlike those of even the least well off social sciences -- perhaps with a more solid wireless coverage, and well placed outlets. The fifty or so attendants spoke to each other face to face. In a format familiar to all academics and facilitating kerfuffle-free transcription, presentations were conducted at the front of the room, one at a time and accompanied by slides on a projector screen. The audience sat on only slightly plush fold-up metal chairs; admittedly coffee was plentiful. One absent principal investigator did telecommute, and showed up projected on screen, but technical difficulties delayed his appearance and eventually poor sound quality on the audio-video Polycom link forced the reappearance of a rather old-fashioned looking conference phone, placed discretely in a corner of the room within reach of the jack. Recounting *this* moment to my video-conference studying colleague, he responded with no surprise.

Those present in the room included an assortment of information technologists comfortable in their home setting of the SDSC, a broad assortment of visiting geo-scientists with retinues of their graduate students and technicians, along with representatives from federal earth science and educational institutions. Admittedly, laptops too were plentiful. Along for this Kickoff of GEON was also an invited social

---

need not dismiss it altogether. To the extent that it is an actor's category, in voice or text, this study will treat it *as seriously* as actor's themselves do.

scientist, recorder in hand, diligently writing notes on his laptop and maintaining his observations acute with gulps of coffee.

It is this privileged position, the proximity to the action, of the ethnographer 'social scientist' that explains the gap between a bewildered microsociologist's silhouette vision of GEON which stretches across national boundaries promising scientific revolution, and my own much more compact impression of the project.

It is not simply that I was present at the formal inception of GEON. It certainly helped to have been in a room and be able to look across thirty feet, rather than the country, and to have seen all of GEON in its days of simpler composition: geo-scientists, laptops, information technologists and coffee. What most helps is that three years later -- as GEON added ontologies, portlets, NASA, and 10,000 compute cycles from the TeraGrid to its membership -- when I looked across the room even then I still saw geo-scientists, information technologists, and an ontology projected on a screen and backlit laptop monitors. Admittedly by then the laptops were all new.

--

My goal in this dissertation is not to deflate GEON. There is undoubtedly something very large in this mobilization of experts, institutions, computing resources and data. However it is a methodological goal of this research to i- retain a particular notion of the observable as it has been developed in the ethnomethodologically influenced lineages of actor-network theory (Callon and Latour 1981; Strum and Latour 1987) and ii- to examine the production and maintenance of scale in action.

In GEON what has been mobilized to achieve interoperability? Above the actors included information technologists, geo-scientists, and laptops (and coffee). This answer is only partially flip. It is a partial cast of characters. Represented at the kick-off meeting were also all the other players we have since introduced. In a single conference room at the supercomputer center the collection of representatives included visits by NSF program directors from the earth sciences division and members of the United States Geological Survey (USGS). Ontologies too were present as information technologists launched on an educational campaign to familiarize GEON participants with the problematic of interoperability and the particular solution which they offer. It is the coordination of this entire assembly which is being mobilized to achieve the interoperability of the geosciences.

We can say that the involvement of the state, the NSF, the compliance of the geoscience institutions such as USGS and Earthscope are necessary to mobilize for interoperability. This is often what is understood as a 'social' explanation of a phenomenon. This explanation certainly does not suffice in this case: finances, policy and politics are never enough to hold science together, and they will not produce a map of the distribution of U/Pb Zircon ages of A-Type Plutons in Virginia. We could also say that there are necessary technological developments and specificities to the effects of particular technical solutions. This technical understanding of interoperability already has its champions in the journals of computer science and proceedings of the technologists. However, these writings speak not at all to the pressing questions of geoscientists such as: 'will I get tenure for making my database compliant with the geological metadata language?' To put it differently, institutional and technical explanations are necessary but not sufficient for understanding GEON's work to achieve interoperability.

Interoperability *occurs* at the site of organizational enactment. To explain mobilizing for interoperability we must move into the human-material arrangement which is GEON, and add to the cast of characters the collective work of the San Diego Supercomputer Center. These are the organizational aspects of interoperability, the practical points of alignment between technologies, domain communities, the institutions of science, and the scientific knowledge which is the content and goal of the entire endeavour:

> Information technologies are not simply purchased and plugged in, even when off-the-shelf products and series are procured for organizations. They are always subject to extensive design of their use within an organization and must be integrated with work processes, communication channels, means of coordination, culture, authority structures -- every central element of an organization.(Fountain 2001)

To speak of enactment (Weick 1969; Weick 1977) is to address the gap between (i) the *vision* of interoperability, the *declaration* of a goal, the *decision* to build GEON, NSF's *approval* of funding or the *choice* of ontologies as solution and (ii) a *functioning* system, *accessible* and *used* by a community of geo-scientists to *conduct* scientific work. Without committing to an outcome, it is this gap, the work of enactment which will be the empirical focus of this dissertation.

Describing this assembly spans scales which are usually the domain of differing subspecialties and methods within sociology (political, micro, organizational &c). The key will be to sustain an analysis which maintains a focus on the practical, discursive, textual and material organizing. The work of GEON then becomes *nothing more* than an ordered description of the lived daily negotiations of the GEON participants themselves.

It is one of the great strengths of STS to have developed a discourse for describing the ties between the smallest of entities and the largest of institutions: linking air-pumps with state-building(Shapin and Schaffer 1985), microbes with national plans(Latour 1988), and now the Wilson-cycle to cyberinfrastructure (Ch. 5). As Latour and Callon note, it not a tying together of small and the large, but rather the tying which makes them small and large. Air-pumps were tied to the early-modern English state in laboratories; microbes to national plans in the experimental farms of France; and the Wilson-cycle is tied to cyberinfrastructure in the conference rooms of the San Diego Supercomputer Center, in rooms no more impressive than a well equipped university classroom.

In the daily life of GEON participants engage in discourses, action, planning, and technical inscription which bring together concerns about cyberinfrastructure, the geosciences, technological detail and organization. In the very same breath and stroke of the keyboard that the future of GEON is planned so too is a future for geoscience, cyberinfrastructure and of the sciences writ large. Thus it did not require three distinct field sites, a stretch of the ethnographer's neck, to see activities in the range of scientific revolution, ontology writing, science policy, funding, and the understanding of pluton chemical composition because these were played out regularly in the same venues, at the same meetings, by the same cast of characters and often in the same breath.

Once again, it is in this sense that this dissertation is forthrightly *microsociological*; in the sense which has been emerging from the term at least since the intersection of ethnomethodology and actor-network theory. I do use this term not as a circumscription to instances of face-to-face interaction – such analysis will only be occasional. Rather it refers to the analysis of emerging structure from the interactional order(Goffman 1974[1986]; Knorr-Cetina 2002). The separation of this dissertation into three parts -- of institutions, organization, and technology -- is counter-posed with the continuous in-situ linkages that actors draw across these boundaries. These three 'scales', then, are heuristics and organizing principles of this text, used to facilitate a topical reading.

While this study is an ethnography, it is not an anthropology. I do not study the 'tribe of GEON' to provide a thick description of the everyday life world– neither norms or cultures, nor beliefs or values are the goal, though in other guises these may make appearances. You will not discover what an SDSC technologist wears, but you will hear of unfashionable web-design. At times a cacophony of disagreement will be heard, while at others GEON will speak in one voice and act in a marshalled stride. This study is best described as wide rather than deep, steep rather than pointed. Mirroring the action in GEON, we will come to see the entire range by following as they step from mountaintop to mountaintop.

– Chapter I –
Methodology

For a while I felt a low-grade thrill at being alive in the moment when this unprecedented thing congealed. But after weeks of jet-setting around the hypermap, I began to see the web as just the latest term in an ancient polynomial expansion. Each nick on the time line spit out some fitful precursor. *Everyone who ever lived had lived at a moment of equal astonishment.*

– Richard Powers

So I am talking about the trade of the sociologist or (since so many people do work that I think of, imperialistically, as sociology even through they themselves think they are some other breed of social scientists or humanist) about the trade of studying society, under the aegis of whatever professional title suits.

-Howard Becker[1]

–

This chapter is divided into three parts: methods, ethnography of scale and ethnomethods. I will take methods to be the researcher's practice of data collection, archiving and theory generation. In this study these three are roughly tied together under the approach which is known as 'grounded theory,' an empirically centered sensibility for developing theory which is deeply tied to the phenomena of investigation. Ethnography comes from a long tradition inherited from anthropology; it is an approach to the study of peoples in their own lived environments. However, in this study it is most closely informed by the canonical 'lab studies' in Science and Technology Studies, and with a concern for studying members work in generating and working across scale. Ethnomethods are the methods of the people. In this study these have been the primary

[1] To Becker's 'other breeds' I would like to add, even more imperialistically, geoscientists, information technologists and the various emerging professions of infrastructuring.

focus of empirical data collection: "by what methods have participants envisioned, planned and enacted an umbrella infrastructure for the earth sciences?"

In addition, because my data collection has been coupled with a participatory role in the GEON project, this chapter has scattered expositions of my role as a social scientist studying and participating in GEON. In grounded theory, and in some branches of ethnomethodology, the research process is tied to returns to the field and testing theory[2] by divulging findings to one's sources. Some of the concerns which inform this study are also of active interest, if with different intonations, to the participants of GEON: community building, knowledge capture, transformations in practice and organization, representing process and method. Throughout the data collection period it has been possible to discuss findings with informants both in casual and more formal settings. I hope that this text will also serve to further this activity.

1.1 – Grounded Theory as Methods

The approach known as 'grounded theory' combines theory, method and methodology in a particular approach to the conduct of qualitative research. My focus in this chapter is on the ways in which it has informed my methods of data collection and analysis. Grounded theory is part of a living tradition. A lineage is often traced from Pragmatism to symbolic interactionism (Shalin 1986) and then grounded theory (Glaser and Strauss 1973) and continues with further elaboration such as social worlds and arenas (Strauss 1993) and situational analysis (Clarke 2005). This section has been primarily

[2] A notion which is native to grounded theory rather than the 'hypothesis testing' of the ideal typical scientific method.

concerned with conveying how grounded theory has informed data collection, how it has assisted in organizing and analyzing that data and how it has informed iterations between field and data. In data collection it is the practical concepts such as sensitizing, sampling, and coding rather than overarching notions such as the conditional matrix (Strauss 1993: esp. Ch.11) upon which I have relied most systematically.

All methods come along with forms of theory. These are tied whether explicitly outlined or carefully hidden (Star 1989). The contribution of grounded theory to this dissertation is in no way limited to the role of 'methods'. However specific applications of concepts developed in the grounded theory tradition will be articulated within the empirical chapters, for example, to name only two, commitment in Ch. 4 and trajectory in Ch. 5.

Grounded theory is not a formal theory, per se, but rather a way of generating theory. The term 'grounded' refers to theorizing from and relative to a specific empirical enterprise. The result could be called 'substantive theory' by which it is meant that findings are tightly tied to understanding a particular object of research rather than attempting a generalization to the level of, for example, 'society,' 'organization' or 'institutions'.

I will refer broadly to my methods as ethnographic, although this should be understood in both the fullness and specificity which are described in the next section, and which in this section includes the stock grounded theory tools of sampling, iteration, testing and constant comparison, as well as an analytic focus beyond 'action' to discourse and inscription.

Grounded theory has usually been identified with a concern for 'action' or with a focus on the 'interactional order.' The advantages of this include a fine granularity that discourages skimming the complexities of situated action, emergence, or the in situ enactment of categories, boundaries, differences and similarities. This remains the interest of a substantial portion of this study. However, in studying i- a technological project, ii- a distributed project and iii- a visionary project with great ambitions and long-term plans it has been necessary to step outside the conventional realm of 'action' and inspect what Adele Clarke has called the discursive, material and visual[3]. The face-to-face interactions at All-Hands meetings, video-conferences, or ontology workshops have been occasions for vigorous activity, but much of GEON has *occurred* in email and list-serve exchanges, links on websites, in the circulation of publications or architecture diagrams and innumerable other forms of inscription.

In commencing a study there is no beginning with a blank-slate. Initial formulations of grounded theory encouraged entering the field 'without preconceptions' and without first familiarizing oneself with extant literature on the topic. More recent formulations have acknowledged the 'theoretically sensitized' investigator who arrives at the field site with a set of research histories, driving questions, and concepts tempered by participation in an ongoing scholarly discourse (Clarke 2005). Rather than standing outside academic discussion in this expression grounded theory engages the outstanding questions of the day in order to generate the pressing questions of tomorrow.

[3] "Our analytic focus needs to go beyond "the knowing subject" and be fully on the situation of inquiry broadly conceived, including the turn to discourse." (Clarke 2005: xxviii)

A virtual fly on the wall: in vivo and in silico data collection

A great benefit for the ethnographer in studying IT projects such as GEON is the continuous presence of technologies and of diverse and constantly changing participants often meeting in groups of a dozen, a score or more. While the primary boundary work in GEON lies on the axis IT/domain, most GEON meetings included a greater range of types of experts: educational specialists, NSF representatives, information mangers, technical support, and administrative assistants. For example, there was a commonly expressed sentiment that NSF program managers were keeping a close eye on the project, occasionally making unannounced appearances at meetings. Similarly these program mangers, along with multiple members of a cyberinfrastructure, supercomputing or geoscience organizations were on the GEON e-mailing lists. Meetings, email lists, and the general progress of GEON were the object of observation and evaluation throughout the project. In short we could say that a sense of being under scrutiny is endogenous to GEON, with the gaze of the sociologist carrying the significance of a drop in the bucket.

Amongst all these participants I, in my role as a 'social scientist' ethnographer, could come close to engaging comfortably in the ideal condition as a fly-on-the-wall, one heterogeneous practitioner amongst many. Almost without fail these meetings included an arrays of laptops, with various peripherals attached (mice, thumb-drives), LCD projectors, video-conferencing cameras, and the occasional tape recorder. Amongst these I could easily pull out my own laptop and digital audio recorder with no notice – only twice in my years of ethnography did anyone ever ask about the recorder, and here with the technical curiosity of geek rather than the ethical inquiry of the concerned.

This is in contrast to many accounts of ethnographic data collection in which researchers must dash to a rest-room, furtively scribble their notes and then spend nights revising, or; circumstances in which participants awkwardly glance at the tape recorder while speaking, or startle at any movement of the ethnographer to take notes. By having my laptop out at all times, hidden amongst a forest of backlit screens, I was able to continuously write notes, code these, and then revise during slower moments of meetings.

It was only during 'informal' moments -- breaks, lunch, dinner, on shuttle-buses and in hotel lobbies – in which I did not have my laptop and, out of respect for the more casual atmosphere, did not record. These informal moments were thus somewhat more tiring for me, requiring inside-the-head memory work and even the occasional dash to the restroom with notepaper.

I have never transcribed the recordings of entire meetings – the task was always too daunting, the hours too many, the voices overlapping, the "[kerfuffles]" too common. Four days of meetings could have been, quite literally, over a month of transcription that only I could conduct; in the best cases the sounds of thirty people across the room would have been too confusing even for voices familiar to my ears. Instead I marked, in vivo, particularly interesting passages for future transcription with time indexes and codes. Data collection and analysis have been facilitated by the qualitative research software suite NVivo.

This program has been designed with grounded-theory methods in mind, and permits a fluid coding of data, the re-arrangement of these codes, and the production of 'memos' which can be linked to particular data, to codes or to entire periods of data collection. In line with iterations of inductive and deductive work advocated by Glaser and Strauss, NVivo has facilitated easy movement to and from raw data (notes and transcripts), summaries, abstractions, codes, concepts and memos through point-and-click links.

'Memoing' serves to explicitly outline the importance of data, and links together notes, transcripts and the meaning of codes. Memos are "the theorizing write-up of ideas about codes and their relationships as they strike the analyst while coding," (Glaser 1978:83). Not only do they preserve the significance of particular coding schemes, they are also generated very close to the data: during collection or immediately following this in it's processing. Memos are informally written, impressionistic and sometimes fanciful; they are a space for free thinking and light commitment. As such they can serve as much to initially organize as to 'theorize' the data.

With the generation of literally hundreds of memos over the years, it become possible to review these in the light of further data collection, consultation with actors in the field or simply the sobriety of hindsight. Memos are summative and semantic encoding techniques which preserve the otherwise unwieldy data archives of qualitative research, rendering them accessible, manipulable and meaningful.

Codes and memos serve as sensitizing concepts which serve to focus the researcher on the collection of particular kinds of data, and also to guide the researcher towards new sorts of data. Glaser and Strauss call this *theoretical sampling*[4] and *theoretical sensitivity.*

--

[4] Glaser (Glaser 1978) distinguishes theoretical sampling from selective sampling. Selection is mining which pulls tendentious data to support an argument, while sampling is meant to guide further collection towards finer concept generation and theory testing.

Ethnographic observation began in November, 2002, at the GEON 'kick-off' meeting (see Ch. 4.1 and Appendix B). At this time I was also placed on the small set of email-lists that existed for the project at the time. This list expanded considerably over the next few years providing an enormous and constant 'data stream.'[5] Eventually the GEON website (http://geongrid.org) also became a significant source of information, including slide-sets, webcasts of meetings, or archives of list-serve discussions[6].

As a distributed project GEON has unfolded in its various sites across the U.S. GEON has also 'occurred' outside the US. I have presented my GEON alongside other GEON PIs in England, for example; some of the GEON participants are Canadian, and some regularly partake in Chinese geoscience, and; GEON has made gestures towards iGEON, a not-yet-existing international collaboration. I will limit my primary focus to efforts in the US.

During the first three years I attended as GEON's participants came together for an annual 'all-hands meeting,' and a sub-group of participants came to the bi-annual 'PI meetings.' Within six months of the Kick-Off meeting a monthly, or semi-monthly, video conference meeting ('ICC Meeting') was setup amongst the four or five of GEON's central PIs, I was kindly invited to these as well. I also attended two meetings of the Geological Society of America where GEON participated through a booth, poster presentations, and on one occasion a keynote symposium on "Geoinformatics and the Role of Cyberinfrastructure in Geosciences Research". The list goes on: Supercomputing

---

[5] I will use e-mail data only very rarely as many of these forums were considered private and highly informal. Email data is used with explicit consent even when maintaining anonymity. All this said, access to GEON's has been invaluable over the years, providing intimate details, and has included preprints of articles, reviews of public statements and draft reports to the NSF.
[6] Most of these documents remain publicly available on the GEON public website.

2003 ('03), visualization workshop, ontology training workshops, Geoinformatics (2003) &c. Finally I have taken a graduate course on ontology and knowledge representation, taught by a GEON participant. In addition to countless informal discussions on various occasions, I have also conducted a set of formal interviews with GEON's NSF representatives in the geo and computer science directorates.

Below I summarize this varied ethnographic observation in three 'empirical concentrations':

> i- *The Public Face of GEON*– Kickoff, All-hands, and PI-meets have been the occasions of much activity. Averaging three meetings a year, and which include most GEON participants.
>
> ii- *The Daily Face of* GEON – *A weekly* 'workgroup meeting' at the SDSC primarily composed of the IT team and (unofficially) top administrators. Geoscience PIs made an occasional appearance in person or through phone conference. In later years parts of these meetings were archived on the GEON website. These meetings dealt with the everyday logistical issues of building GEON, co-ordinating and communicating with participants, or planning future events. They offered an excellent vantage point from which to observe the general organizational emergence and functioning of GEON.
>
> iii- *Ontology Workshops*: These workshops have been foci for the production of scientific workflows and ontologies; they are one of the points of fine grained technical interaction between computer and geo- sciences.

--

Grounded theory generation is coupled with multiple obligations for iteration which include constant comparison, theory testing, returning to the field, and consulting with one's sources on findings. Iteration refers to returns to the field following grounded theory generation (coding, memos); in this study it has taken many forms. The most significant iteration was simply the coupling of rapid theory generation and long-term ethnography. My particular role as a 'social informatics researcher' in GEON occasioned several points of feedback to the GEON community, to the SDSC and to a broader community of informatics researchers. More common versions of returns to the field have occurred as casual conversations with GEON participants during meeting breaks, in random meetings on UCSD campus or in hotel lobbies while waiting for a shuttle. Many of these 'consultations with the field' were prompted by participants themselves, aware of the nature of my research they would regularly open a discussion with 'so what have you found out about us?' Due to their informal nature I have never recorded these encounters, but they have served to prompt further thought or reconsideration of particular findings.

A core precept of the grounded theory method is of 'constant comparison.' Those unfamiliar with this method should understand the theoretical underpinnings of these comparisons. This is not the formalist method of John Stuart Mill, in which the goal is to unearth particular causal variables through comparing discrete historical instances of difference and similarity (Mill 1843/1874). In the method of grounded theory, comparison is a vehicle for the production of richer and more elaborated meanings, to build empirically grounded analytic tools for understanding the phenomena. To be methodologically precise the technique is best defined as cross-case analysis, but comparison here will be used to show precisely that the various cases mentioned are *not* discrete, rather they should be understood as a the larger portions of an assemblage: the construction of GEON is at play in a larger field of cyberinfrastructure development, but also within advances in computer science, and movements for data interoperability that precede GEON or CI by decades.

In this study constant comparison has itself been a form of iteration and return to the field. *Comparison is endogenous* to the action of GEON and CI. For example, from the Kick-Off meeting the CI project BIRN, the Bio-Informatics Research Network, has been framed as GEON's 'sister project'. This has prompted many kinds of comparisons between the two projects, in terms of the specificities of domain knowledge, contrasts on the scope of each project, the technologies to be deployed, the character of the participants or the nature of the funding institutions (NSF vs. National Institutes of Health). Many of the GEON participants, in particular the IT PIs, participate in other CI projects, such as SEEK or CUAHSI and so comparisons emerge here as well. In this study I have drawn on these endogenous comparisons (and added some of my own, see part III): these sorts of auto-analytics have been crucial in defining GEON's identity and characterizing the nature of work for its participants.

Finally comparison is endogenous o the field of CI, which has set itself the goal of collating the collective learning of various CI projects. This is the rationale which informed the day-long 'Building Communities in Cyberinfrastructure' seminar at Supercomputing '04 which included representatives of many CI projects, including BIRN and GEON, but also NEES and LEAD to name a few: "The expected outcome is a set of best practices that can be applied by the many emerging IT initiatives," (NSF Officer). As with grounded theory's constant comparison, endogenous comparisons are by no means formal analysis, but should be taken to be, as Clifford Geertz notes, "stories that people tell themselves about themselves". Epistemic judgment should be suspended in favour of following the consequences of endogenous comparison. That members are themselves the primary actors in their own comparisons, scaling and analysis is the methodological terrain of actor-network theory and ethnomethodology. These are discussed in the next two sections.

1.2 An Ethnography of Scale

There is an enormous range of activities which come under the rubric of 'ethnography' today. Ethnography has been the staple method of the anthropologist at least since the 1910s with Malinowski's work in the Trobriand Islands. He called for a focus on a "member's point of view," through long term immersive studies that include both observation and participation. Ethnographic data has been used within sociology at least since Durkheim (Durkheim 1976), and became a method of the Chicago School (later symbolic interactionism) during the 1920's. I will not attempt a systematic review. However, I do wish to distinguish this study from the more general usage of ethnography within social informatics research circles and locate it a closer to the tradition established in STS of the 'lab study.'

Within computer supported collaborative work (CSCW), or participatory design (PD), and most often in application within human-computer interface design (HCI) ethnography is taken to be a methodological tool in the *requirements testing* process (Anderson 1994; Jirotka and Goguen 1994). Often, the goal here is to contribute something to the design of a particular software or hardware feature through a detailed, in situ knowledge of existing user practices, needs or contexts (Button 2000). Many of my interventions in GEON have been framed in this manner, as have some of my publications (Ribes forthcoming).

This dissertation draws from studies in CSCW and PD, but its structure is closer to the substantive theory generation of the 'lab studies' of science and technology studies (STS). The goal in this dissertation is grounded theory development, the understanding process and consequence and then also perhaps to contribute to broader dialogue on design and policy, on a vision of interoperability or plans to infrastructure the sciences.

--

There are many similarities of this study to the 'lab studies' which emerged in science studies during the early 1980's. The four canonical studies, all conducted in California during the 1970's but with distinct impetus and style, are Latour and Woolgar's *Lab Life*, Lynch's *Art and Artefact*, Knorr-Cetina's *The Manufacture of Knowledge* and Traweek's *Beams and Lifetimes*[7]. Each of these studies was based on extended periods of ethnography conducted in laboratory settings: biological sciences for the first three, high-energy physics engineering for the latter.

It is difficult to pin-point the exact style of these ethnographies, but the range of explicit influences included ethnomethodology, symbolic interactionism, semiology, and, of course, threads of cultural anthropology. These studies took a particular lab site (or two in the case of Traweek) and investigated the 'shop talk,' practices, and the written communications within these laboratories[8]. Topical emphasis varied widely, however

[7] (Latour and Woolgar 1979), (Lynch 1985), (Knorr-Cetina 1981), (Traweek 1992).

[8] Each of the canonical lab studies employed a strong language of empirical investigation e.g. "direct observation of actual practices". These phrases were repeatedly emphasized throughout the texts. Although the purposes of the lab studies varied, much argumentative emphasis was placed on countering, for example, naïve understandings of scientific method, of direct access to phenomena or, of rule-following

general themes emerge across the studies such as counter-posing the empirical observation of scientific practice to philosophical idealisms (of philosophers or scientists alike); strong support for social constructivist arguments about the nature of facts, truth production or scientific method, and; of the local, material and contingent nature of laboratory practice.

For example, Latour and Woolgar followed, over many years, the emergence and stabilization of the entity TRF, a hormonal factor. More accurately they conduct an ethnographically enabled semiotic analysis of the *statement* "TRF is Pyro-Glu-His-Pro-NH_2" by observing how through instrumentation, texts and interpretive practice all 'modalities' (or historical, practical, and contextual referents) were removed from the statement over time, thus resulting in what Woolgar and Latour (re)define as a 'fact.'

Latour and Woolgar's approach placed particular emphasis on the circulation of 'literary inscriptions' within the lab. They followed the flows of inscriptions (or texts, visualizations, charts and diagrams) around and out of the lab and argued that the primary focus of scientific work is on the production and inspection of concatenations of inscriptions rather than directly on biological substances. Instead, and in what has become a typical pragmatist reversal within actor-network theory, 'substance' came to be understood as the outcome of all these practices, inscriptions and linkages.

---

behaviour. With such goals in mind posing ethnography as providing 'direct observation' of what was 'actually' going on in labs was an opportune rhetorical move. However, this argumentative style has not stood up to the continuously reflexive elements of the field. A good example of the hyper-reflexive turn appears in Woolgar's own look back on his lab studies (Woolgar 1988). See also below, Lynch's comments on attempts at a 'metalanguage'. The reflexive discussions amongst anthropologists that have been percolating through that field since at least the 1970s became clear within science studies: notions of 'direct ethnographic observation' can themselves be placed under the same sort of sceptical scrutiny which first motivated these lab studies. The 'reflexive' literature on the subject is seemingly endless within anthropology, however to cite only one, Clifford has offered some historical treatments 'on ethnographic authority' which are congruent with analyses of scientific methods and instruments within STS (Clifford 1983).

Similarly this ethnography traces, over time, the stabilization of various entities. These include a 'push and pull' to cyberinfrastructure, ontologies and the method of their production, the future users of GEON, and GEON as an institution itself – none of which achieve anything like the facticity of TRF in the span of this study.

--

Immediately upon the success of the canonical lab studies came the impetus to study 'broader contexts,' to move outside the laboratory walls, or include the broader scientific communities and institutional influences. In fact many of the pioneers of the canonical lab studies have since spearheaded research trajectories to understand the enactment of a 'macrosphere':

> No matter how great the interests of many social groups for what is being done in one laboratory, there is nothing to stop interests from fading and dispersing if nothing more than laboratory studies happens. (Latour 1983:150)[9]

These trajectories of research are particularly important for this study, as they were conducted by 'micro' scholars with interests in explaining the 'macro.' The approaches which have most influenced this study appear under the rubrics of ethnomethodology and actor-network theory (ANT).

Ethnomethodology takes as its 'object' the methods of the people. Its studies can focus on the most mundane of activities, such as crossing a street (Garfinkel 1967), or the most technical, such as doing polymerase chain reaction (PCR) with standardized kits (Jordan and Lynch 1998). Ethnomethodologically influenced strains of actor-network

---

[9] This is, in fact, a clever symmetrical double-entendre; is Latour speaking of Pasteur or Science Studies? The claim applies equally, if either remains at the level of 'micro' lab study, achievements will be of a passing interest.

theory[10] take as their ‘object’ the methods by which something comes to be, or not. For example, the focus might be how we know whether a forest is receding (Latour 1999), how the electric car comes to fail (Callon 1986) or how heterogeneous actors are coordinated in airplane design(Law and Callon 1988). From the perspective of ethnomethodology ANT provides a ‘gloss’, skipping over vast swaths in the practical details of doing. However, ANT studies tend to cover more ground, can tie together vastly heterogeneous methods and are able to focus on longer term consequences than ethnomethodological studies[11]. This dissertation shifts between the granularities of analysis which typify ethnomethodology and ANT; the traversing of broad activity terrains is coupled with a detailed analysis of talk and practice.

Both ethnomethodology and ANT are ‘meta-methods’, or more simply methodologies which guide in understanding the methods used by actors. ‘Practice’ in these approaches ceases simply to be ‘what people are doing’ and become part of a rationale, a tying together action, materiality, knowing, and outcome. Practice is an instance of a specific method. Very little more can be said about ‘practice’ in general, as

[10] The term ethnomethodology in this study is used with not a little trepidation. From the vantage point of the turn of the century the short history of ethnomethodology seems fraught with internal (non-indexical) finger pointing and endless boundary work through disavowals or associations: “I know of no discipline, which has suffered more at the hands of its expositors than ethnomethodology”(Livingstone 1988). This is unfortunate as it seems to me there are bigger bugaboos than creative theoretical appropriation. I could also, perhaps, have borrowed the term post-analytic ethnomethodology (Lynch 1993) to describe my approach, or perhaps could have claimed to adopt an ‘ethos of the ethnos’. However, in the end, I have chosen the awkward phrase ‘ethnomethodologically inspired branches of ANT’. Despite its own battles for internal stake, ANTmethodologists have generally demonstrated themselves to maintain a catholic tolerance and greater concern with theoretical agility through substantive research. Most importantly this research is clearly not an ethnomethodological study of work. There is still much value to be gained from granular analysis of local accomplishment (‘haecceity’), which here I will substantially gloss. Thus, I do not argue that ethnomethodology needs to ‘move on’ according to the stipulations of this research, but more simply that this study is inspired by a creative application of ethnomethodology’s canonical texts.

[11] See also the hybrid notion of ‘lamination’(Boden 1994; Taylor and Van Every 2000) which attempts to bridge across interactional encounters without losing the sense of local accomplishment. This concept is developed and applied in Ch. 4.

its doings are always specific. From both approaches the methods of the people remain in principle observable whether vernacular or technical. This is the keystone in an ethnography of scale.

--

One of the main goals of this dissertation is to stretch an analysis across the scales, what I will very loosely call the technical, organizational and institutional activity of GEON. Too often are studies of organization, institutions and policy divided from those studying the interactions, the technical work and routines of scientific practice.

Bruno Latour has been particularly adept at articulating the problematic of scale as we have framed it in social science: "We [social scientists] use a model of analysis that respects the very boundary between micro- and macroscale, between inside and outside, that sciences are designed to not respect" (Latour 1983:153). From such initial reifications of scale – from the acceptance of an ontological distinction between micro and macro – Latour argues that what follows is a bifurcation of the scholars and methods for addressing these now distinct spheres: qualitative research will cover micro spheres of meanings, interactions and negotiations while quantitative methods will bring to bear statistical strength in covering greater territory. With this setup of divided research traditions the traditions of social science have themselves devised a ground upon which to justify the production of 'key problems' in social science along with volumes of their solutions: context/content, external/internal, micro/macro.

Both ANT and ethnomethodology (Hilbert 1990; Boden and Zimmerman 1991; Taylor and Cooren 1997) have instead framed this approach to the bifurcation of the macro and macro as the concern of actors (including social scientists); the size, context, relevance of action is itself negotiated by actors in situ:

> We cannot distinguish between macro-actors (institutions, organizations, social classes, parties, states) and micro-actors (individuals, groups, families) on the basis of their dimensions, since they are all, we might say, the 'same size', or rather since size is what is primarily at stake in their struggles it is also, therefore, their most important result.

By treating the size of actors as an *outcome* analytic focus is shifted to the work of making them so. Scale becomes a product of members methods, and accessible to the researcher as practical and local talk and action.

For example, Latour has pointed to the laboratory, not as a site for micro activity, but rather as locale for mediating scale:

> the very difference between the 'inside' and the 'outside', and the difference of scale between 'micro' and 'macro' levels, is precisely what laboratories are built to destabilize or undo.(Latour 1983:143)

Laboratories are able to do so by 'bringing the world to them' through accountable transformations into interactionally accessible and manipulable materials: summaries, samples, cultures, photographs, statistics, and maps to name a few (Latour 1986). 'Bringing into the lab' is done through specific practices and material arrangements that simultaneously maintain ties to the world 'out there' and translate it into a more tractable 'small' forms, such as samples and charts. These are 'translation chains' (Callon 1991). A lab in this formulation is simply a particular step in a long chain of translations; a moment of material organization permitting isolation, intervention, and double-edged conquest. To the extent that the world-made-accessible-in-the-lab it is rendered docile, something has been learned, knowledge has been produced. To the extent that this can be made to hold outside the lab – a new technique, a new substance, a new technology – then the world too can be rendered docile.

--

The size of GEON is an *everyday* concern for its participants, for non-participant geo-scientists, for the SDSC, and for its patrons at the NSF. The explicit goals of GEON include building an information infrastructure for the geosciences, innovations in computer science, integrating databases and computational resources, connecting disparate disciplines, mapping scientific knowledge, supporting everyday practice and interoperating with an emerging cyberinfrastructure, to name only a few. If we decide in advance which of these are ‘micro’ and which are ‘macro’ activities, or which are technical and which are organizational, then it is becomes impossible to trace members’ work. However, if we take this as an achievement in practice, talk and material arrangement, this work will remain observable.

GEON participants regularly define their current size, their planned size and organizational plans to bridge the two (see esp. Ch. 4.3 ‘the two-tier approach’). GEON must be made as large or small as those entities it will come in contact with: the geoscience community, geoscience knowledge and data, everyday geo-scientist’s practice and a national cyberinfrastructure. GEON is an ongoing technological project designed to ignore the conventional boundaries of scale, and instead permit its participants to make use of scale in its construction. GEON is to become a ‘large’ infrastructure but it must be made accessible to ‘small’ everyday geoscience users, it must collect ‘large’ databases, and capture ‘small’ schematic differences in ontologies.

There are aspects of the *reduction* that Latour describes as a lab and which mirror activities in GEON, most notably the ontology workshops. In these workshops small teams of geo- and computer scientists are placed in a single room. In this rooms it is possible to bracket out the complexities of 'the geosciences': not the field and its debates but experts and their knowledge, not theory and its various interpretations but individuals and their articulations, not communities and their internal variety but spokespersons as representatives. The heterogeneities of 'geoscience knowledge' and a diverse constituency of heckling geoscientists are bracketed out in order organize for the possibility of creating a clean computer accessible semantic representation (see Ch. 5).

A lab metaphor is most applicable to the ontology workshops where small sets of geoscience representatives are able to speak the knowledges of geoscience and computer scientists can encode these in representations. However, the metaphor of the laboratory is too rigid for the majority of work in GEON, its organizational routines, its technical practices. What 'GEON as lab' fails to capture is how very ordinary is work of collecting information, of just how mundane are the procedures for its assembly and evaluation, of just how casually conclusions are translated into effecting consequences. For example, in the interaction below, taken from a GEON workgroup meeting, we can observe the construction of both a technical and institutional set of demands made from materials at hand:

> Chad: Ok so we need to respond to the NSF's request for an update [pointing to an email printed out in front of him].
>
> Doug: Did you tell them that we are preparing for the GSA [Geological Society of America] meeting coming up, that we'll have a functional demo running by then?

> Chad: Well no, I think Zed wants to hear about something finished, something he can make a slide from. I think he's going to visit a bunch of Texas U's [universities] next week.
>
> Doug: Ok well, Kim can you just forward him some of the geochem ontology viz[ualization]s that we already have. [Kim nods]. It's going to take everything we've got to finish the demo to the point where a real geo guy can come in and play with it at the booth. Not crash as soon as he moved outside what we tell him he can do [smiles]. Then after we can send something to Zed.
>
> Kim: I'm still working on a GUI [graphical user interface], along with Kelly's new comments on the old polygons. It will be ready for GSA. [Geological Society of America]. At least working.

In this excerpt we have a crafting of an institutional demand, a technical trajectory, and an alignment of the two by establishing a line of communication. This begins as a reference to the NSF as whole, 'a request for an update,' and is built up from a particular email available for inspection in the room. The reference is then 'personalized' as Zed, this is sender of the email, and all present know him a geoscience program manager at the NSF. In conversation, references to the NSF and Zed will shift fluidly; for all intents and purposes GEON's responsibilities to the NSF come to be articulated through Zed.

The NSF, later Zed, are demanding an update from GEON, but it is not particularly specified what sort of update. Chad then imputes to Zed an interest for something to be able to display while travelling in Texas, 'a slide', rather than an update of current events, or the most recently completed results. We, as analysts, do not need to inspect the interests of Zed or the NSF; the actors have done so for us. The NSF, or Zed, is travelling to Texas and would like to demonstrate the work of GEON: it is not 'progress' but slides that need to be sent to Zed.

In this excerpt we also have a crafting of technical development. This begins with a reference to a demo, but then becomes personalized by the programmer working on the demo program, Kim, who is in the room. Thereafter this technical activity has a spokesperson: Kim. Note, however, that there is nothing particularly 'technical' about the observable activity or discussion in the room. The 'small' and complex details of the ontology demonstration software are encapsulated in conversational talk. The technical trajectory, or Kim, is also ascribed an interest: to keep working on the development of the graphical user interface (GUI) in the face of the upcoming Geological Society of America (GSA) meeting.

There is an equal hand-waving at institutional demands ("Zed wants to hear about something finished, something he can make a slide from") as at technical activity ("can you just forward him some of the geochem ontology vizs that we already have"). A quick solution is articulated in situ by participating members through the everyday means of conversation. This solution is formulated to meet both sets of interests. A quick connection is built between the interests members have locally defined: the 'technical' geochem ontology slides will satisfy the NSF's 'institutional' needs. By moving from abstractions to individuals an alignment is achieved across institutional demands and technical activity: the NSF gets an update (or, Zed gets an ontology representation to show in his slides), and technical efforts can remain focused on the upcoming demo for the GSA (or Kim works on the GUI and the polygon updates).

It is my argument and the treatment in this dissertation that it is in this manner scale is performed and then managed by members. By bringing together resources at hand, by assigning spokespersons and representatives, or by rendering abstract entities observable in situ, members come to organize the scales of activity: institutional, organizational and technical. Thus this study relies on the treatment of scale as developed within ANT but, in the spirit of ethnomethodology, with an increased emphasis on the situated and everyday practical negotiation of members. We will come to see that phenomena such as multidisciplinarity, basic research, long-term development, and especially infrastructure, are enacted in situ, and framed by actors as endeavours with notions of scale in mind.

--

The tripartite division of the scales – technical, organizational, and institutional – are not actor's categories. Rather these are an organizing principle for this study, reflected in the three parts of this dissertation[12]. However scale itself is of great concern for participants in GEON; it is after all a project with the goal of building a national information infrastructure. Scale has a particular meaning within computer science: computer scientists will speak of 'scaling-up' (and occasionally 'scaling-down') to reference the activities, system design, and hardware requirements to support, for example, an increased number of users, a greater capacity to process or store data, or a broader geographical coverage of service. The actor's term scaling, although technically focused, parallels my own usage by emphasizing the work of accomplishing transitions in

[12] 'Technical' is certainly a term of usage, and is usually laminated with a sense of a detailed practice such as writing software or 'coding'. It is also often used to describe the 'content' of domain knowledge. Organization has come to be used within GEON, although least frequently of the three. Its usage differs substantially from that in this dissertation. Often organization is a term noted with distaste as referring to the business sphere. The term institution is used very similarly to my own. Actors will refer, for example, to the NSF or USGS as well as the various home 'partner institutions' of the GEON PIs. All this said, while there are multiple distinctions in GEON, technical/organizational/institutional are not the organizing principles of actors, but rather they act as such for this dissertation.

service, capacity or coverage. Scaling for computer scientists points to i) a kind of planning (often including research such as requirements testing), ii) securing of material resources (including human resources), and iii) the activity of implementation. In this sense scale is very much an actor's category and going concern in GEON[13]. In this sense, GEON actors share in my usage of scale as an accomplishment.

The argument that scale, or 'the size of actors,' is an outcome should not be interpreted as an attempt to establish a microstructural foundation for macrostructure. Rather, it is that any sort of scale should be considered an accomplishment. Many commentators take ethnomethodology to be focused on the micro and take its advocates to argue for a macro which emerges from face-to-face interactions. However this misconstrues the radical quality of the ethnomethodological, and what has come to be the ANT, approach. I take all scale as an accomplishment, and remain strictly agnostic about the particular 'scale' of *interaction* (except to take it to be observable)[14]. Rather scale emerges out of routine inferences, designations and summary procedures – this applies equally for micro or macro levels, or in this study technical and institutional activities. Scale is in potential reach of every actor, it is in a change of register, a stretch of technology, or in giving voice to institutional action.

---

[13] Scale as a topic has often been problematized for ethnomethodologists: "The notion of integrating micro- and macro-sociological activities is a construction of the researcher and, therefore, is not a concept we can attribute to those members of a group whose normal activities create micro-macro integration." (Cicourel 1981:52). However, as we will see, in many senses scale is an endogenous concern for GEON members, generated by the talk and practice of participants.

[14] The internal consistency of ethnomethodology should not be over-represented. Many ethnos have embraced the label of being focused on the micro. In particular Aaron Cicourel has made many explicit statements on the notion of the micro and macro and has argued that the macro emerges out of micro interactions by various processes of 'decontexualization,' aggregation, or interpretation (see esp.(Cicourel 1981)). In this study of GEON there is just as much work dedicated to cutting out, and then aligning, what comes to be a (micro) technical practice with a (macro) institutional set of demands. However Cicourel's work of observing the production of macro still remains exemplary, and substantially informs my method.

1.3 – Taking Ethnomethods Seriously

To repeat, ethnomethodology is the study of 'ethnomethods,' or the everyday but specific methods of people in producing forms of order, coordination and accountability. The term 'everyday' is of special import. What is considered key in ethnomethodology is the topicalization of specific achievements in situ: how do musicians play together, how do jurors deliberate, how is a photograph made relevant scientific evidence? In each case it is the 'doing' which is centrally placed, rather than independent analysis of meaning, identity or surrounding customs. It is these achievements, the doing, which are taken by the ethnomethodologist to be of greatest concern for 'members' rather than, for example, the reproduction of order, structure or society. Moreover, it is not possible to reduce to such abstractions without losing the phenomenon itself.

In his *studies of work* Garfinkel has described these as the "unique adequacy requirement of methods" which make up concrete instances of activity such that there can be no translation of ethnomethods to the various common denominators of social science e.g. *ideology*, *interest*, *structure*, or for that matter, a generalized notion of *practice*[15]. Rather, the domain of ethnomethodological inquiry consist solely of actor's situated

---

[15] Garfinkel is said to have accused most of sociology as 'missing' the very 'what' in its study of phenomena. Mike Lynch recounts Garfinkel's critiques of, for example, Becker's study of dance band musicians. Becker speaks of "the linguistic and customary practices by which dance band musicians attempt to distance themselves from the "squares" who make up the typical audience" and Garfinkel has accused him of never discussing how they manage to play music together. "The interactional and improvisational "work" of playing together […] was somehow 'missed'," (Lynch 1993:271). In STS (particularly the sociology of scientific knowledge) it has become a common phrase to explore the 'content' of science, Lynch notes that this is a mystification of the term, while in contrast Garfinkel's 'missing what' encourages an empirical analysis of action without demarcating it as 'content'. Similarly, practice can become a generic term for all kinds of activity (Vann and Bowker 2001), whereas ethnomethodology forces a concentration on the unique adequacy and consequences of an action.

actions and talk which perform collectively objective phenomena such as order, structure, scale and do so in a manner particular to that activity.

Within ethnomethodology there is no set of methodological precepts that can be generalized to the study of particulars, in fact ethnomethodologists have been actively resistant to such attempts while also recognizing the occasional need for such theoretical summation. Rather, a corpus of studies have been produced which can be made canonical relative to particular studies, or as Michael Lynch has noted "This basis [in previous studies] is not a firm and stable outline of analytic rules and techniques, as the shape of its suggestiveness often emerges in-the-course-of or consequent to particular researches," (Lynch 1985:6). It is in this sense that I describe ethnomethodology as a methodology. It is not a specific set of tools and actions for research, but rather a sensibility and approach to inquiry.

Below I outline the three primary analytic sensibilities drawn from ethnomethodology which have substantially informed this study; they are *indifference*, *the endogenous social,* and the ethnomethodological treatment of *context*.

Indifference and Sciences

A central principle in this study has been the ethnomethodogical notion of 'indifference' (Garfinkel and Sacks 1970). This is a levelling of epistemic privileges across kinds of lay and expert accounts. The term 'indifference' can be misleading; it does not refer to a lack of interest or personal involvement on the part of the researcher so much as a methodological stance with the intent of rendering activity observable. This is particularly useful in the face of practices that are often demarcated as special, such as science or even social science.

Ethnomethodology does not permit an explanation of scientific activity as a special case of action, but instead requires a treatment that while specific (to the activity, here a scientific one) remains congruent with practical -action, -rationalities and -accounting more generally. There have been strains of ethnomethodology which distinguish science as a special case (which at times has seemed to include Garfinkel). Michael Lynch has traced this to social phenomenologist Alfred Schutz's 'scientific exceptionalism'. Schutz's work played a significant role in Garfinkel's formulations of ethnomethodology, and he often credited with having reintroduced Schutz's work to active discussion with sociology. Lynch calls strains which demarcate between science and non-science 'proto-ethnomethodologies'(Lynch 1993). In these cases what are otherwise phenomenal or practical descriptions fall into a cognitive idiom upon encountering terrains defined as 'science' or 'social science.'

However ethnomethodological research within the set 'science studies' has usually displayed a *symmetrical* indifference to explanations based on scientific exceptionalism, or, at times, has topicalized these as accounts. The key point here is not that the sciences do not involve particular activities or methods but rather that analytic descriptions should call for *no more* specific an explanation than of any other endogenously constituted accountable setting:

> indifference is not equivalent to denial or opposition … Nor does it imply that there are no distinctions to be drawn among sociologists', coroner', physicist's, or any other lay or professional methods. Rather, it [ethnomethodological indifference] states that any such distinction is contingent,

> locally organized, and in a peculiar way discoverable. (Lynch 1993:142)

Lynch's approach to indifference is similar to Gieryn's boundary work (Gieryn 1983; Gieryn 1999), not a denial of a boundary (non/science) but rather a tenet to observe its in situ performance, maintenance and consequences. Lynch broadens the kinds of distinctions to those which are of greater relevance in this study, not simply science or not science, but the boundaries of kinds of sciences, expertise or professional knowledges.

There is a strong parallel between ethnomethodological indifference and what is known as the symmetry principle in STS. Arguably these two lineages intersect at ANT, with Latour's formulation of a 'second symmetry principle' and the empirical studies of 'sociologics' (Latour 1987) in science and technology projects. Put briefly, the 'first' symmetry principle argued that sociologies of knowledge should treat the production of truth in the same manner as we explain what comes to be considered false. This was formulated in response to what Bloor has called a *sociology of error* in which 'social influences' had come to be understood as contaminants in an otherwise ('pure') scientific activity. The second principle extended this symmetry of treatment to the production of other phenomena, such as agents and objects, and most importantly for this dissertation, social and. technical. In this study of GEON and CI, an a priori distinction social/technical would have proven doubly concealing. Symmetrically, or indifferently, I have studied across these categories and observed GEON's practical work of organizing, knowledge capture, or aligning with larger visions of CI.

GEON is cyberinfrastructure for the sciences, and it self-avowedly includes the participation of many sciences. However, most of the activity in GEON has not usually received the label ‘scientific,’ that is, in most cases work in GEON has not been given any special label at all but rather more vernacular terms such as ‘workgroup meetings’ or ‘All-hands Meetings’ (see Ch. 4.1 and Appendix B). Even the ontology workshops, which most directly touch upon scientific knowledge, were not themselves considered scientifically informed activities but rather as an art, or perhaps an organizational routine Ch. 5.2). GEON, however, is thought to touch upon many scientific or expert enterprises, namely sub-disciplines of computer, geo and social sciences.

As Marilyn Strathern has noted in her studies of (and participation in) collaborations considered multidisciplinary, that there is a tendency in these groups to impute a model of expertise:

> In a many disciplined context, each expert becomes a representative of his or her discipline. Indeed, experts will be turned to for their ‘traditional’ knowledge, for specialist wisdom assumed to be already in place (Strathern 2004:5)

This has certainly been the case in GEON. GEON has expert representation in geodynamics, information visualization, geophysics, active tectonics and knowledge representation. Each of these is a topical harbour of expertise: a domain. GEON also has a sociologist. The model of expertise – a representative of sociology, an expert on the social – has been distinctly uncomfortable for this ethnomethodologically informed researcher who simply does not think much can be expertly said about a generalized social.

The Endogenous Social

A running theme and empirical concentration of this study is the generation of 'a social' at all scales of action in GEON. 'Social,' 'cultural' and 'community' are all categories regularly invoked within CI circles, as are the methods and findings of the various 'social sciences.' The participation of 'social scientists' has also become common, and is increasingly more so. Participants who in no way consider themselves as social scientists regularly draw the languages endogenous to sociology, history, philosophy or communication. In her studies of disciplinary borderlands, anthropologist Marylyn Strathern has noted:

> For one consequence of this inter-folding of expectations is already evident in certain types of investigation. Anthropologists once regarded it their job to elicit reflectivity from their research subjects, but nowadays they are often presented with a high degree of already cultivated self-awareness and self-consciousness […] presented with what one might call indigenous social analysis (internal interest in analysing the structure and role of the organisation), and presented with a desire to engage with the social environment in a responsive mode (openly advocating learning).
>
> (Strathern 2004:10)

This comes as no surprise to ethnomethodology which is concerned with the methods of the people and in that research program has always focused on endogenous analysis. Similarly within ANT, which has described the phenomena as a 'socio-logic'(Callon 1980; Latour 1987). However there is a specific configuration in the case of CI where the Atkins Report has actively promoted the importance an explicitly defined social sphere:

> The conduct of science and engineering is a social activity, pursued by individuals, collaborations, and formal organizations. Any enlightened application of information technology must take into account not only the mission of science and engineering research but also the organizations and processes adopted in seeking these missions. (Atkins 2003: 14-15)

Furthermore, the report defines social scientists as key participants in a *tripartite partnership* for CI which also includes domain and computer scientists:

> Building, operating, and using advanced cyberinfrastructure must be done in a systemic context that exploits mutual self-interest and synergy among computer and information, and social science research communities who see it as an object of research, and other ("domain science") research communities who see it as a platform in service of research. (Atkins 2003: 7)

CI has actively taken up the category 'social,' seeks partnerships with 'social scientists' and appropriates the findings, language and methods of multiple social science traditions. Within GEON this has included my participation of an ethnographer sociologist, but as we will see, is also constituted as a pervasive concern with 'the social and cultural' aspects of *community* building, the *adoption* of novel technologies, *organizational and intuitional* differences across the geosciences, or *reward structures* for geo-scientists, just to name a few.

--

Just as ethnomethodology has come to be indifferent about 'science' as a privileged category of action, 'sociology' too has also placed under the same scrutiny. The history of ethnomethodological studies is replete with turning attention back upon sociology: positioning ethnomethodology as *indifferent* in its valuation of professional and lay sociological explanations, not for the purpose of denigrating either, but to make both empirically observable.

One of Garfinkel's nemeses has been the notion of a 'cultural or judgemental dope'. This is any explanation which portrays people as determined, whether by structure, background or demography; or, as proscribed by norms or prescribed through the following of scripts. More specifically the 'cultural dope' is the rhetorical tool which makes possible a sociological explanation based on a notion of a common culture with predefined roles, rules or norms, it is the "man-in-the-sociologist's-society" (Garfinkel 1967:68). Garfinkel's writing on the cultural dope are infused with a sense of indignation about the explicit and implicit models of agency in, then dominant, Parsonian sociology. This sense remains in ethnomethodology today which often celebrates the various agilities of its 'members'. People are understood to be continually:

> engaged in having to decide, recognize, persuade, or make evident the rational, i.e., the coherent, or consistent, or chosen, or planful, or effective, or methodical, or knowledgeable character of […] activities to their inquiries. (Garfinkel 1967:32).

In ethnomethodology the phrase 'lay sociological explanations' is meant to shift focus to the member's work of accounting conducted in the everyday activities of summation, coordination and delegation. Everyday accounting becomes a mode of enacting the social sphere and the cultural dope is re-written as the lay-sociologist. For the ethnographer this re-writing is only possible with a sensitization to actor's practical reasoning in the data collection process.

One of the strongest 'levelling' arguments from ethnomethodology points to the shared vernacular language between sociological and lay analysis. This line of reasoning is influenced by readings of Wittgenstein which link language to forms of life and styles of reasoning. Attempts to construct meta-languages outside the vernacular, independent methods and extrinsic criteria for the evaluation of evidence simultaneously rely on the languages of objectivity but at some point must also make ties back to a 'shared vernacular' base of knowledge and reasoning. This argument was put forth empirically in Garfinkel's study of sociologists coding clinical records.

Run in parallel with a study of intercoder reliability, Garfinkel's goal was not to evaluate the truth of the codings but to tie actual codings of data to accounts of that action. The practices and accounts combined would reveal, to the ethnomethodologist, a set of "games" (Garfinkel 1967:20) or methods which link the two. One finding was that the methods for coding relied on a particular theory of representation. The data in this case were records in folders which were made in the process of clinical activities at UCLA's Outpatient Clinic. In order to effectively code the data the coder

> must treat actual folder contents as standing proxy for the social-order-in-and-of-clinic-activities. Actual folder contents stand to the socially ordered ways of clinic activities as *representations* of them; they do not describe the order, nor are they evidences of the order," (Garfinkel 1967:23)

This activity necessarily involved assumptions about the means of data production: "the coder must know the order of the clinic's activities that he is looking at in order to recognize the actual content as an appearance-of-the-order" (Garfinkel 1967:23). Another way of stating this is that coders had to formulate a 'context' for the production of the records.

In order to interpret the data it is necessary to posit a regularized routine of inscription, and that this is done so with the intent of record archiving[16]. This context was built using information that was not contained in those folders (the data) themselves but rather relied on a 'common vernacular understanding' of what goes on in record producing settings such as the clinic (the ethnomethodological treatment of 'context' is elaborated in the next section). Garfinkel called this 'ad hocing' and argued that this was inescapably so; 'improvement' in study design would invariably shift ad hocing to another scale. This was not intended to undermine 'objective' sociological research per se; in fact the ethnomethodological study does not directly speak to the issue at all[17]. In contrast the intercoder reliability study, which compared agreement of multiple coders to the same record, was explicitly formulated to generate accuracy, objectivity or lack thereof in the study design.

In discussing ethnographic studies of science Lynch makes a similar argument, pointing to shared vernaculars across social and natural scientific domains:

---

[16] Later in the same text, Garfinkel argues for 'good reasons for keeping bad records.' Similarly other studies of record keeping and producing practices have shown that archives serve many purposes aside from those which are formally or explicitly stated and available to a 'vernacular understanding' : (Bowker 1988; Star 1988; Berg 1996).

[17] This study has been used through various rhetorical strategies to undermine positivist understandings of sociology or quantitative/statistical research. If we take the study to be about how coding is 'really done' this is a reasonable conclusion. If we take it that "whatever they did could be counted correct procedure in *some* coding 'game'" (Garfinkel 1967:20), then the study speaks not at all to the issue of objectivity. Garfinkel himself appears ambivalent. While I am generally in support of those arguments developed within ethnomethodology which undermine fundamental distinctions between lay and professional sociological analysis this is not the place to attempt to generalize to a methodological principle. Rather, here I am interested in demonstrating how these tenets can add meaningfully to the accounts of practical action within GEON, the SDSC and CI-circles more generally. It is the composition of expertise and organization found in this study which have made this toolset particularly relevant.

> the problem is that most of the terms of the tribe[[18]] are ours terms as well, since (1) they are integrally part of the discursive grammar through which scientist's activities are conducted and made intelligible *in situ* and (2) they are embedded in social science vocabularies for making coherent descriptions and explanations of trivial activities of diverse kinds," (Lynch 1993:97)

As a methodological technique this is doubly important in the case of GEON because of my role as a participant sociologist, but also because outside my own interventions social scientific research and findings have been endogenous to GEON's activity.

Throughout this ethnography of GEON there has been talk and some preparations to add a social networks analyst to the GEON team. Social networks is a primarily quantitative, and partially qualitative, social science research approach which tracks kinds of human-human connections: 'social networks' (Chin, Myers et al. 2002). The purpose of this research would have been to understand: who is using GEON? Who do people in GEON talk to? What are the natures of the existing linkages across the disciplinary linkages of geo-scientists? How is the growth of GEON's user base proceeding? This study was never enacted within GEON, but it still may be conducted in the future.

When the possibility of this research approach was mentioned to me by other participants, it was with some trepidation. Would this research be akin to stepping on my social-science turf? My answer has always been a resounding no. A social networks analysis of GEON seems a natural extension of the trefoil setup of Atkins' cyberinfrastructure: computer, domain and social science collaborations in designing

[18] The particular 'tribe' of reference here are scientists. Lynch is making a backhanded reference to Latour and Woolgar's famous call in Lab Life for a metalanguage which will "explain the science … without resorting to any of the terms of the tribe"(Latour and Woolgar 1979).

large-scale infrastructure for the sciences. In fact, I looked upon the possibility of an additional 'social scientist' with some relish and prepared myself to extend this study to encompass a social networks analysis:

> There can be nothing to quarrel with or to correct about practical sociological reasoning […] Ethnomethodological studies are not directed to formulating or arguing correctives […] Although they are directed to the preparation of manuals on sociological methods, these are in *no way* supplements to "standard" procedure, but are distinct from them. (Garfinkel 1967:viii).

As with GEON's methods for making geoscience knowledge visible, or for aligning with NSF's institutional vision of CI, social networks analysis would have been an additional method in making visible the emerging GEON user base and the geoscience community.

The practices and rationales of a social network analysis would have become another method in the repertoire of GEON's construction. The object of a social network study would be to make visible the user communities and their access patterns to GEON participants; the means for producing these knowledges and the accounts of their production would have remained an observable and in situ practice accessible to my own methods.

'The social' is shorthand for the particular configuration of a sphere of knowledge and practical activities understood to be critical in the development of CI. Social scientists are active participants in shaping the understanding of the social, the roles of social science within it, and place for its interventions: see for example Kling and Scacci for a definition of information technology as organization (1982), or Kling on a programmatic call defining role of social science research within informatics (1999), Button on the contribution of ethnomethodology to human computer interface (Dourish and Button 1998; Button 2000), and Dourish on the purposes of ethnography within information system design (2006). Social scientists are participants in defining 'the social' of GEON and CI. However, they are only one set of actors amongst many. The language of the social has infused efforts to build cyberinfrastructure, today the entire of participants regularly frame talk and action in terms once considered the domain of the various social sciences.

Actor's Context

In denying the 'cultural dope' model of the actor ethnomethodology places a great deal of analytic power in the hands of its informants. All the resources of the researcher – what we call method, theory, empirical evidence and reflexivity (Lynch 2000) - must also be also attributed to the actors. While ethnomethodology focuses on talk and practice it does not take the everyday unfolding of the life-world as simply a stream of unreflexive action, but also as a continuous impartation of meaning through the negotiation and attribution of categories, the production and transformation of coordination mechanisms, and the establishment of durability, or order, through material arrangement. To the model of the actor as sociological analyst in the previous section I would add also the actor as lay-historian (Ch. 3.1) and lay-science studier (Ch. 5.1), this is to say that histories of GEON, CI or the SDSC, along with theories of knowledge actively play a role in GEON's daily construction.

Here I will focus on a single actor's category of particular importance for this study: context. In this interview excerpt below this GEON PI is discussing the collaboration between information and earth scientists in GEON. He poses geoscience as the 'context' for the application of information technologies:

> But the thing with GEON trying to work with scientists ... The scientists think they're going get something that they can use right away, to produce science. And the IT guys are thinking … well they are using the scientists to get some interesting *context* and 'we'll now build this nice little tool.' 'And I want a nice tool, if the tool isn't cool then what the point' [small chuckle at the rhyme].

The GEON geo/IT collaboration is premised upon building tools that are useful within geo-scientific practice, but for information technologists geology itself is understood as 'a context'. But briefly, here context means a) a set of rich computational problems; b) which are meaningful for particular actors outside computer science, and; c) which will be addressed by information technologists through particular 'applications' of computer science. In computer science a more common, but in many ways quite similar, terminology for 'context' is 'the domain' and in GEON 'the geosciences' or simply 'science'.

As Garfinkel showed in the case of sociologists coding clinical data, the key point here is that 'contexts' are themselves at play within the repertoire of actor's planning and abstracting activities. More than theoretical tools of the sociologist the everyday in GEON can be characterized by the continuous *framing and reframing* of 'contexts' (Goffman 1974[1986]; Callon 1998). Technologies are built in relation to contexts such as targeted future users (see Ch.4.2, or Woolgar 1991) and the formulation of computable geo-scientific research questions (see Ch.2.3, and Fujimura 1987). The vigilant NSF is regularly looking over GEON's shoulder ensuring that technologies are not built once again ‘re-inventing the wheel’. To do this GEON must be developed ‘in the context of’ concurrent cyberinfrastructure projects and the multiplicity of information projects ongoing within the geosciences. Planning the future for GEON is conducted ‘in the context of’ congressional funding of the NSF, the projects reception by geo-scientists, and prospects for attachment to geoscience institutions.

As Bruno Latour has prescribed for study the emerging technologies “interpretations of the project cannot be separated from the project itself, unless the project has become an object,” (Latour 1996:172). The context of a project is part of the ongoing action of enacting the project.

Another way of expressing the same methodological tenet about context is to return to the notion of the observable in ethnomethodology:

> relevant contexts should be procedurally related to the talk said to be contingently related to them. That is, there should be some tie between the context-as-characterized and its bearing on ‘the doing of talk’ or ‘doing the interaction’ (Schegloff 1987, p219).

A strong way of restating this would be to say that ethnomethodology does not permit a use of ‘context’ as it has been traditionally used in sociology: a stable frame in which the action takes place. In sociology this stable frame is often *made of* history or larger events of a national or global scope and actors themselves may be (explicitly) unaware of these. These contexts are the constructs of the analyst, and in ‘traditional sociology’ they are understood to add explanatory force to the ongoing action. A more subtle reading of Schegloff is that contexts are emerging in tandem with the doing of the interaction, Schegloff demands that contexts be ‘in the talk’ of actors, or as Garfinkel might say, ‘inspectably so’.

In every sense, the particular techniques available for GEON's actors to build contexts are the same as those in the sociological repertoire, they do so on a continual basis as a part of everyday action, or through more formal methods drawing on survey, statistical or ethnographic technique. As noted, for a period negotiations were underway within GEON to have social networks analyst study relations with non-participant domain scientists. The goals of these professional sociological investigations would have been to furnish materials from which to build the context of GEON's reception in geoscience.

For the most part, however, building contexts does not require social science interventions. Many of documentary methods of the sociologist or historian were regularly used within GEON as par for the course such as gauging NSF's disposition towards future financial support by reviewing public statements or documents on congressional funding for the NSF, or; tracking the internal reorganization of the NSF relative to CI both through formal feedback from the NSF or through casual conversations with those program directors connected to GEON.

These are the same means available to historical/ethnographic data collection. Admittedly, GEON collection is conducted with less concern for archiving, formal analysis and future communication in a manuscript – their purposes are practical and specifically interested. On the whole daily participants have more regular interactions with the NSF, with the technical development, and with other members – this is the privileged view of the participant.

In re-embedding understanding of 'actors building contexts' within the larger argument of this dissertation we can see that building a universal informatics is reliant upon not only the production of a common infrastructure but also the production of multiplicity of local contexts: “in fact, the trajectory of a project depends not on the context but on the people who do the work of contextualizing,” (Latour 1996:150).

This is powerfully expressed the boundary work: IT/domain. For participants ‘contexts’ (such as: ‘science’ or ‘domains’) must be understood empirically and infrastructure must be applied to these. In the introduction we saw that the Atkins Report refers to the ‘conduct of science and engineering.’ The Report poses this as crucial to producing practical information technology applications for the geosciences, but it also falls outside the focus of that report. This is because scientific conduct is considered particular. In GEON, and other CIs, knowing the domain becomes the active endeavour of not only of information technologists but also domain scientists themselves. The means for knowing the domain come in various forms.

Perhaps more consequentially and certainly more enduring modes of context building are the formal procedures of ‘user studies,’ ‘requirements testing’[19] and GEON’s own knowledge representations through ‘community ontology workshops’. See Ch. 4.2 and 5.2 respectively. These studies come to have the force of method behind them, and most importantly their end products are inscribed in presentations, reports, and articles.

GEON actors have been avid context builders of various kinds. The tripartite analysis of institutions, technology and organization will allow us to see kinds of ‘external’ and ‘internal’ contexts that are built within GEON, which include

[19] For example social informatics researcher Finholt et al. were invited to conduct requirements testing on the NEES project, various publications are pending, but more important to the activity within NEES these reports have been presented at its various meetings and been incorporated into many slide sets. I become aware of this research when one of NEES’ lead PIs presented at the SC’04 community building workshop, drawing on many of the findings from NEES studies.

a) Images of the current and future trajectory of cyberinfrastructure. How best to secure mid or long-term funding. Considerations of CI's failure or perhaps projects which may come in its wake. The San Diego Supercomputer Center must continually align its rhetoric, service model, and technologies with emerging agendas at the level of the NSF. GEON must consider future funding sources as its short five year cycle has always loomed near.
b) Geo-scientist computer users, who are they, what do they want to do, and what computing resources might they already have available? At times these users are discussed outright by computer scientists, at other moments GEON's geo-scientists come to stand-in for the community, and routines have been developed building representative samples of the geoscience community through outreach or participatory design.
c) Incentives for work: why might computer scientists or geo-scientists not be interested in participating in cyberinfrastructure? What are the rewards for contribution in computer and earth science? What are the career paths available for graduate students contributing to GEON or CI? These concerns permeate CI activity, it is one of the primary frameworks by which NSF calls for funding are designed, a concern of domain scientists about their own careers and their graduate students and solutions are built into the very organization of GEON activity in the two-tier model.

Conclusion: Lamination, Elaboration and the Limits of Actor's Categories

As a form of a conclusion to this methodology chapter I would like to address some of the limitations of the approach I have outlined, as well as some of the compromises I have chosen to make in writing this dissertation.

The emphasis on actor's categories and resources, ethnomethodological indifference, the endogenous constitution of the social, and scale will serve to establish a theory of activity which can account for the range of *methods* at play within GEON. Participants in the project include a range of experts such as philosophically informed knowledge engineers, theoretical computer scientists, practically oriented programmers, practicing geoscientists and managing geoscientists. CI is generally framing as a collaboration of computer, domain and social sciences. This has led to particularly reflexive set of discussions within GEON.

Participants take GEON to be new, experimental, and consequential: GEON has no road path, or more precisely it has had many; GEON is a "prototype" with a built-in five year funding cycle, at the end of these five years a new plan must be formulated; with calls for revolution and paradigm change discussions of implication have ranged from paralyzing the geosciences to its complete reinvention. In GEON there have been periods of immersive activity in which goals and means seem clear, while at others moments the project's purpose, identity and approach have come under radical questioning. An analysis of GEON requires an inclusion of this uncertainty and questioning, and of its self consciously bootstrapping activity. In this sense ethnomethodology's conceptualization of action which includes continuous negotiation of categories of 'what are to be taken as foundations' is particularly apt for this study.

However, there are clear limitations for a study focused on 'following the actors,' tracing the negotiation of categories, or deeply engaged with member's practices. For example in following actors one must make choices about who to follow. That is, it is not practically possible for me possible to follow all GEON actors in all their various directions. One possible outcome is a privileging of the actors considered most important. In discussions of ANT this has been criticized as a 'managerialist' approach (Haraway 1988; Star 1991; Law 1994). This is a focus, even glorification, of particular actors in acknowledged positions of power. The critique is that the work of the 'little people,' such as support staff or technicians is rendered invisible, regardless of their contribution (Shapin 1989; Orr 1996).

In some senses this criticism can be applied to my research. Thanks to the length of the data collection period and my location in San Diego near the SDSC I do not believe this has been a significant problem relative to the sections on organizing and the technical (Part II and III). My access and relations with the 'little people,' such as administrative staff, technicians and nose-to-the-keyboard programmers has been greater than to the managerial GEON members. On the other hand my contact with the institutional aspects of GEON life, in particular at the NSF or USGS, has been distinctly limited to a series of 'key actors' such as program, division and assistant directors at the NSF. Similarly, while I have had extensive contact with earth science PIs and their retinues brought to earth science meetings, I have had little access to the 'distributed little people' at each PIs home institution.

Another limitation to following the actors and their categories is that there very well be places they chose not to go. The crevices of GEON's history may not be accessible to this methodology. These criticism have been set forth, in particular, by analysts in the social worlds tradition (Fujimura 1988; Star 1992; Casper and Clarke 1998). For example, a first proposal of GEON was rejected. In the process of re-writing and submitting the second, successful, iteration many earth science PIs were dropped. Over the years very little has been made of this in GEON; in the course of everyday practical work members rarely discuss these once-GEON participants. In strictly 'following the actors' it has been difficult to draw out consequences in member's terms (this example is discussed in greater detail in Ch. 2.3.).

--

Following from these sorts of criticisms a question often arises about the role of the analyst in ethnomethodology and ANT: where does all this attribution of analytic resources to the actor leave the professional sociological analyst? Supporters of a 'strong thesis of actor's resources' deny the analyst usage of conceptual resources such as scale, context, or the social. They argue that in order to follow actor's accounts and faithfully portray members practice the analyst must sustain an extended effort to exclude exogenous commentary.

However, it is distinctly unclear how one conclusion follows from the other. After all the argumentative and empirical work in arguing for indifferent allocation of analytic resources between lay and scholarly actors, why would we then turn around and ban the same capacities from the social scientific repertoire? Despite strong ethnomethodological influences in his own work, Latour, in his review of Michael Lynch's ethnomethodological lab study, noted:

> Why on earth should we be less free than the people we study/work with? Aren't they constantly changing instruments, focus and scale? The very neurobiologists whom Lynch studies happily mix together anatomy, physiology, electron microscopy: why can't we go from shop talk to photographs, to scientific writing, to science policy, and then back to instruments? Why don't we have the same rights as the scientists we observe? And, above all, why don't we have the freedom to be 'self-explicating'? Why should we be the only scientists to be stuck at one level, and with only one focus? […] He really believes that there exists some sort of inherent distinction between small-scale and large-scale phenomena. (Latour 1986, p.548).

Locating interactions in context, characterizing a mode of scientific activity, or speaking of scales are some of the most powerful tools in the sociologist's toolbox. I argue that to do so, as an analyst, does not foreclose a careful attention to comparable endogenous work. Thus this study is not an orthodox ethnomethodological study. Rather I seek to draw on the sensibility developed within ethnomethodogical to address the empirical matter of this study. Actor's accounts are always partial, highly indexical, and rely on a great deal of layered interactionally built understandings. To fill the gap between actor's accounts and a need to communicate to a broader audience through this text, I will rely on two notions 'elaboration' and 'lamination'.

Elaboration is the creating of fuller picture of laminated actor's accounts. The notion of lamination is borrowed from ethnomethodologist of organizing Deirdre Boden. She notes that in organizations accounts proliferate, they are repeated endlessly, adjusted to the situation and come to acquire built-up meanings through overlapping interactions. This is case with a series of stock phrases that have appeared over the years in GEON: "we must learn something about the Rockies!" (Ch. 2.3.) "low-hanging fruit" (Ch. 4.3) and "we don't make standards" (Ch. 5.2). Any GEON participants will recognize such phrases, as they have been repeated in meetings, conferences, and daily conversations.

The notion of lamination demands a treatment of the immediacy and situatedness of an utterance, while also acknowledging that its entire content may not need to be generated locally but comes come from many previously enacted interactional episodes. It is a mediating notion between the persistence of meaning (in culturalist approaches) and demands of a situated account for the generation of meaning (in ethnomethodology). Laminated categories are built up over time through multiple interactions by the same actors. Such a laminated category is still understood to be deployed locally and its boundaries negotiated in situ, however lamination helps us understand how meaning may overflow a bounded interaction.

Each of the laminated 'stock phrases' in GEON raises the spectre of greater depths of meaning than are inspectable in a particular set of utterances. In the face of such activity, I have generally chosen to 'elaborate' laminations rather than using large tracts of text to trace their emergence. In Boden's work she traces over multiple encounters the lamination of a phrase – a great deal of her book is dedicated to such tracings. In this text I would like to cover greater ground, and so I elaborate laminated meanings rather than tracing them in detail.

For example, members of the SDSC often draw a laminated trajectory for the centers: "just as with IBM we are moving from supercomputing to services" (Fran Berman, director SDSC, town hall meeting 2004). An unstated (and potted) history of IBM runs in this comparison: it is a company which is understood to have recently pulled itself from the brink of financial collapse through switching emphasis from hardware and software production and sales to services in the form of consultation. Today, the understanding is: IBM's servers and software suites are sold and implemented in tailored packages as recommended by technical and organizational consultants. With this in mind the SDSC as providing 'services' to the sciences and engineering can be understood as a combination of consultation and technical development and implementation. This is an elaboration of the comparison between SDSC and IBM. I have collected this understanding over years of attending SDSC 'town hall meetings,' hallway conversations, slide-shows and dinners.

Elaboration is a complement to the methodology of following the network (Latour 1987), permitting greater analytic range than extended tracings of particular laminations in situ. It differs from the notion of context in that elaborations do not 'explain away' actor's accounts, but rather simply collect them creating fuller images for the reader.

In this sense I have returned full circle to the approaches of grounded theory. Produced in iterative relation to empirical data collection, with close attention to the enactment of actor's categories in talk and practice, the ultimate goal of this thesis is in generating substantive theory.

– Chapter II –

Generating a 'Push' for Cyberinfrastructure: Enacting and Organizing Disciplinarity at the NSF

–

A system of power reveals the world in something like the Heideggerian sense. It opens a certain angle of vision and defines a corresponding realm of objects. This foundational work of power does not contradict the pursuit of truth but makes it possible by orienting research in a specific direction

--Andrew Feenberg

Government is the right disposition of things, arranged so as to lead to a convenient end.

-- Guillaume de La Peirre[1]

–

> I don't know what you call it but it's one of the different terms we use here at the [Supercomputer] Center level and we call it 'the push and the pull'[…] Do you get the applications to pull you? […] One way you can say the domain are pulling on us to deliver some technologies for them, or deliver some functionality. The push is where we tell the domain, 'Now here's this cool thing. You really should use this.'

-- GEON IT PI

The notion and terminology of a 'push and a pull' are actor's categories in GEON and cyberinfrastructure (CI) circles more generally. Their meanings are roughly drawn from the language of economics: push is equivalent to 'supply' while pull reflects 'demand'[2]. So for example, in the epigraph to this chapter an IT PI is equating push with technology developments 'looking for' an application. Here computer scientists are

[1] 1567, quoted in (Foucault 1991:93)..
[2] This said, the uses of these terms are colloquial. They are often articulated completely in the reverse and occasionally their meanings do not map onto conventional uses in economics.

attempting to convince the domain to use a novel tool or technique: “Now here’s this cool thing. You really should use this”. Pull are the ‘needs’ of the domain, a request on the part of scientists for computational resources or data visualization tools to be applied to basic research: “the domain are pulling on us to deliver some technologies for them, or deliver some functionality”.

As with economics (Callon and Muniesa 2005, see also Callon 1994 on the public good), in CI circles push/supply and pull/demand are relatively unproblematized. In GEON and at the Supercomputer Centers they have become part of a commonsensical vernacular. In both Ch. 2 and 4, the use of these terms are *as actor’s categories*; this dissertation *is not* dedicated to the problematization of these terms but rather to their respecification (Garfinkel 1991). The goal is to show the *generation of a push and pull* for cyberinfrastructure in the earth sciences by institutional actors: how have the geosciences come to have a *need* for a unified information infrastructure? And, how have the new information technologies been made relevant for basic research about the earth? This chapter focuses on the generation of push at the institutional scale of the National Science Foundation (NSF).

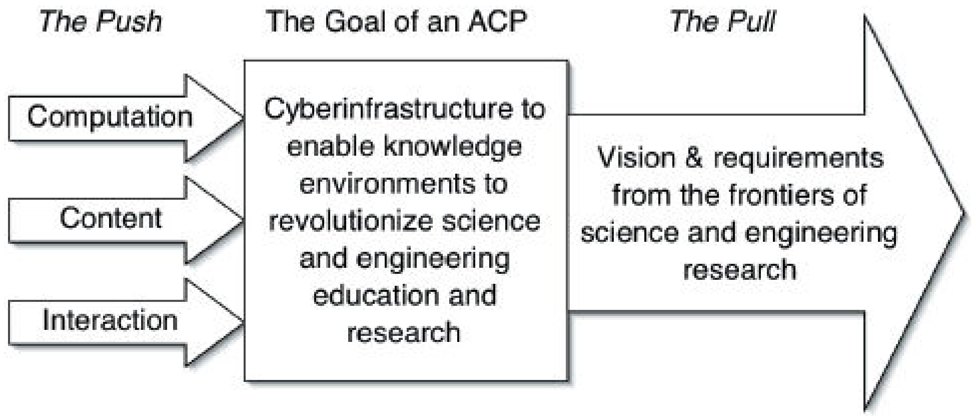

Figure 5: The push and the pull of the Advanced Cyberinfrastructure Program (Atkins 2003:33). Although the particular usage varies the notion of push and pull are endogenous. The Push is usually technological advance or increased complexity and pull is a need in the domain.

--

GEON was born out of a debate about the nature of information technology research. It is a debate about the organization of science, of relations between the disciplines and the support of science through funding bodies and coordinating institutions. Cyberinfrastructure is a particular crystallization in that debate. It includes a model of support for scientific research, a trajectory for altering relations between disciplines and particular roles for computer science and information technology in the provision of 'infrastructure'. As an explicit statement about the functioning of scientific research the Atkins Report constitutes a substantial shift from previous models of scientific practice, organization and support. In this vision two sciences will be advanced simultaneously -- computer and domain – *and*, in addition, short term advances in these sciences will harmonize with a long-term investment serving future scientific endeavours.

Cyberinfrastructure has not emerged fully formed from the Atkins Report. Its history is linked to an emerging model of science funding, to the practical organizing of multidisciplinary collaboration and enacting an intersection of science research trajectories. It is the rise of a vision tied to future technical capacity, the maintenance of an archive and of divided disciplines coming together through these. CI has begun to open a space for itself within the institutions of science by scaling up the technical programs and vision of infrastructure and interoperability. In the case of GEON this has meant attachment to the NSF's Information Technology Research (ITR) program, an opportunity site for large-scale multidisciplinary funding.

In turn, GEON's vision has not emerged fully formed from an established model of cyberinfrastructure. Rather, it has defined itself, and been defined, as an 'experiment', a 'prototype, ''version 1,' and in its most grand moments a 'flagship' of the San Diego Supercomputer Center (SDSC). A 'formal' GEON, as represented in its two proposals (1999 and 2000) to the NSF, went through two writing iterations which drastically reshaped its membership and goals. The institutional work to achieve a funded proposal for GEON has meant enrolling NSF's Geo directorate in the development of high end information technologies, and the computer science directorate in supporting application as a site for knowledge production.

--

This chapter will follow the links back to the institutional funding and organizational arrangements which made possible the project GEON. We do not need to take these as a context-built-by-the analyst for GEON; rather, this is a context which participants build on a daily basis. The data in this chapter is collected by tracing the links from these contexts: GEON was funded under the NSF program Information Technology Research, and on a regular basis GEON was compared to other ITRs. On a regular basis GEON participants framed the NSF funding program Knowledge and Distributed Intelligence (KDI) grants as the historical predecessors to the ITRs; on a regular basis GEON participants have looked to the NSF to help fill out its formal plan for conducting research and building infrastructure for the sciences. Structures to the past, for the present and to a future are in play in the regular activity of GEON. In order to maintain an interactional and historicist understanding of 'a context' we need only follow the work of GEON participants – a network which includes NSF officers – and their continuous strategizing and reflection upon their own endeavour. It is the shorthand 'institutions' and their collation into this chapter which are the attribution of this author[3].

This chapter is divided into three sections. The first section outlines the changing vision of science funding as described in the Atkins Report and enacted through the ITR

---

[3] March and Olsen note how institutionalism shares with the 'regular organizational theory' the interest in "the beliefs, paradigms, codes, cultures, and knowledge that support rules and routines," (March and Olsen 1989, p 22) but differs in their interest by adding core ideas about contemporary public administrations: goal results, normative performance outcomes and purposefulness (DiMaggio and Powell 1991). The key determining characteristic of an institution is its tie to the state, rather than to the business sphere, and as such 'its formal framework of preferences' are at least partly 'endogenous,' or formed within political institutions. Public organizations are understood to stand in a special relationship to the people they serve, can invoke the authority of the state to enforce their decisions, and may claim legitimacy through a larger, indivisible, contribution to public interest. 'Contributions to public interest', while a typical justificatory strategy for the public administrative institution, has also been a bane to organizational theorists most interested in metric formalisms because it has been found to systematically confound attempts at quantification. (Frederickson and Smith 2003, see also Callon 1994 for the public good).

March and Olsen (March and Olsen 1989) list 7 key definitional characteristics of institutions:
i- institutions are formal bounded frameworks of rules, roles and identities
ii- within formal frameworks preferences are inconsistent and changing, and at least partly endogenous, formed within political institutions
iii- institutional theory emphasizes 'logic of appropriateness' : individual behavior within the organization is normative, and norms are historically produced
iv- logic of appropriateness is based both on norms and situational adaptation
v- for some theorists an institution must emphasize community and the common good
vi- while other theorists understand (v) as an emergent property of individual maximization (rational choice, etc…)
vii- there remains an open debate as to the explanation of 'order': a- as a product of rational action; or more relevant to this paper b- as a historical process without equilibria, taking extended periods of time, leading to non-unique equilibria, or resulting in 'unique but suboptimal outcomes'

program. Those cyberinfrastructure projects funded under ITR – such as SEEK, LEAD and GEON – have each been structured by funding arrangements, proposal and vision to satisfy the 'basic research' requirements of their home directorates, computer science, and to construct infrastructure for the domain. This is what I call the *generation of a funding push* for cyberinfrastructure. The second section tracks the organizing of multidisciplinarity at the institutional scale of the NSF. Officers at the NSF have modelled the institution and built up routines, techniques and technologies for navigating enacted disciplinary boundaries between directorates such as the Computer and Information Science and Engineering (CISE) and Geosciences (Geo). These members have collected experience in working across boundaries from previous experience with cross-directorate funding programs. In this section multidisciplinarity is cast as an actor's category and a practice, its consequences are the result of framing disciplinary differences and enacting relations across these. The third section follows the failure and then success in awarding GEON's first and second ITR proposals, respectively. The construction of a successful funding proposal repeatedly crosses any received boundary between practicing scientists and their NSF representatives. Over the two iterations of proposal writing GEON was reworked to meet requirements for earth and computer science basic research and as a viable infrastructure project for the geosciences.

2.1 – Organizing an Institution for Infrastructuring: ITR as an occasion for CI

The NSF has a mandate to support research in the sciences and engineering. The definition, organization, and means by which the NSF provides this 'support' has been the object of considerable research and debate (Nelson 1992; Slaughter 1993, see also

Callon 2004 the question of the public good). A common (arguably, founding) formulation is that the NSF supports 'basic' or 'pure' research (Bush [1945] 1960). This is high risk and expensive science that the private sphere in unlikely to assume. In turn, in this formulation applied research and development is left to industry (Jasanoff 2003; Jasanoff 2005). However, it has been argued that basic research is too narrow a category and excludes the need to formalize on-the-ground knowledge production generated through application (Kline 1995; Stokes 1997). Or, that the 'basic science' model mischaracterizes the trajectory of science, and that science-industry partnerships have always been the cornerstone of modern scientific knowledge production (Rosenberg 1994). Others have argued science itself is shifting into a new 'mode 2' operation (Gibbons, Limoges et al. 1994), requiring support for interdisciplinary collaboration, and research which is less 'basic' and more 'socially relevant' (Gibbons, Limoges et al. 1994; Elzinga and Jamison 1995; Elzinga 1997).

In all these discussions the topic remains how best can the NSF support scientific research. Some outcomes of these debates have included redefinitions of the NSF's mandate reframing "basic so that it had an applied dimension" (Slaughter 1993:290), or; inclusion of requirements in proposals for considering 'broader impacts' outside the narrow confines of the scientific field or outside science generally to reflect that "basic research has become intimately intertwined with production of goods and technological development of relevance for all realms of society" (Elzinga 1997:420). However, in all these reformulations, NSF continues to be the support for science. NSF is 'the infrastructure' for science, providing funds, venues for communication or collaboration, and sites and centers in which to do research, whether apart from or in conjunction with industry.

Cyberinfrastructure is a reconceptualization of infrastructure and the support of science. Here the NSF (or the NIH, or DOE &c.) in addition to being an infrastructure of science -- supporting science -- also comes to support *infrastructuring* for science. CI does have its buildings, such as the supercomputer centers, and it has its physical 'tubes and wires', such as the TeraGrid (Beckman 2005, see also Appendix A) and the Internet2 (Kratz, Ackerman et al. 2001). The Atkins Report maintains a continuity with concerns about science as a site of progress and as a 'public good'(Callon 1994); the preservation of open access and a site for the free exchange of knowledge (Merton 1942; Slaughter and Rhoades 1996); and redefinitions of the understanding of basic research[4]. However the difference, the crux of CI, lies not in providing infrastructure to the sciences but rather in enabling the sciences to infrastructure for themselves: "Only domain scientists and engineers can revolutionize their own fields," (Atkins 2003: 50). From the Atkins Report, rather than *science* requiring support, it is more accurate to say that *scientists* require support to build infrastructure.

The NSF, however, is understood in the Report and by participants to continue on an existing trajectory for supporting basic research. While definitions of basic research may be shifting, CI is a project for today:

> The ACP [Advanced Cyberinfrastructure Program] will be retrofitted to an NSF organization whose primary mission, the conduct of science and engineering research and education, remains unchanged. It will be important and

---

[4] "there should be no artificial distinction […] between research and development; the best enabling and application infrastructure projects, almost without exception, include both" (Atkins 2003:29)

> challenging to pursue major changes in the organization and processes underlying NSF's primary missions to promote innovative application of information technologies, while avoiding significant organizational disruptions. (Atkins 2003: 50)

Thus Atkins views the NSF as both a resource for building CI, and a site for transforming the understanding of basic research. In the case of GEON, this occurred in several iterations of pre-proposals, proposals, and the active engagement of GEON participants to define the project as twice basic research, computer and geo- science, as well as infrastructuring.

Shifting the terrain of NSF, of its individual directorates, of the divisions within the directorates, towards infrastructuring has been a fraught terrain. As an organizational transformation it has been site of constant negotiation and debate (Fligstein 2001): whether and how to push an institutionalization of infrastructuring. As we shall see, in even a few years the ‘allies of infrastructure’ (‘Cyber-’ or otherwise) have switched sides, shifted trajectories, and been shifted across directorates and offices. As one IT GEON PI noted “it used to be CISE [Computer and Information Science and Engineering] that really pushed for GEON, now we’re looking much more towards GEO for GEON’s future. They’ve become our greatest supporters!” As the savvy GEON participants well know, ‘institutionalization’ is not a passive activity to watch from a distance. They do not sit quietly awaiting the results of decisions ‘made at the NSF’ before acting; rather, it is a deeply situated and dynamic endeavour (Hallett and Ventresca 2006). The PIs are regularly engaged with the NSF by providing information, responding to questions, sending relevant articles or simply arguing their case over the phone.

This is not a ‘transformation’ of the NSF from stable organization to stable organization. Rather a formal transformation – organization chart to organization chart (Orlikowski 1996)– is coupled with multiple trajectories of change, preserved human ties, multiple interpretations of the formal transformations, and a pervasive sense of uncertainty.

–

This section follows NSF’s funding of GEON through its priority area Information Technology Research. It is a tracing of the generation of ‘push to infrastructure’ through an opportune use of the ITR funding program and a crafting of organizational boundaries and relations – a ‘collaboration’ -- between geo-scientists and computer scientists. In order to find an (initial) home for infrastructuring in the NSF GEON had to be defined as basic research (Calvert 2006) for both geo and computer science. The consequences of this dual definition have substantially shaped GEON’s goals, its methods for evaluation, and at a finer scale, its participants, technologies and physical sites.

From 2000 to 2004 the NSF awarded five rounds of funding under its Information Technology Research (ITR) program. The solicitations explicitly called for proposals with “applications and effects throughout the sciences, engineering, education, the economy, the humanities, and society in general” (NSF 2001). When the program is complete it will have disbursed approximately one billion dollars in funding, with projects ranging from three year ‘small ITRs’ of $500K to five year ‘large ITRs’ with cumulative totals up to $15M. GEON submitted two ITR proposals, the first was not funded. The first proposal was submitted during ITRs second round during fiscal year (FY) 2001, and the proposal that came to be funded was submitted for FY2002.

The term 'infrastructure' appears in Atkins in various forms, but ITR was not a cyberinfrastructure program; the latter term does not appear in any of its five solicitations. Out of the 453 total medium and large ITRs only a select group explicitly used the term infrastructure in their proposal title (18), only a handful used the term cyberinfrastructure in their abstracts (5), and only one in its title: "GEON: A Research Project to Create Cyberinfrastructure for the Geosciences"[5]. Some of these projects have later come to call themselves CI, and many more have been labelled as such by others. The ITRs are characterized as cross-disciplinary collaborations researching computing science and/or applications, and many of these were multi-institutional and physically distributed projects (Olson and Olson 2000; Teasley and Wolensky 2001; Finholt 2003; Finholt 2004; Baker, Jackson et al. 2005; Lawrence 2006 (forthcoming); Ribes forthcoming). Although the ITR program was not 'for CI', it was the largest funding program for IT running at the time of the Atkins Report, and thus became an occasion for CI. ITR, and in particular the large ITRs, became an opportunity for enacting the Atkins vision of multidisciplinary collaborations and *generating a push to infrastructuring*. In the case of GEON, work on CI came to be defined both as a domain and computer science contributions, *and* a long term investment in the form of community infrastructure.

The Atkins Report has come to be much more, but it is important to keep in mind that initially it was commissioned as a review of the Partnerships for Advanced

---

[5] This is out of the total 453 large and medium ITRs.

Computational Infrastructure (PACI) program. PACI is the NSF program which included the two 'leading edge' supercomputer centers (SDSC and the National Center for Supercomputing Applications - NCSA) and was organizationally housed within Computer and Information Science Engineering (CISE) directorate of the NSF. CI is envisioned as multi-institutional and multidisciplinary, but CISE was the original 'home' for cyberinfrastructure[6]. Similarly the ITR program stretched across all the NSF directorates but CISE was selected as the lead. Their officers chaired the meeting to organize planning of solicitations, review, awarding and evaluation. As a CISE NSF officer describes the convergence of the two programs:

> Remember that ITR has this Coordinating Group that has representatives from each of the Directorates. So these Directorate representatives are also learning about Cyberinfrastructure and what it might do for their communities or researchers. They're saying hey, we've already got this ITR program so we should be funding some of this stuff now. […] So it's some kind of convergence of ideas. Ideas circulate pretty well within NSF, so if there are new ideas that people agree upon…I think cyberinfrastructure was one everyone agreed upon but just doesn't know how to do, quite. (NSF CISE)

ITR was 'crosscutting,' which means it included all seven NSF directorates for the sciences and engineering, and the three offices. The 'Coordinating Group' mentioned in the excerpt was composed of chairs from CISE and representatives from each the seven NSF directorates and three offices. Held on an almost weekly basis, these meetings came to be an ideal forum for sharing informatic developments across the NSF. Each program referenced the other. The Atkins Report itself was published in the fourth year of ITR,

---

[6] The term 'cyberinfrastructure' is most commonly traced back to former CISE assistant director Ruzena Bajcsy (1998 – 2001).

just as 'cyberinfrastructure' was becoming a buzz word and, in turn, the Report favourably reviews the ITR funding model[7]. Within the cross-organizational forums of ITR a funding line came together with a vision to infrastructure.

--

ITR became an NSF 'priority area' following the influential President's Information Technology Advisory Committee (PITAC) Report of 1999[8]. The committee was established by President Clinton in 1997 in order to provide "advice and information on high-performance computing and communications, information technology, and the Next Generation Internet." The committee reported its findings in 1999, claiming (i) that federal funding for IT research was either flat or had been declining for over a decade, and; (ii) that existing support was focused on narrow short-term goals and immediate application. It recommended the creation of a "strategic initiative in long-term information technology" which would encourage research of a "visionary and high-risk" nature through diversified funding mechanisms and with multi-year support[9]. Organizationally it recommended the development of new management and implementation strategies for IT research and development.

The PITAC Report led to several responses in the US government. President Clinton and vice-president Gore proposed an initiative for Information Technology for

---

[7] In speaking of the 'fluidity of idea exchange within the NSF' in the excerpt above, the officer is describing something akin to *isomorphism*: a convergence of organizational goals and means through communication within an institutional field (DiMaggio and Powell 1983; Frederickson and Smith 2003). However in a situated analysis the methods to achieve ITR as the site for CI require a finer granularity than are encouraged by the analytic tools of neo-institutionalism.

[8] This lineage is relatively uncontested. It is cited by NSF officers, on the NSF webpage (accessed July 2006 http://www.nsf.gov/news/news_summ.jsp?cntn_id=100686) , and 'in-action' by GEON participants to defend the experimental nature of the project, see section 3.

[9] Within social informatics circles PITAC is also often cited as 'the first' major report to point to sluggish support for social scientific research on computing and information technologies, workforce and education. The report calls for increasing funding for "socio-economic impacts" and "public policy issues".

the Twenty-first Century (IT^2)(Service and Malakoff 1999), Congressman Sensenbrenner introduced the Networking & Information Technology Research & Development Act (NITR&D) which passed the house but not the Senate (Bromley 1999), and eventually Congress passed a federal budget that included an NSF FY00 appropriation increase of $126M for 'ITR'.[10]

In its first year the calls for funding were divided between 'small' and 'large' proposals[11]. In addition to the usual NSF review criteria the ITR program also noted that proposals would be evaluated by the degree of planned integration with existing research and education programs, and by "Innovation in Information Technology and its Application to Science or Engineering". Research would be "novel, high-risk, and high-impact research,' with an emphasis on innovation rather than an "incremental improvement on standard ideas." (NSF 2001). In the year GEON was awarded an ITR,

---

[10] I have not been able to follow the details of this appropriation. Interviews and discussions with NSF officers indicate that the NSF may not have been the only considered site for the ITR grants and that another possible contender or collaborator could have been DARPA (Defense and Advanced Research Projects Agency) which has a strong track record in IT research (Roland and Shiman 2002). One interviewee from CISE noted that in the early periods (~1999) ITR funds could have been allocated to DARPA but NSF was chosen because of its mission and review process. The extended extract below reflects the formal attitudes and definitions of the NSF towards basic research generally, and in the ITR program:

> NSF is basic research, very far downstream, really doing stuff that might fail, funding things that we don't know where they're going to go necessarily, whereas DARPA really has a mission that it has to accomplish out and about in the field. [...] their peer-review process is very different at DARPA. I don't what people would say about its "ethical-ness", but let's just say that NSF is the gold standard in the merit-review process. And that is because of all these things I was telling you to begin with about the confidentiality of our reviewers […] We also handle conflicts of interest. No one can review a proposal who has a conflict. The most serious ones are fiduciary conflicts, but also if you are a collaborator with someone who's on it, then you cannot be a reviewer; you cannot be part of the recommending of awards or declines if you have a conflict. […] It's all documented and we have these systems by which we do this – get the reviews in, the reviews go back to the PI as anonymous reviews without the names and so forth. At any rate, a lot of people feel our process is much more conservative than the DARPA […] (NSF CISE Interview)

[11] Program Announcement NSF 99-167: small-ITR: less than $500K for 3 year total, and large-ITR: up to $3M a year, 5 years total. Following this, year two through four of ITR divided grants into small, medium (group), and large, and in year five only included the categories medium and large.

the proposal also included an emphasis on activities that would "enable research and education in multidisciplinary areas, focusing on emerging opportunities at the interfaces between information technology and other disciplines" (NSF 2001).

Although ITR was crosscutting there was no requirement built in to the program for cross-directorate collaboration. A directorate could chose to spend funds internally on information technology projects with a narrower disciplinary focus. For example one CISE NSF officer summarizes the case with the Social, Behavioural and Economic Sciences Directorate (SBE) which had the smallest apportionment of ITR funds:

> Other Directorates like SBE […] looks at its budget quite differently because they don't have much money for their regular programs –they're just trying to fund the sociologists that are tenured and full professors just to get a little bit of money. So they really don't want to play in these big things because they don't see the value of it for social scientists. Nonetheless they still put money into it and they were very good at being able to leverage the money […] So you can imagine people in SBE saying, "well, we've only got $3 million in the game, so we want to spend all of our $3 million on 10 small awards that are just sociologists, or geographers, or whatever..." We don't want to play in these other things over here [large, cross-directorate ITRs] because our tiny amount of money will get lost.

As we will see, collaboration with, for example, CISE, could mean a substantial increase in the pot of money available for an ITR project, however, securing collaboration also meant redefining research goals. Funding a project within a directorate would still require 'high-risk IT research', as per the solicitation, but this allowed for a narrower definition of disciplinarily informed research interests.

Instead of through formal proposal requirements, collaboration was encouraged by the distribution of funds in the NSF. Directorates with smaller portions of ITR funds could collaborate with another directorate to maximize funds. ITR funds were variously distributed across the seven directorates and three offices as determined by the director of the NSF, each directorate receiving an apportioned sum. In turn each assistant director[12] would have some discretion as to how to distribute these amongst the various divisions in each directorate. Because funds were not given a fixed allocation across divisions, this led to some competition *within* the directorates as well as across.

For example, during the GEON proposal's second year review process, it was reported to have come up against another large-ITR project within the GEO directorate: Linked Environments for Atmospheric Discovery (LEAD). GEON is funded by the Earth Sciences division (EAR) within the Geoscience directorate (GEO), while the hurricane prediction project LEAD is funded by the Atmospheric science division (ATM) under GEO. Both LEAD and GEON were described as receiving very favourable reviews and in the panels securing support from the proposed co-funding directorate, CISE. However, given the GEO ITR budget, and the distribution of finances within the directorate, it was clear that only one project could be funded. A GEO/EAR officer describes the tension within a directorate and across its divisions in regards to the ITR funds:

---

[12] The NSF is formally hierarchically structured as follows: A director for the entire NSF; followed by a division into seven directorates (and three offices) that are each headed by an assistant director (AD); each directorate is further split into divisions headed by a division director (DD); which is in turn divided into programs headed by a program director (PD – these are often also called program officers). For example, in 2001 the AD for the GEO directorate was Margaret Leinen, the DD for earth sciences was Herman Zimmerman, and the PD for GEON was Leonard Johnson. I will use the generic term NSF officer to refer to a member of the NSF, regardless of position, and cite these along with their directorate. When division or program is relevant, this information will also be provided.

> Had we not gotten GEON, we may not have gotten as much of that money. This is how the funds are distributed: the GEO Directorate gets that money and then Margaret Leinen [GEO AD 2000-present] makes the decision. Earth Sciences will get this amount and Oceanography and atmospheric [ATM] will get that amount. If we didn't have GEON, and ATM had their big project [LEAD], they would have gotten a bigger percentage of the ITR money. Although it's "our" money [GEO], GEON helps *us* [EAR] out in some ways. (NSF GEO / EAR Officer – emphasis in original, quote marks made as finger gestures)

This officer is neatly describing the generation of a funding 'push'. With curling fingers around his head he described GEO directorate resources as 'our' money, and then with a more careful specification indicated that GEON helped his division, EAR, secure more of it from the shared base across three divisions. Without the GEON large-ITR ATM would secure a larger portion of the available pie. Even with multiple medium or dozens of small ITRs, EAR would be hard pressed to secure both directorate support and collaboration with CISE; in the case of GEON ~$6M is at stake within the directorate alone and double that if you include the entire project budget[13].

To be clear, this is not 'all about money'. It is, however, one consideration. Mobilizing the geoscience community to infrastructure is also bringing to bear i) extant geoscience 'community needs' of the ii) heterogeneous nature of geoscience data and practice, and iii) endogenous comparisons within geoscience and with other sciences. In other words earth sciences comes to define a set of IT needs within the community, based on a characterization of their data as diverse and heterogeneous, and through comparison

[13] LEAD was funded the following year. The project has its technical core at NCSA, the 'other' leading edge supercomputer center within the PACI program (although it does have a larger team of distributed computer scientists) (Lawrence 2006 (forthcoming)).

with other sciences such as atmospheric (see Ch. 3). For example, another GEO/EAR officer describes his methods, and a panel reviewer's argument, in pushing for GEON over LEAD:

> The point that I tried to make through this whole discussion period was if [LEAD] were funded…it's certainly a worthy project and worthy of funding. But if it were funded it would be an *incremental* addition to the atmospheric science community because they already had NCAR [National Center for Atmospheric Research] and that tremendous computing capability. And earth sciences had nothing of that scale. So my argument was, and as you read in the memo, the argument of one of the panel members: funding [GEON] would give a huge step up to the earth science community and you'd get more bang for the buck if you did that. (GEO/EAR Officer)

The argument relies of a characterization of the atmospheric sciences as *already* possessing an organizational/technical platform for data integration endowed with substantial computing resources; it also relies on an endogenous comparison between atmospheric and earth science, and a conclusion that an integration project is most needed in geoscience by virtue of organizational and technical lag. We will come to see the characterization of the geosciences as informationally deficient as a common strategy for generating a 'pull' to cyberinfrastructure in the earth sciences (Ch. 3).

With the exception of CISE the uneven division of ITR money mirrors the overall distribution of funds across the NSF: SBE received the least, and the Directorate for Mathematical and Physical Sciences (MPS) the most. The exception of CISE is crucial as it provides the fulcrum to encourage cross directorate collaborations. Approximately 50% of the annual ITR budget was in the hands of the CISE directorate, and 25% with MPS. Thus, for example, as with the three large ITRs (and CI projects) LEAD, GEON and

SCEC, the GEO directorate could fund significantly larger proposals in collaboration with CISE than it could possibly do on its own[14].

A co-funded large ITR with CISE, while requiring the greatest investment from GEO also provides the greatest return; in the case of GEON this meant several millions dollars from outside the geoscience directorate. However, in turn, a collaboration with CISE also meant a greater involvement with computer science research goals. The modes for these collaborations came to include the SDSC, cyberinfrastructure and new models of IT/domain collaboration informed by social science research.

The meeting points of ITR, CI and the SDSC

Before GEON, the first formulations of earth scientists were in term of 'a database' for geoscience. This would be a clearinghouse for all publicly funded data. However, in order to enrol the interests of the computer science directorate the GEON proposal was written as more than 'a database for geoscience'. Such a database was not considered in line with ITR's call for the cutting edge and high-risk research (see Ch. 3 for some of the technical formulations which anticipated GEON previous to engagement with the SDSC). Three strategies were employed in making themselves an interesting research project for CISE: the GEON proposal included prominent computer scientists with explicitly stated cutting edge research goals; the overall goals of GEON were framed in terms very similar to those of CI; and GEON would have its technical core at CISE's 'leading edge' site for the PACI program, the SDSC. All three of these strategies

[14] Formally , funds remain listed in their respective directorates regardless of cost-sharing. GEON does not increase the GEO budget 'on paper'.

intersect at the SDSC, which is the nominal subject of the Atkins Report and which employs high profile CS researchers working on the application of ‘leading edge’ IT.

--

The SDSC should not, a priori, be considered ‘the natural’ fit for a project such as GEON. This too is an accomplishment. At all scales of GEON’s action the occasional doubt is still expressed. At the time of the GEON proposal the SDSC was the PACI ‘leading edge site’ mandated to remain ahead of the computational curve by one or two orders of magnitude over the regular computing resources available to universities. More recently the SDSC has come to be a research institution increasingly defining itself in terms of the application of cutting edge information technologies to the ‘sciences and engineering’.

As researchers SDSC computer scientists publish a steady stream of scholarly articles, participate in national and international conferences, and are often well known and respected in their fields for their invention and innovation. In short, the SDSC is not primarily populated with ‘technicians,’ web designers, information managers, or support staff—and certainly these describe not a one of GEON’s IT PIs. And yet for geo-scientists GEON is meant to be a stable platform in the provision of data resources for conducting everyday science.

There is substantial ambivalence amongst earth scientists as to whether the SDSC is up to the task of developing IT for everyday scientific research. Even four years later, in 2006, well into the GEON project, concerns about the practicality of the SDSC remain, as one GEO officer asked:

> Why would an IT, a guy or woman with a PhD in information technology or computer science, simply want to do just what the paleontologist wants? So that's the barrier: to make it interesting. I think for them [the SDSC] GEON is interesting because there are these grid-enabling things to happen. But once that's all set up and working, it's unclear to me whether they [the SDSC] will become bored. (GEO Officer)[15]

The work of implementation, of 'working the bugs out' and of the highly detailed work of fine tuning stable and usable human interfaces is not usually considered research. It is, however, acknowledged by all to be extremely detailed and involved (Jirotka and Goguen 1994; Star and Ruhleder. 1994; Trigg 1994; Mackay, Carne et al. 2000; Oudshoorn and Pinch 2003). The problem of a development gap between high-end computer science application and its everyday usage is an endogenous concern for GEON participants. In fact, it would not be an exaggeration to say that it is one of the central organizational problems that GEON has framed for itself, and it has done so in the familiar terms of disciplinary difference as the problem, and infrastructure production as a vehicle for its resolution (See 'the two-tier approach' Ch. 4.3).

The later work amongst GEON participants, though, does not explain how an organization such as the SDSC came to be part of the project. Rather, the SDSC was part of the early strategy to include more computer science research elements in order to enrol CISE. As we will see in Ch. 3, it is during the 2000 meetings that the SDSC first came to participate in Geoinformatics, and then become the leading institution for the GEON grant. More abstractly though, the SDSC and its sister center NCSA, both had particular roles outlined within the Atkins vision of CI.

---

[15] For grid computing see Appendix A and Ch. 4.1.

The problem of a development gap between CS and useful geoscience computational tools is endogenous to CI. However the Atkins report parses the matter quite differently than the geoscience officer in the extract above. The Report concurs that computer science is interested in cutting edge research, but to this problem it adds that domain scientists do not have a sufficient interest in long-term technological thinking. Rather than CS as the sole contributor to the problem, both domain and IT are seen as short-sighted in different respects. Thus Atkins frames the development gap problem in terms of a balance:

> Taken together, these two issues present a serious challenge to any organizational structure. If the organization is weighted too heavily toward the domain scientists, the focus overemphasizes procurement of existing technologies, and computer scientists become viewed as "merely" consultants and implementers. If the weight shifts too heavily toward computer science, the needs of end users may not be sufficiently addressed, or effort shifts too heavily toward creating new technologies with insufficient attention to stability and user support. (50-1)

This dual problem, as a matter of balance, can be resolved by carefully arranging points of intersection between the two groups and this is to be achieved via organizational means. In counterpoint to the GEO officer above, the Atkins Report does not frame the SDSC along with computer science. Computer science is theory, and only the most rudimentary of programming for proof-of-concept[16]. Meanwhile the SDSC is a center, populated with computer scientists, but also a support staff, and with almost two decades

[16] As Lawrence notes proof of concept is "sometimes unstable of highly customized software which often required the people who created it to run it" " (Lawrence 2006 (forthcoming):10). See also Ch. 4.2.

of experience working in close partnerships with domain science. From the Atkins Report it is the site of application rather than basic research narrowly conceived:

> In particular, some PACI-enabled collaborations among domain scientists and computer scientists have been exemplars of interdisciplinary interactions in which information technology becomes a creative, close partner with science (62)

Contrary to a view from geoscience, from within CS the supercomputer centers are often seen as close to the domains, as sites of application research, and as points of collaboration *amongst* diverse computer scientists. They are the ideal sites for enacting the vision of CI.

CI is not the first attempt to systematically introduce information technologies to the sciences – rather it sits as the newest program in a long line of models, the most recent of which include the Supercomputer Centers Program, the Science and Technology Centers Program and Digital Libraries, to mention only a few (Hughes 1999; Borgman 2000)[17]. The Atkins report collects scattered insights from state and scholarly reviews of these programs, as well as from computer science, sociological research in informatics and the personal expertise of technologists that have designed large-scale information systems with both successful and fatally flawed results. The recent past of computer science is dotted with a landscape of electric ghost-towns, acclaimed but then abandoned high-end computing project and applications. PITAC, ITR, and the Atkins Report reflect facets of these understandings. In particular Atkins tackles the challenge

[17] See also Weedman (1998) for another formulation of balance in IT/domain collaborations during the early 90s. Weedman's focus is on the practice of multidisciplinary collaborations established by such programs.

by encouraging new management structures and implementation strategies in collaboration with social scientists:

> Much of the effort under way to use cyberinfrastructure for collaborative research is not giving adequate attention to sociological and cultural barriers to technology adoption that may cause failure, even after large investments. (Atkins 2003:13)

Those 'social' barriers are to be understood as mismatches between technical development and existing routines and conventions within the community. Thus, although the overall goal-set of CI is the development of a unified informational infrastructure, the possible modes of its implementation are substantially varied even in the technically sparse Atkins Report. It includes some possibilities for centralized management, federated databases, and standardized development, but also for local control, distributed collections, and for relying upon tools for integrating heterogeneous development trajectories.

Rather than 'top-down,' Atkins calls for a multiplicity of technical approaches each of which would have to carefully tended for integration. The approaches should be specific to community needs, as determined by social science studies, but most importantly, as is understood by practitioners themselves. The matter of technical trajectory is to be determined at the site of application, and relative to existing technology and community configurations:

> The conduct of science and engineering research is built (in part) on these applications, which are tailored to the specific needs of people, groups, organizations, and communities conducting that research. […] Some applications are generic (such as distributed collaboration),

> and many others are discipline specific (like distributed community access to a specific scientific instrument). (Atkins 2003:48)

The Report dedicates considerable space to distancing the approach of CI from top-down design[18].

The meaning of 'top-down' appears in common parlance and published sources within informatics circles as decision-making by committee or decree (such as from the NSF), and the imposition of standards or other forms of centralized uses of authority. These mismatches of ordained technology and existing community need are often understood to be at fault in failed large-scale projects. 'Top-down' is defined as the initiation of projects by CI centers rather than by practicing scientists themselves:

> this is the belief that disciplinary experts, in close partnership with computer scientists, are best able to judge the merits, impact, and importance of applications and specialized cyberinfrastructure focused on their field, and that these projects should be peer reviewed rather than initiated by the centers.(Atkins 2003: 64)

Eschewing top-down design, the ethos of the ACP is to fund new projects initiated by scientists themselves, to promote collaborations and to harness the multiplicity of already existing informatics projects. This approach is often dubbed 'bottom-up' and suggests projects which are initiated from within the domain, are led by scientists, and have a community endorsement. More than an ideal of bottom-up there is an explicit suggestion of mechanism through peer review, the NSF's primary funding model. The final

[18] "these new projects are building out in terms of broader scientific application, and they are building up in terms of function and performance." (Atkins 2003:7)

publication of Atkins was during ITR's third solicitation the Report looks upon its funding model quite favourably.

As we have seen ITR encouraged a kind of organization, and model of collaboration, which neatly mirrors the Reports recommendations. Atkins outlines a model for NSF shaping the domain fields through funding and priority definition. It also sets the limits for the interventions and the place where domain activity should begin. The model is simultaneously 'top-down' and 'bottom-up':

> Two complementary activities are to be organized. The first is programs within NSF, which prescribe how resources are allocated to the various activities, evaluate proposals and make awards, and assess outcomes. […] The second involves the science and engineering community itself – the researchers, developers, and operational organizations that carry out the missions defined in the ACP. NSF can have significant influence on the organization of the community through setting priorities, defining programs, establishing evaluation criteria for proposals, and then evaluating proposals."

ITR became an opportune occasion for just such an endeavour. To review the ITR model thus far (see also Fig.6): it includes a 'top down' funding solicitation crafted at the level of the NSF, with input from all disciplinary directorates. All five solicitations encouraged cutting edge, high risk and long term (five year) research. The distribution of funds amongst the directorates created a substantial incentive to encourage proposals that are i) large and ambitious and ii) in co-funded collaboration with CISE. This encouragement would give CISE the needed leverage to enact the visions of CI: interoperability, contemporary database technologies, and a technical core at a CI center such as the PACIs. As a review of the PACI program it not difficult to impute the Report's suggestion of a model for multidisciplinary collaboration, technology development and basic research as applying to the Supercomputer Centers:

> There is no intention that these activities be strongly separated; development, generic, and disciplinary activities may be co-located or even grouped within common centers. One appealing organizational model, for example, is a development or generic center that maintains and integrates a collection of disciplinary group. (57)

Most importantly the actual proposals themselves would be solicited. That is, they would be generated from a 'bottom-up' internal momentum in the domain community, with domain PIs, and support from the respective directorate. Research would be framed in terms of domain contributions: 'what will this information infrastructure do for us?' Infrastructuring would be framed as basic research for both domain and CS. Lastly, peer review would ensure that a heterogeneous group of computer scientists and domain scientists would evaluate the proposals relative to CS and domain research.

--

In this section we have seen the generation of a funding push. Most broadly this is simply the ITR program and its substantial funding allocation to the directorates. At a finer granularity we have the dynamics within and across directorates. Within the directorates a light competitive environment amongst divisions coupled with a genuine desire to develop community resources generate a more subtle push to produce viable yet ambitious ITR proposals such as GEON. The allocation of 50% of the ITR budget to CISE itself generated a push to cross-directorate collaboration, with clearest funding benefits emerging from the large ITRs.

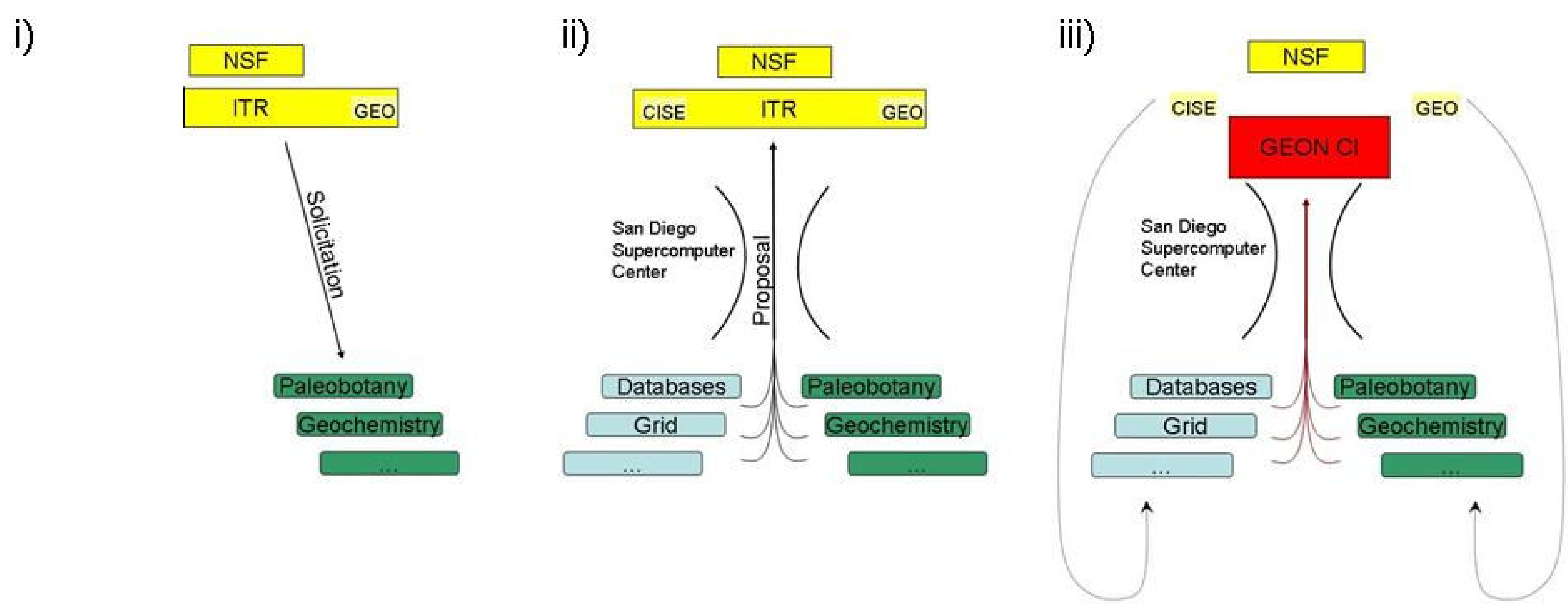

Figure 6: The generation of a 'push' to CI through the ITR funding model i) the request for proposals (RFP) to domain communities, generating a 'funding push' (Ch. 2.1); ii) through a careful 'practice of multidisciplinarity' proposals are co-authored with CS, thus bringing in CISE and the SDSC (Ch. 2.2); iii) proposal is peer reviewed, and evaluated at the NSF; permitting further reframing as 'basic' domain science (Ch. 2.3). The awarded funding enables 'bottom up' work for domain and CS. At each stage a new actor is enrolled – geoscience, computer science/SDSC, 'geoscience community' respectively. In the final iteration something very much like the Atkins vision of CI is enabled.

Collaboration with CISE, however, required framing the ITR in terms of CS advances as well as geoscience contributions. This collaboration at the large ITRs provided a novel window in the NSF to fund infrastructure projects. Thus ITR became an occasion for cyberinfrastructure: an ambitious bottom-up, geoscience led, CS partnered, NSF funded, community peer-reviewed and endorsed, project.

However, the push to CI is not exclusively a matter of funding. CI is not simply ‘add funds, IT and stir’. Even before funding GEON, negotiating a multidisciplinary collaboration – whether framed as a cross-geoscience or across geo and IT – was also a substantive organizational undertaking.

2.2 – Organizing an Institution for ‘Multidisciplinarity’

Across GEON’s scales, whether we look at the institutional, organizational, or technical dimensions, there are two primary axes of multidisciplinarity which we see again and again. The specific configuration of these axes is locally enacted but the rough outline is as follows: i) the geosciences are internally disciplinarily diverse as is to be seen in their methods, language, culture, data structures and knowledge itself; ii) computer science (or IT) *and* the geosciences are disciplinarily diverse from each other as is to be seen in primarily in their topical/technical interests, in their goals and reward structures, but also in their language, culture, and methods. These two axes follow the summary diagram from The introduction reproduced in figure 2. Boundaries of difference are constituted across both axes, domain/domain and IT/domain. These boundaries are problematized relative to goals of infrastructure, of collaboration and of data interoperability. This problematization becomes the platform for the active organization of multidisciplinary efforts.

Figure 7: Relations of difference in the logic of interoperability (see p.28)

This section has two objectives: i) to come to see 'multidisciplinarity' as a practical achievement at the institutional scale, and; ii) make visible the practices and organizing which are informed by 'multidisciplinarity'. Importantly, this is not a '*study of*' multidisciplinarity. Nor is it a diatribe against the buzz of post-normal or Mode 2 science. This is a study of enacting GEON at three scales. At each of these scales 'multi-' 'inter-' 'cross-' and 'trans-' 'disciplinary' is a an actor's category. On a regular basis its meaning is generated and re-generated locally. It is a key-term, an organizing principle, and a goal. This study takes the activities informed by 'multidisciplinarity' as object. I argue that multidisciplinarity is a term, concept and model *along which* CI is built in the double-forked action of boundary work and relations.

--

In explicit opposition to the hype surrounding 'Mode 2' science (Gibbons, Limoges et al. 1994) Weingart has argued that the rhetoric of multidisciplinarity is a standing discourse "proclaimed, demanded, hailed, and written into funding programs" for decades. However, parallel to this 'flag waving' he notes that the practice of research itself is better characterized by an increasing specialization (Weingart 1997). Mode 2 science its characterized 'transdisciplinarity' are new wrappers for old ideas, and, as Weingart fervidly asserts, are closer to ideologies than descriptors for the state of today's science.

I do not wish to contest his claim – this is distinctly diverging from the goals of this study. Rather, I argue that to the extent that Weingart is correct in claiming that today's language of multidisciplinarity is "old wine in new bottles" we appear to have failed to understand the composition of that old wine. By black-boxing the mechanisms of funding, dismissing it as a rhetoric, or, worse, as political legitimation, Weingart frames his research questions in the binaries of truth: 'are we in Mode 2?' 'Is science now transdisciplinary?' (see also, Hicks and Katz 1996 for a more moderate 'test' of the Mode 2 hypothesis). These questions leave no room to take into account the outcomes of 'multidisciplinarity' as an enactment beyond rhetoric and ideology. In short, whether successful in the long-term or not, or whether it is *really* multidisciplinary or just 'formulated for purposes of political legitimating [to] camouflage normal disciplinary research," (Weingart 1997:598) here I argue that within the NSF a great deal more than flag waving is occurring to make multidisciplinarity happen 'on the ground'. Rather than taking avid positions on the ideology and truth-state of multidisciplinarity we must understand its usage and consequences in action.

In the broadest brush strokes, I conclude that multidisciplinarity comes to be boundary work between the disciplines and then the formation of relations across these. More specifically, in the case of ITR multidisciplinarity is framed by actors as acquired logistical and practical skills for effectively negotiating existing disciplinary difference. In funding GEON this is a framing of infrastructure building as a balance between basic research in computer and geoscience. The specificity of practice and organizing follows below.

Multidisciplinarity as Organizational Skill and Learning

The ITR program itself can be considered no small organizational achievement. The NSF is built on a base of proposal solicitation, peer review, committee and program officer review and then awarding. ITRs did follow this model. However the scale of the program and the range of research goals represented a set of marked changes for many of the directorates, changes which required practical management as an organization.

First, ITR had to span across the disciplinarily linked directorates of the NSF. ITR was a cross directorate program in the NSF, including all seven of its domain divisions. Second, the financial size of ITR as a whole: when the program is completed it will have disbursed approximately two billion dollars for research funding[19]; its individual apportions ranged from $150,000 to $15, 000,000. Third, is the sheer response in numbers to the solicitations, over 2000 in its first year. Fourth, is finding reviewers for all these multidisciplinary proposals, substantially qualified in a particular field yet significantly comfortable with the technical expertise of others. To complicate this, 'disinterested' reviewers had to be found – that is, a reviewer could not themselves have submitted a proposal to ITR[20]. These reviews come to be significant resources for the panels – they inform but do not determine the outcome. Assembled by expertise to assess several proposals simultaneously the panels are sites of detailed review and agonistic debate. It is here that directorates begin making final allegiances to support or co-support particular projects. With the financial and temporal commitment of the large- ITRs,

---

[19] This is distributed over 9-10 years. Portions of these funds were allocated outside the ITR funding line, for example in 2000,126M was in the budget, 36M was allocated to the TeraGrid Program (Beckman 2005)and 90M to ITR.

[20] At 2000 proposals, with three reviewers per proposal, this means thousands of scientists were immediately unavailable. NOTE: in any given year, the proposals were divided into large, small and medium, an individual could, for example, submit to large, and review for small.

program, division and directing officers are often involved in the final decisions. Finally, the awarding of the proposals themselves, that is, the awarding of funds following iterative reviews, committee debate and final decisions by upper level officers: ITR program officers describe these formal steps as only partly representative of the immense logistical work in the solicitation, distribution, organizing of actual proposals, pushing reviewers to respond within specified time-frames and in creating lines of communication across directorates.

--

Actors involved in ITR describe the *organizational learning* within the NSF as members attempted to encourage and then accept funding proposals made to stretch across directorate boundaries. Broadly sketched, organizational learning describes actors as they work to "adapt to changing environments, draw lessons from past successes and failures, detect and correct the errors of the past, anticipate and respond to impending threats, conduct experiments, engage in continuing innovation, build and realize images of a desirable future," (Argyris and Schon 1996: xvii). Organizing is seen as practice, as embodied in human skills and the enactment of formal structures through action[21]. Ethnomethodological studies of work, informed by science studies, has in addition come to be "distinguished by the way in which they topicalize embodied practices with 'paper,' 'pages,' 'blackboards,' and 'typewriter symbols,'" (Lynch 1985:294). The stability 'across' organizing activities is to be found in material arrangements (Latour 1991), such

[21] Note: In Ch. 3 I focus on the simultaneous *enactment* (Weick 1977) or framing of an 'environment' and organizational response through informational, material and practical resources. In this section I focus analytically on the learning of practical skills for multidisciplinarity. This said, a great deal of enacting the environment can also be seen e.g. in framing the growth of KDI to ITR, in conducting a social science study of KDI, and in applying its findings to ITR.

as forms and files, and in summative declarative statements such as reports or even memos (Boden 1994; Taylor and Cooren 1997; Taylor and Van Every 2000). Thus, from organizational learning we must consider even a model of 'organizational change' as too static, too distanced from practical work, talk and writing (c.f. earlier models of organizational learning Argyris and Schon 1978). In those formulations organizational change was taken to be 'a problem' to be explained. In situated/practical/action oriented perspectives organizing is an everyday accomplishment. The only continuity is to be found relative to the work of (re)creation and material arrangement. 'Persistence' is a matter of actor's accounts, and so too is organizational learning. Thus rather than sketching the 'objective lineages' of ITR, I trace *actor's lineages*: accounts of organizational learning and practical application of this learning.

Participants in ITR speak of two skills sets: i) a logistical understanding and the creation of administrative support for a large multi-directorate funding program, and ii) a practical organizational skill to foster multi directorate collaborations in the NSF. The ITRs did not emerge blank slate as a cross-directorate funding line. We have already mentioned some of the Foundations' previous cross cutting programs, such as the Supercomputer Centers, PACI, and Digital Libraries. From these programs a degree of organizational learning was acquired both in the reproduction of logistical forms and in personal practical experience.

Actors have traced ITR itself to various lineages. Here I will focus on the most commonly drawn and temporally immediate lineage: the Knowledge and Distributed Intelligence (KDI) Program. Many of those researchers who received KDI grants later worked on ITRs, similarly many of the CISE NSF officers had direct experience on both research programs. Direct experience and learning from KDI program could be as an NSF officer or as an awardee, but there is also a legacy of for program which has come to be recorded in the organizational memory of the NSF through more formally commissioned social studies. I treat direct experience and then formal encoding in turn.

--

Much smaller in scale, and primarily centered around CISE, KDI is sometimes considered “the beginning of this agency’s attempts to support interdisciplinary research on a grander scale.” (Cummings and Kiesler 2005:15). The competitions for funding were conducted in the two years previous to the beginning of the ITRs in 1998 and 1999 awarding a total of seventy-one grants, averaging three years each, with an average of $1.5M per annum. The solicitations were phrased to encourage cross-disciplinary and cross-institutional collaborations which would focus on advances in computer science and its applications:

> To achieve the aims of KDI, proposals are solicited from individuals or groups for research that is inherently multidisciplinary or that, while lying within a single discipline, has clear impact on at least one other discipline. (KDI Solicitation 1998)

Already in the KDI solicitation we find the rhetoric of a revolution for the sciences and engineering, and already computation and interoperability is tied to these goals (here referred to as Knowledge Networking or KN):

> The recent growth in computer power and connectivity has changed the face of science and engineering. The future promises continued acceleration of these changes. The challenge today is to build upon the fruits of this revolution.

> KN will focus on attaining new levels of knowledge integration, information flow, and interactivity among people, organizations, and communities. (KDI Solicitation 1998)

There is some emphasis on 'infrastructure' within the KDI solicitation, however the framing of this is much narrower than what would emerge later in the ITRs and then with the language of cyberinfrastructure[22]. In Richard Zare's 1997 editorial to *Science* magazine, announcing the KDI program, he links this upcoming line of funding to President Clinton's plans for 'Next Generation Internet' and the nation's competitive science infrastructure (Zare 1997)[23].

In an interview, a CISE NSF officer noted how the NSF had had little experience with institution-wide funding programs, and that the ITRs had been significantly shaped by the experience of the KDIs:

> (I): the first year of ITR, we did have some lessons learned from what we had done with KDI. […] we had some experiences with foundation-wide interdisciplinary programs. We had learned from KDI a lot about *what not to do*.
>
> (DR): Such as…
>
> (I): […] the program director who ran KDI […] he had his regular program assignment *and* had to run a priority area: KDI […] So everybody gets parceled out their programs and then they ["NSF"] said "and now thou shalt also run a Foundation-wide thing". And I guess that drove everyone nuts. How did they do the money? At KDI, I

[22] There was some overlap in awardees of KDIs and the ITRs, and most relevant to this study the SDSC participated in two KDIs and specific individuals were common across these projects and the GEON ITR. The EAR division and GEO directorate, and 'geo-scientists' did not participate in KDI, although several KDI projects did focus on the ocean or environmental sciences. One KDI did involve geo-physicists, however the project was framed primarily in terms of physics rather than geology.

[23] See also Metzger and Zare (1999) for a programmatic call to interdisciplinarity research. These pieces also note some of the structural barriers to these forms of collaboration.

> guess the money was in some central pot or something. And there was this terrible bickering about how to spend the money. With ITR we had learned a little bit about how to do interdisciplinary panels, about how to parcel out the money, and about giving [ITR] to someone as their full-time job to do it. In 2000 it was Dick Hildebrand. He had very little support, clerical or administrative. He did a lot of things himself. There were very little tools to help. (Interview NSF CISE)

This excerpt is insightful but convoluted. The NSF officer is describing what has been learned from the KDI program, and the first year of the ITR program (FY1998-9 and 2000 respectively). The learning is primarily logistical: how to assign responsibility, how to organize interdisciplinary panels, how to distribute funds and material means to keep this all in order.

For KDI and the early ITRs the responsibility for organization laid primarily on the shoulders of a single program director (PD) who, in addition, was also responsible for his funding program. Following what came to be understood as the sheer size of ITR in FY2000 relative to KDI (not fully experientially clear until after a full year iteration), a PD was assigned exclusively for each year of ITRs FY2001-4, with a substantial increase in administrative support staff, and the creation of administrative 'tools' to assist in this. Tools included organizational forms for composing the solicitation (a complex cross directorate task cf. the Coordinating Group, this chapter section 1) review and panels, lists of possible reviewers and their expertise, templates for letters to prospective panellists, categories for kinds of proposals and a file system to organize the thousands of submissions.

Officers in GEO describe the chaos of the first year of ITR, the various pains in working with cross directorate evaluation panels, and the informal learning collated to improve next years functioning:

> there were some panels that were just yelling. And, God, they just went on about money this and money who! Some people were a nightmare, just trying to get as much as they could for their little discipline. We learned not to invite them onto panels the next year. There were people from tectonics [speaker's specialty] that I didn't invite back the next year just because they were so detrimental to the entire process. Yes, we wanted funds, but no, it was nor worth their headache. (GEO Officer)

The size of ITR and the number of submissions, meant panel members themselves were necessarily requested to participate from year to year. Over time panellists came to understand the mechanics of evaluation, the modes of cross directorate funding, and the valuation of reviews: "I remember one guy who just freaked when we didn't fund a project that had five 'excellent' reviews. Five! But the next year he calmed down when he learned that reviews are just one step" (GEO Officer).

The experience of KDI, and the first run of ITR, had led to the creation of a more refined system for organizing responsibility, for mobilizing support staff, in experience with multidisciplinary panel organization and personal experience in participating in multidisciplinary panels or how to exclude those that are deemed incompetent. The crucial materials for supporting organizing, such as file systems, lists of reviewers and panels, categories for arranging proposals had also been developed (Yates 1989; Berg 1996). Overall it is clear that the second year of ITR was considered a substantially smoother logistical process: "in the second year, things ran much more smoothly," (CISE Officer).

NSF officers describe the experience of KDI, and the early years of ITR, as an accumulation of resources for managing the large-scale cross-directorate program. We have seen the production of a structural 'funding push,' but to understand this *in situ* we have also followed some of the organizational practices, routines, and material arrangements to which explain its functioning in action. To make a cross-directorate program *run* is a logistical question solved in a situated manner with resources such as 'structure' and 'tools'. Just as in CI social science is considered a component along with IT and domain, with the KDI and ITR programs another resource in managing and evaluating these grants is social science research. The findings of this research come to inform the practice of future programs.

Multidisciplinarity as a Site of Research and Formal Organizational Learning

This personal and organizational learning as skill and material arrangement is coupled with a more formal set of findings from 'social science'[24] research. Following the close of the KDI program a study was commissioned from within CISE/NSF to evaluate the program and its grants. The researchers were Jonathan Cummings (then of MIT, Sloan School of Management) and Sarah Kiesler (of the Human-Computer Interaction Institute at Carnegie Mellon); their research is published as an NSF report (Cummings and Kiesler 2003) and a peer reviewed article appearing the science studies journal Social Studies of Science (Cummings and Kiesler 2005).

[24] Within CI and CS circles the term 'social science' is often used as a generic header for economics, sociology, and communication but also management, organization and information studies. While initially uncomfortable with the term and its blanketing of *disciplinary difference* it does occasionally simplify interactions. Identifying one-self as a sociologist is likely to lead to raised eyebrows while social scientist increasingly receives interest, curiosity and even familiarity from CI participants.

Cummings and Kiesler framed the research as being *about* multidisciplinary projects which were conducted in multiple institutions and primarily involving research and development of information technologies. This framing closely parallels the language and goals of the ITR grants. The study was primarily quantitative and survey based, focusing on coordination mechanisms used and outcomes from the KDI projects. KDI outcomes were measured by self-evaluation of project members relative to NSF's explicit goals as a science funding institution: generation of new ideas and knowledge; generation of tools and infrastructure for research; training of scientists, and; outreach and public understanding of science and engineering.

In an interview, a CISE officer set up the study and summarized its findings:

> the main finding is there was that multi-institutional projects were less likely to produce new ideas, people or tool-related outcomes unless there had been coordination mechanisms put into play by the project. So if they did do things like videoconferencing each week and regular PI meetings where the whole group would come together or exchanging students or faculty would go there – to the extent to which they did those things, that mediated the relationship between the independent and dependent variable of the number of institutions and then the number of new ideas produced.

While this line of discussion was prompted by my questions about the organization of ITRs, and my identity as a social scientist, I did not specifically ask about this study or other formal findings.

The findings of social science are endogenous to the practice of multidisciplinarity at the institutional scale. Here is one of the blind-spots of Weingart's argument and any study which only asks in clean binaries 'are we multidisciplinary?'

'Are we Mode 2?' That is, despite the enormous profusion of theoretical definitions of multidisciplinarity, studies of multidisciplinarity, public calls for more multidisciplinarity, policy statements for multidisciplinarity, funding solicitations with encouragement for multidisciplinarity, and projects, program, and centers founded as multidisciplinary -- following all of these programs it is significantly misguided to continue to ask 'are we multidisciplinary or is it merely legitimating ideology?'.

We must now, instead, acknowledge the existence of the phenomena 'multi-' 'inter-' 'cross-' 'trans-' 'disciplinarity' as active organizing categories within academic circles, within extant research projects and the institutions of science. Our question then is to observe its configuration in thought, its operations in practice, and it's embedding in organization, technology and institution.

In direct continuation from the excerpts above listing the findings from the KDI study, the officer immediately noted some of the changes she initiated in the ITR program as informed by Cummings and Kiesler's study. In particular she focused on the finding that multi-institutional studies were found to be effective only with the establishment of regular coordination mechanisms:

> So I learned that. There had been Management Plans that were required of the largest awards in the early years [of ITR]; in the last year [of ITR] I said 'no', let's have a Coordination Plan because it's not just 'how are you going to manage these things' but how they are going to coordinate across the institutions and across the disciplines. In the problem of integrating, we know that distance matters. (CISE Officer)

In summarizing the study's findings above, the officer was bringing to bear the resources generated by a social science study of KDI on the ITR program[25]. In the phrasing of ethnomethodology, abstracted concerns about distanced collaboration were made locally relevant (or endogenous) to the ITR program (see Ch. 3). Here abstracted findings about multidisciplinarity –'collaborations function only with regular use of coordination mechanisms' – is translated into a policy change within the ITR program: in FY2004 rather than a management plan large-ITRs required a coordination plan.

In a further nod to social science, the clever final catch-phrase in the excerpt above is a reference to an article that has come to be quite popular within social informatics: Olson and Olson's "Distance Matters," in which they argue against naïve utopian notions that in science distanced collaboration is unproblematic (2000).

Multidisciplinarity as Actor's Category and Organizing Principle

One of Weingart's strongest criticisms against the reality of post-normal (Funtowicz and Ravetz 1993) or Mode 2 science (Gibbons, Limoges et al. 1994) is the need for greater empirical evidence supporting the phenomenon. He notes that the majority of examples in Gibbons et al. are drawn from those spheres of science closest to policy, political consequence or the public spotlight, such as climate change or technology assessment. Thus little surprise that there are close relations to social economic and political sciences, and due to controversy great collaboration amongst the

[25] This same study was mentioned during a GEON PI meeting, October 2005, suggesting that it could offer much advice for communication and coordination issues in GEON (the reference was not by me). As we will see in Ch. 3 ,5 and 4 social science findings percolate, in a variety of manners and in various degrees of formality, into the activities of Geoinformatics, knowledge representation or GEON's organizing activity.

natural sciences[26]. Instead, Weingart's 'litmus test' of change in science requires a 'generalized' claim across the sciences, not those limited to obvious contact with public and political spheres[27]. He then defines the criteria by which multidisciplinarity should be tested:

> The crucial criterion to determine, if, indeed, new transdisciplinary lines of research are institutionalized in this way, is their relation to established disciplines. This is defined, on the one hand, by the stability of the contexts of application as locations of knowledge-production and, on the other hand, by the independence of their quality standards from those of the recognized disciplines. This is the litmus test of all theses about the emergence of new science (Weingart 1997)

If a research activity is conducted in close relation to established disciplines then we have "only the variation of the existing one" (Weingart 1997:600). A discipline not in 'close relation' is defined by the stability of contexts for scientists to conduct this new research (e.g. a center, a university discipline), and independence from quality standards of established disciplines (e.g. peer reviewed publication in major journal). In short, multidisciplinarity for Weingart is defined by institutionalization and endogenously generated standards for 'what is a contribution to the field'.

---

[26] Gibbons et al. can easily be read as a programmatic call to research, under the framework of Mode 2. In this sense Weingart is perhaps unfair: the bombastic claim to new modes of science are in fact calls to research to understand this new mode. That there is little systematic research to support the claim is why Gibbons et al. must make a programmatic claim.

[27] Certainly, as we have seen, ITR and previous IT application projects such as KDI, or PACI could be sites for investigating Weingart's claim: they are (defined as) multidisciplinary collaboration 'not in the spotlight' of politics, and which span the sciences. In the ITRs alone: GEON/geosciences, LEAD/atmospheric, SEEK/environmental, GriPhyN /physics, SCEC/seismology. Each of these, in their own terms, spans IT/domain and domain/domain expertise. However it is not clear that these projects would fit Weingart's a priori definition of multidisciplinarity.

For this dissertation Weingart's criteria are distinctly counterproductive. Once again the framing of the question on multidisciplinarity as a two-way 'litmus test' has cut out the innovations of ITR: those very activities at the institutional scale that actors define as multidisciplinary. If we define 'multidisciplinarity' as necessarily the activity of scientists participating in research we miss the generation of an organizational skill-set and material arrangement for configuring multidisciplinary research within the funding institutions of science[28]. To the extent that there is an 'after-ITR' program, continuing CISE officers and their staff will have gained logistical skills in scaling up from KDI to ITR, other directorates will have had extensive experience in evaluating cross-directorate proposals and thereafter overseeing large-scale long-term IT projects, and reviewers and panellists will have participated on the 'inside' of an NSF cross-cutting funding project.

Weingart's second criterion is the independence from the evaluation standards of the existing disciplines. Yet this a priori definition completely misses how CI came to be funded as 'basic science.' The solution within CI projects funded under ITR does not require a 'new' set of criteria, per se, but rather structures projects such that the *already existing* criteria of *two* disciplines are to be met. This clever arrangement is a product of extensive negotiations between NSF officers in CISE and the domain, as well as participating scientists submitting proposals. This is the topic of the next section.

--

This section is an anthropologization of multidisciplinarity. As Marylyn Strathern has noted in her studies (and participation) at the borderlands of disciplines:

---

[28] A 'stable context' also seems in direct contradiction to infrastructure as is seen from within CI. Is the stable context the supercomputer centers? Most of the work there is still 'pure' CS. Is the stable context GEON itself? It is not clear if a context is a 'place,' or includes distanced collaboration, as in GEON and most ITRs.

> Take various scientific disciplines, or disciplines from the humanities or from social science form that matter, and you will find interdisciplinarity practice well entrenched. This does not deter current rhetoric, which sees new opportunities in new combinations of interests. (Strathern 2004: vii)

To this statement I would add new practices, organization and tools. Strathern takes multidisciplinarity to be an actor's category and instead relies on the term 'interculturality,' by which she means "the condition of already inhabiting one another's cultures," (1). The perspective is somewhat monist, observing as lines are (re)drawn across a whole. Information technologies are deeply entrenched, in myriad ways, with the geosciences. To speak of a multidisciplinarity, the creation of a balance and twice basic research is to operate on an existing state of interculturality by drawing out boundaries and then crossings those differences.

2.3 – The GEON ITR: Twice Basic Research and Infrastructure to Boot

> You're interested in how things work, how things get started. And my point is it is ultimately the quality of the science and the people that is the largest factor. But also as with any human enterprise, there's an element of who knows who, who knows what, and are you there at the right time, and all that kind of stuff. (GEO Officer)

Since the kick-off meeting GEON's geoscience participants have commented on a *community opinion* that placed doubts upon GEON as a contribution to the solid earth sciences. In many senses this has been an ongoing formative controversy (Collins 1981; Scott, Richards et al. 1990) previous to GEON, and continues in its daily life. Within computer science a similar, if muted, debate has played out over whether the infrastructure building is a contribution to the field. Within NSF this controversy, 'writ large,' eventually culminated in the creation of an Office of Cyberinfrastructure (OCI) formally divided from CISE[29]. This debate is drawn along and across disciplinary lines: is GEON geoscience research, an exercise in computer science, or perhaps neither (Hakala and Ylijoki 2001)? The lines of argumentation can be summarized into two prevalent disciplinary arguments:

i) GEON is not engaged in computer science research, but merely in the *application* of computer science theory and invention to geoscience problems and research;

ii) GEON is not engaged in geoscience research but in experimenting with information technologies not yet sufficiently developed to contribute to practical earth-science questions;

And since GEON is part of a larger effort to carve out a third, hybrid, response, another position is often articulated:

iii) GEON dovetails applied computer science research with the development of tools to be used in addressing today's geoscience questions.

In this latter response GEON has aligned itself not only with computer and geoscience but with a third camp, that of cyberinfrastructure. In doing so it has opened itself to the demands of an emerging and undefined set of concerns in addition to the narrower precincts of disciplinary advance. GEON must craft itself at the intersection of computer and geoscience as well as in line with the ambitions of cyberinfrastructure.

---

[29] OCI was formed in June of 2005. I will not trace this reorganization as this relates more to GEON's future than its formation. However, points of note include that OCI is composed primarily of those officers formerly within CISE in the division of 'shared cyberinfrastructure.' A substantial portion of CISE's budget moved along with these officers (depending on the calculation approximately one quarter – OCI's FY2005 budget was $127M, requested budged for FY2007 is $182M). The Supercomputer Centers (no longer of that name or PACI) are under the directorship of OCI, as is the TeraGrid. Perhaps most relevant for this dissertation as of June 5th, 2006, Dr. Dan Atkins assumed a position as director of OCI.

In this section I focus on the debates over GEON as a geoscience project. There are three objectives: i) to open up the view of the NSF as a distanced peer-review funding institution and instead take it as a work of ongoing organizing; ii) to understand (what come to be) GEON actors as intrinsic to the process of NSF funding, and in particular iii) to understand them as participating in the larger process of defining GEON as a 'balance' between CS and IT basic research.

We have already seen in the section above the mechanisms of the ITR program for generating a funding push to IT research, and incentives for the directorates to collaborate with CISE. This is a balance of directorate participation. However 'balance' in the Atkins report, is conceived in terms of contributions to the field rather than the ostensive participation of a directorate: "we need balance (and better yet, real synergy) between extending the frontiers of computing and extending the frontiers of science using computing" (Atkins 2003:40). In section 2.2 we have seen 'multidisciplinarity' as organizational work and learning at the institutional scale, complicating any a priori definition that understands it as solely the work of practicing scientists.

In this section we will see how actors have come to define multidisciplinarity as intrinsic to the standards of disciplines, as simultaneously meeting existing multiple definitions of basic research. To begin with a definition of multidisciplinarity as requiring new modes of evaluation (Weingart 1997) is precisely the opposite of how actors, through iterations of GEON proposals and negotiations over funding, came to define their goals. In the case of GEON this occurred over a series of iterations in writing, 'selling' and evaluating the ITR proposal.

GEON as Computer Science

The first GEON ITR proposal submitted in 2000 was not awarded funding. This rejection has remained heavily on the minds of both IT and domain PIs. The strength and character of GEON's tie to the domain has remained significantly at play in the design, enactment and outreach efforts of GEON. Thus, one everyday method has been to consciously and continuously foreground 'science' in the various facets of project development. In particular this has meant returning to the first proposal and its failure. Much has been made of the first proposal within GEON, and its primary significance has been along the boundary between CS and geo:

> We have been told by the NSF again and again, and this is why we didn't get that first GEON, that we need to focus on the science, we need to learn something new about the Rockies! And we have to be careful about how to show our [geoscience] side of the work. (All-Hands 1 Geo PI)

In casual discussions and conversations just about every GEON PI has opinions on the cause and the significance of the first proposal's problems. In summary, by these accounts the first GEON proposal was rejected because it was i-insufficiently focused on the earth sciences, ii- attempted too great a breadth of the geo-domains, iii- had a lack of clear 'science questions' and empirical sites of investigation and/or iv - had too ambitious an information science/digital library vision.

Just as with the proposal, in the daily life of GEON much has been made of GEON's funding: how and why the ITR was awarded? What are consequences of funding GEON for geoscience research? How the funds and resources themselves are disbursed? In meetings or gatherings GEON PIs have regularly conveyed a sense that the geoscience community believes the project was funded at the expense of other geoscience research. There are two basic variations on the argument:

i) GEON has taken money from the geosciences, funds and resources which could be better spent on pure geoscience research, and

ii) GEON's funds could have been better apportioned to multiple geoinformatic projects.

The first argument is often dismissed by GEON participants and NSF officers alike through a differential accounting of the funding mechanism. However it is perhaps the most consequential as regular efforts are dedicated from within GEON to frame the project as geoscience research, and from the NSF, to ensuring the community that money has not been withdrawn from basic research:

> But the next [accusation] is you're taking research money away from these other accounts. But that's not true. The truth is that that [ITR] money is a separate line item and you don't have to explain it. People just don't quite understand how those monies are manipulated. I've never injured one of the other programs to do something like this [GEON]. (NSF Officer)

The second argument is more complex, and often deployed by 'insiders': prominent representatives of geoscience institutions who are closer the mechanics of the NSF, ITR or funding more generally. This argument has proven more difficult to dislodge, and in my interviews with NSF officers I encountered it regularly. I will treat these arguments in turn.

--

In casual discussions with geo-scientists at, for example, Geological Society of America (GSA) meetings, the former of the two opinions is most common. The multiple concerns we have outlined of CI, those have informed the KDI or ITR program and that came to be Geoinformatics (see Ch. 3) are not naturally the concerns of the 'everyday geoscientist.' Data, in and of itself, has not usually been parsed as "interoperability" and "preservation" but rather is a resource of the local scientific team and perhaps immediate collaborators. While many scientists attest to sharing data, it is usually done for specific research tasks, and in close partnerships. In contrast, interoperability is framed as infrastructure: a general 'community resource' enabling data finding, sharing and tools for analysis. Infrastructure, then, is general, leaving particular applications as concerns for the domain. This leaves 'a lot to the imagination' for the domain scientist.

During attendance at GSA 2003 geoscientists I spoke with responded to GEON's posters and presentations as "jargon" and dismissed GEON goals as technical or computer science, as "not my science". For example, a poster presentation on ontology at this conference which represented a software architecture design for integrating paleobotanical and geochemical datasets baffled a passing paleobotanist: "and what the hell am I going to do with hundreds of geochemical datasets?!" In a discussion following the presentation I asked him how he saw GEON and he expressed his concerns in terms of 'big science':

> our sciences are not suited to becoming 'big', I've been doing fieldwork on fixed parcels of land for quite some time, and I've watched all the other sciences go big, and

> always said this isn't how we should go. That's what worries me about your[30] project.

The particular understanding of infrastructure described in CI, and enacted in GEON, is not easily conveyed in a language available to geo-scientists. This poster, for example, was composed in the iconography of computer science: a software architecture. While this visual language is clear within computer science and IT circles, it was abstract to many of the 'practicing geoscientists' at GSA (Rudwick 1976; Ashworth 1991; Galison and Daston 1992). Furthermore while 'data sharing' and even 'data preservation' have some currency within the geosciences there are many ways to conceive of its operation and 'interoperability' remained outside these bounds.

The model of interoperability and infrastructure in GEON is particular. We will see in Ch. 5.2 that learning the problematic of interoperability is a substantial investment in technical understanding, a shifting vision of scientific, and in long term planning. At GSA, GEON was often summarily dismissed as 'not geoscience.' To the extent that funding was considered at all, the implicit and sometimes explicit response to the project's funding through GEO was looked upon with disapproval. In drawing on the language and visual imagery of computer science, by grounding their presentations in abstract applications, and because of the complexity of CI's vision of infrastructure, GEON fell outside the conventional bounds of geoscience.

While most common, and perhaps most consequential in active efforts to disprove it, criticism based on 'GEON as taking funds away' is also most quickly dismissed by participants and NSF officers alike. They note that GEON is not funded out

[30] In many such situations I was often considered to be part of GEON. I have never objected to this.

of GEO directorate funds, but out of those in ITR: these are additional funds to the directorates with the stipulation that they must be spent on IT research. As one GEON CS PI noted:

> GEON from a funding point of view is half funded by CS and half by Geo, even though the Geo guys keeps thinking it's 100% GEO. And so they're attitude is "all this money that's being put into CS! Our money!" […] A bunch of computer scientists would never have worked with you [geoscientists] if NSF had not created this thing called ITR, which is now becoming Cyberinfrastructure. So in those terms, it's a new resource for geoscience.
>
> GEON IT PI

GEON is made accountable to the geosciences by pointing to it as an external line of funding to the GEO directorate's base and in collaboration with computer science research.

The second form of argument, that GEON funds could have been distributed *differently* amongst Geoinformatic projects, is more telling. GEON, and many of other CI projects, are often dubbed 'big science'. The term is used loosely, with only a light sense of reference to the big science of the mid-to-late 20thC: government-sponsored laboratories, employing thousands of technicians and scientists, and managed by universities (Price 1963; Galison and Hevly 1992). Instead, within CI circles big science is often used as a synonym for a distributed laboratory (collaboratory) or a short hand for distanced access to large-scale instrumentation (c.f. Welsh, Jirotka et al. 2006).

The term is sometimes used favourably ("geology has finally entered the stage of big science" GEON PI) but more often it appears with ambivalent tones of evaluation:

> R: Do you think GEON is big science?

> J: To the physicists, hell no. It's just a little pimple. But to the earth science community, it's approaching large-scale science[…]if you're out in the boonies with your $70,000 a year grant, it looks like big science.
>
> GEO Officer

While many of the geosciences participating in GEON have had histories of large-scale single project funding from the NSF – e.g. geophysics and seismology – many of the geological field sciences see the project as particularly large, its goals divergent from domain science concerns and, most significantly, that its computational resources will be irrelevant to their scientific practice.

This second argument appeared repeatedly in interviews and informal discussions with GEO NSF officers – specifically because they were well aware that GEON 'could have been otherwise'. Particularly those close to the awarding of the GEON ITR are acutely aware of possible alternate configurations of information technology research because of the actual proposals submitted: "ITR was very popular. Everyone in geology knew about it. We had every kind and size and way of putting research together that you could imagine." (GEO Officer).

Alternatives included funding multiple small or medium sized ITR projects, perhaps in collaboration with CISE. As we saw in section 2.1 there is a certain logic to the ITRs which encourages collaboration through a promise of greater funds, this possibility seems most attainable through the large-ITRs. The large ITR, while requiring the greatest investment from the directorate also provides the greatest return; in the case of GEON this meant several millions dollars from outside GEO/EAR. GEO could not fund proposals the size of GEON without cross-directorate support, and would not simply have the funds of GEON to allocate to other projects. Furthermore, as several NSF officers and GEON PIs have pointed out, it would have been very difficult to secure as many collaborations with CISE with small and medium ITRs.

In a final iteration of the argument towards alternate geoinformatic configurations, a possibility would have been to create another large scale ITR project which has a narrower geo-scientific focus. This is exactly what occurred in GEON's first attempted ITR proposal. In discussion with GEON PIs the leading account of the first proposal's rejection is a failure to balance CS and geo research. However from NSF Geo officers' 'insider' accounts the matter is framed somewhat differently: in terms of GEON's *competition* with the seismology large- ITR: SCEC. SCEC is the Southern California Earthquake Center and the name of a large-ITR awarded to that institution in 2001.

Making GEON Geoscience

While GEON PIs have usually referred to the first iteration ITR proposal as leaning too heavily in terms of CS, interviews within the NSF frame the discussion as a decision between competing proposals. As we have seen, in its second year the GEON ITR proposal was in competition with LEAD, an atmospheric research program in another division of the GEO directorate. NSF officers describe a (light) competitive environment in which divisions compete over directorate resources e.g. solid earth vs. atmospheric sciences. In its first year GEON was in competition with SCEC another GEO directorate large-ITR; however, this proposal was *within* the same directorate: solid earth science (EAR).

SCEC was founded in 1991 under the NSF Science and Technology Centers program (Hughes 1999) and housed at the University of Southern California; it is considered a highly successful ‘synthesis’ effort within geoscience circles (Aki 2002). The ITR grant of the same name, and awarded to that institution, was specifically geared to “developing an integrated modeling framework that automates the process of selecting, configuring, and executing models of earthquake systems. We will achieve this ambitious goal via an innovative integration of knowledge representation, knowledge acquisition, Grids, and digital libraries,” (SCEC-CME Proposal http://www.scec.org/cme/sectionc1.html - accessed June 7, 2006). The language of this proposal mirrors that of GEON’s; both draw on the keywords of KDI, ITR and CI. As with GEON and LEAD, the grant is collaboratively funded with CISE, and is routed through a supercomputer center (in this case also the SDSC).

--

In the studies of multidisciplinarity a common analytic framework distinguishes between the external and internal factors affecting success. The external factors are the institutions, the support and reward structure of science, and funding opportunities (Berkenkotten and Ravota 1997; Jeffrey 2003). Internal factors may include motivation, interest, coordination or communication (Klein 1990; Klein 1996; Grinter, Herbsleb et al. 1999; Chompalov, Genuth et al. 2002). For example, in an extensive social networks study of multidisciplinarity Diane Rhoten divides academic research into three spheres of attention and organization: extrinsic, intrinsic and systematic implementation (Rhoten 2004). She argues that each of these dimensions requires attention for a successful transformation of the academic sphere, and in particular that we are lacking in university level structures for implementation. Extrinsic attention includes the NSF, other grant institutions, and the extent to which they fund, support and encourage multidisciplinarity. Intrinsic attention includes the motivation levels of students, researchers and professors, including factors such as career rewards (Ziman 1981). Finally, systematic implementation is the site of practice and organization, and the established university structures to support local multidisciplinary collaborations (Fig. 8). All three are required for successful change in university based research operations[31].

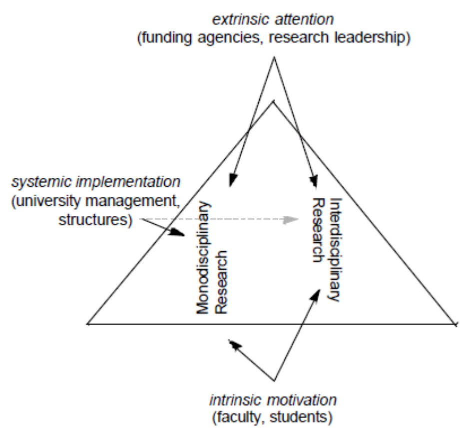

Figure 8: From Rhoten 2004. Three dimensions of multidisciplinarity. Rhoten notes that is a modification of 'the triangle of change' in Huy and Mintzberg (2003).

These distinctions can be useful for a structural analysis and intervention; however, for a situated understanding of a practice of multidisciplinarity they are distinctly misleading. While parallel to the notion of scale in this study, Rhoten's three dimensions or categorical distinctions between internal and external are not sufficiently flexible in portraying the fluidity of actors and arguments across the three dimensions. Rather participants travel freely across the insides and outsides of these projects: GEON comes to the NSF and the NSF goes to GEON.

[31] For the full details of Rhoten's extensive study see (Rhoten 2003)

The walls of the NSF are quite porous, heavy with interactions over future funding, ripe with a continuous flow of 'community input,' and perhaps most importantly, maintaining a continuous rotation of domain scientists as officers. The GEON proposal had substantial backing previous to submission; it had two iterations of formal submission (four if you include the pre-proposals), the generation of community support through CI and Geoinformatics workshops (see Ch. 4) and a visit and workshop at the NSF by GEON PIs, all previous to funding.

The institutional, organizational and technical scales, as defined in this study, are not composed of distinct sets of actors playing particular roles, but rather simply serve as this analyst's organizing principle for grouped kinds of activities and characterizations. The same sets of actors and places often overlap across the scales. To classify the NSF as an 'extrinsic' motivator of multidisciplinarity is to misunderstand their deep participatory role in arranging collaborations. In the previous section we have seen NSF officers organizing for multidisciplinary and in the next chapter (Ch. 4) we will see them standing in for the community of earth scientists, and redirecting technical efforts to CI. To understand GEON actors as 'intrinsic' (outside the NSF) and NSF officers as extrinsic (inside the NSF) is to ignore both groups participating role within the NSF and in GEON. Throughout the two iterations of funding (those that came to be) GEON participants were engaged in extensive negotiations for securing an ITR and defining GEON as earth science research.

--

Nominally proposals for ITR are solicited, geo-scientists write and submit, and the NSF organises review and funding. In practice the roles are not nearly so distinct. Even within a more standard funding model than ITR there is often regular contact between a PI and the respective program officer, careful tailoring of a proposal to the framing of the solicitation and current funding trajectories in the field (Calvert 2006). With the stake of large ITRs, and particularly with a plan to create a 'universal platform for the geosciences', negotiations between PIs and various actors in the NSF were vigorous.

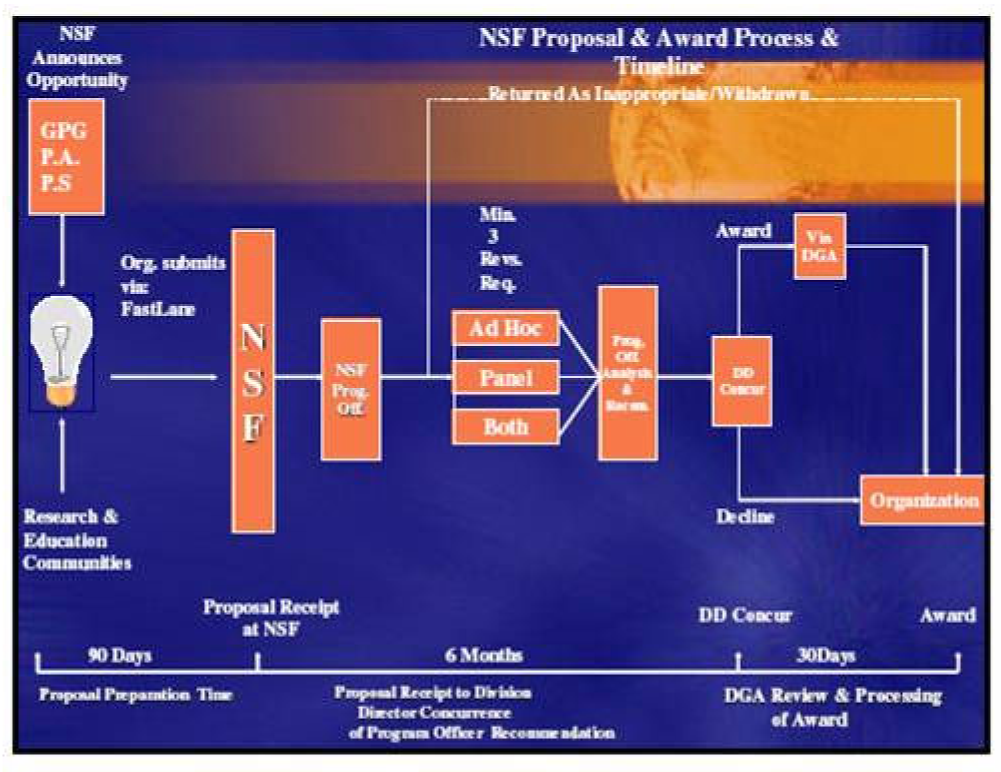

Figure 9: A formal organizational chart of the NSF proposal solicitation, and review process.From CISE NSF Officer slide-set.

The ostensive process for large ITR is itself complex[32]: a large ITR requires a pre-proposal, permitting a round of suggestions (and possible foreclosure) previous to a full proposal. Full proposals are submitted to double-blind review. These reviews are

---

[32] Strum and Latour (Strum and Latour 1987) distinguish between an ostensive and performative accounts of social order. An ostensive account relies on reified understandings of structure and order; society operates upon a relatively stable foundation of persistent arrangements. As with ethnomethodology, a performative account focuses on the continuous negotiation which constitutes the enactment of structure and order. Unlike canonical ethnomethodology, for Strum and Latour a performative account also considers material arrangements ('technology') and how these come to act in the reproduction of social order. See also (Feldman and Pentland 2003).

available to the multi-directorate panel proceedings, and then final decisions travel up a chain of program, division and assistant directors. Informally, negotiating GEON at the NSF involved a flurry or inter- and intra- directorate negotiation; future GEON participants were part and parcel of this process.

There is a great deal of sensitivity about the internal process of proposal evaluation and selection at the NSF. In my interviews and informal discussions with NSF officers, and occasionally with GEON PIs, all such information was given with trepidation and qualifications of what can or cannot be revealed. Part of the concern is a fear that knowledge of a failed proposal submission will become public and thus possibly affect the prestige or future funding possibilities of the submitting PI(s). I have removed details of particular reviews and individuals, but attempted to maintain a sense of a remaining debate following review, as program and division directors in panels make decisions informed by reviews. However, given the lived significance of GEON's first failed proposal it is not possible to avoid treating the matter.

Also, to understand the institutional setup of GEON I have found it necessary to leave in the names of GEON's competitors for the first and second proposals (SCEC and LEAD, respectively). In the broadest terms, all three projects were considered highly competitive, and were sufficiently well received by reviewers to be considered by a panel. Furthermore, each project received some degree of support from CISE with regards to financial collaboration in funding.

--

Because both SCEC and GEON had support from CISE, the debate over funding was primarily within the GEO directorate, and mostly involved the earth science research division. Accounts of the panel discussions frame the competition between SCEC and GEON in familiar terms: as a balance between CS and Geo, and as a proper coordination across these boundaries.

> the success of SCEC [is that is has] a really good IT department that they worked with. But […] some of these are geoscientists with some computing skills trying to develop stuff. The reviewers were saying these are really the right people to be doing it. The challenge here is the interface of these very different communities, the IT people. What are they? Are they engineers or are they scientists? How does their brain work and think? Selecting the best geoscientists or working with the best IT people? I think SCEC got that right. But at the time I questioned whether GEON got that right. (GEO Officer)

SCEC in 2000 had already established itself as a successful center for synthesis, and as a clearinghouse for seismic data. It contained 'in house' technicians; we have already seen the wariness of geo-scientists towards the SDSC and its computer scientists (Ch. 2.1). While the SDSC is attempting to define itself through an expertise with data, this officer leaned towards a vision of integration in which geoscientists remained the primary investigators, and technicians served to provide tried and true means for integration. However, GEON and its SDSC partners did not sit idly back and allow themselves to be defined out of geoscience.

The first suggestions for a single clearinghouse for geoscience data emerged from NSF sponsored workshops for 'Geoinformatics'. The Geoinformatics workshops were also the originating point for the GEON proposal and the collaboration with the SDSC (and thus CISE). Following the workshops and the bulk of the work in composing a proposal, GEON PIs arranged a seminar at the NSF: "GEON: A Defining Opportunity for

Earth Science Research". The seminar was held only two weeks before the proposal deadline and focused on what remain many of the GEON mantras:

> The vision of the GEON consortium is to create a distributed, interoperable, and scalable **Earth sciences information system** […] With a its strong emphasis on ease of access and use […] GEON will have a revolutionary impact on the way geoscience research is conducted (Seminar Announcement, emphasis in original)

GEON's first iteration proposal had already mustered significant support from within and without the NSF. A pre-proposal had been approved. Also, USGS had agreed to contribute substantially with 'in kind' services. Letters of support included institutions such as NASA and the US department of Commerce, and private industry such as ORACLE and ESRI (best known for databases as GIS, respectively). However, accounts from NSF officer supporting GEON describe a poor response to the seminar: "I didn't like the discussion. I didn't think they were getting their due." (NSF Officer). Officers who were less favourable describe a scepticism to visions of CI and large scale information infrastructure projects for the geosciences:

> I hate to see the way we train geologists, people sitting back and looking at databases, pulling out data, strikes and dips and structural geology and geochemistry and so on, and not actually going out into the field. I get a feeling there is going to be a lot less going out in the field. […] I like to see people trained hands-on in the field. I'm a little concerned that this is going to take away from that.[…] I just hope that the people who are looking at strikes and dips on a computer know what the hell they're talking about. We're still buying geology picks, not just for show and tell.

Large scale and informationally centered projects are seen as distant from the 'hands on' practice of the field-based earth sciences. The project visions of SCEC-ITR and GEON are distinct[33]. While both have come to be called CI, only GEON identified itself as such at the time. SCEC-ITR has a domain emphasis on seismology, and while still undertaking an umbrella project (seismology, geophysics, tectonics), these disciplines have already been known to collaborate on well defined scientific questions (SCEC is founded upon such enterprises, and has a strong track record of publications and long term funding). Meanwhile GEON's vision is relatively new, 'community cyberinfrastructure,' with what to geoscientists appeared as clearly defined information technology research goals, and only vague links to familiar geoscience[34]. In its first year, while up for consideration, GEON's vision for CI had not yet been embraced within GEO NSF.

It was also clear that the earth sciences division (or for that matter the GEO directorate more broadly) could not fund two large-ITRs even in collaboration, "it would have been more than a little hard to get 2 of these" (GEO NSF officer). In a competition for resources GEO directorate officers were unwilling to choose a vague 'platform' activity, appearing to lean heavily in the direction of computer science, over a more narrowly defined and geoscience led project.

As I have noted, in the first round GEON was rejected in favour of the SCEC. Officers and GEON PIs concur that proposal reviews were sufficiently clear in favour of

[33] "GEON is the only what I would call truly platform activity. Every other proposal we have over here is a single domain, specific project […]now there isn't anything else like GEON," (GEO NSF Officer).
[34] Later endogenous comparisons of GEON and SCEC are still informed by these notions: "SCEC is earthquakes in southern California. […] It's paid off; it's been very productive. They've gone after low-hanging fruit; they've got it; they've published it. It's been used by the [NSF] Director all the way to the top as an example of 'this is how you use supercomputers, this is how you use the TeraGrid'. But it's very different from GEON. It's not a platform, it's a project." (GEO NSF Officer).

a greater weighting towards geoscience as to merit a substantial reframing of the GEON proposal. One method for this came to be the inclusion of the empirical application sites which came to be called ‘test-beds’:

> the first competition there were 2 primary Geo-related proposals, the SCEC proposal and the GEON proposal. The SCEC proposal got funded that first round primarily because it had an identifiable piece of science to do along with the IT part of it. There was some cutting-edge science and that, of course, was hazard assessment. So the message went back to GEON that they almost made it. […]But the message was that you need some identifiable science in the proposal, so when they came back the next time they had these two test-bed science applications, and it made it. It was touch and go that time between GEON and an Atmospheric Science proposal [LEAD] (GEO NSF Officer)

This GEO officer links the success of the second proposal to the inclusion of test-beds. While GEON’s two test-beds (see Ch. 4.1) were enacted by participants in re-writing the proposal, these are in fact common across many of the ITR CI project. BIRN, SEEK, and LEAD each draw on the language of ‘test-beds’ as sites of empirical application[35]. GEON’s test beds are specific to the project, but the notion of a test-bed appears to be an effective trope in engaging domain scientists in informationally driven technology projects.

In the second proposal GEON IT research was tied to specific sites of physical scientific activity through two geographic sites; the mid-Atlantic and the Rocky mountain test-beds. These test-beds have been salient sites for geo-scientific research for over a

[35] The meaning, or referent, for test-beds across these projects varies significantly. Each usage shares in common a focus on the application of novel information technologies to empirical domain concerns. However what a test bed may actually be varies considerably: domain datasets to integrate, specific science research questions to tackle, or institutions in which to foster collaboration. GEON’s test-beds are geographic locations.

century. They have well documented science questions which span various disciplines and they have multiple datasets in each area. For example, data integration and visualization in the mid-Atlantic can serve to help bring together data, evidence and scientists in current controversies with seismology:

> continental growth through accretion of approximately twelve terranes has been proposed, but current models do not agree on criteria for identifying the terranes and present various accretion histories. To assess this mechanism of continental growth, the models must be constrained by multidisciplinary observational data sets within the region that will be assembled by the GEON geoscience team. (Keller 2003)[36].

The test-beds came to organize multiple PIs as they re-wrote the proposal: "they gave us something to work together on. Our research questions became more coordinated, and more precise. I think our [half of the] first proposal was vague, unlike the IT side," (GEON Geo PI).

Furthermore in the second proposal a greater emphasis was placed on 'routine technical work.' The first proposal had been seen as lacking sufficient planning for technical activities outside of CS research. Experienced reviewers and panellists noted that the application of novel IT required substantial mundane (or non-research) work on the part of information technologists. The second GEON proposal allocated funds and time for such work. It acknowledges that some low-level service provision by the IT team will be necessary for non-research aspects of integration: "While not viewed as IT

[36] Terrane: a geological term for a crustal block or fragment that preserves a distinctive geologic history. A terrane is different from the surrounding areas and is usually bounded by faults. Historically oriented geosciences often trace the movement of terranes, which may be considered subsections of a plate or crustal block. The term should not to be confused with 'terrain', which is a much more commonly used term used to describe the physical, usually visible, features of a land formation.

research, format and schema incompatibility issues will be addressed specifically for the test-bed related databases." (GEON Proposal p.5).

A rewriting of the proposal emphasizing geoscience research, through the inclusion of the test-beds, and a shifted emphasis away from computer science research and towards technical support aided significantly in GEON's second round of ITR. However an outstanding issue remained: the scepticism and unfamiliarity with the notion of a 'general informational platform for the broader earth sciences'. In discussions at GSA, or even at the NSF, this notion remains alien to many geoscientists e.g. identifying himself as a paleobotanist and pointing to a nearby poster entitled 'A 3-D fault interaction model to investigate the topographical process' one geoscientist said this to me about GEON "I don't understand how its going to help my research and his research." Rewriting GEON's second proposal was only half the battle in enacting a vision of cyberinfrastructure within the geosciences. The next chapter traces the growth of Geoinformatics and the extension of the notion of a general informationally umbrella for the broader earth sciences.

A Note on Methods: Exclusion and the Limitations of Following the Actors

My research focus has been on the interactionally generated consequences of rewriting the second proposal in light of a failure of the first, and the received commentaries from review and NSF officers. I will not conduct an extended comparison of the first and second proposals. However there is one outstanding question rarely discussed in GEON, but of great importance to social scientists: this is the question of the 'exclusion' of geoscience PIs from the first to second proposal.

As noted, one of the received criticisms of GEON's first proposal was of scope: "there were too many disciplines represented," (GEO officer). This was understood to complicate coordination, communication and collaboration. In response to this the second proposal reduced the number of geoscience PI and co-PIs from 19, to 10.[37] In conjunction with this the range of listed expertises in the proposals for each remaining PI increased substantially from one specialization to three or four. For example, one PI listed initially as an expert in "Geodynamics" in the first iteration, became and expert in "geodynamics, GPS vector database, strain modelling, active tectonics" for the second. In this manner it was thought that GEON could cover a broader range of geoscience expertises while reducing the number of PIs, thus lessening communication and coordination concerns. Also a greater emphasis was placed on education and outreach. Below is a comparative chart of the listed expertises (table 1), PIs 'excluded' from the second proposal are underlined; many of their disciplinary expertises remain.

In active discussion GEON participants have not made much of the change in PIs across proposal iterations. There have been brief mentions that some of the PIs that 'did not make the cut' were unhappy. This said, it has not been a regular public issue or even a topic of discussion during more informal moments such as during breaks, lunch or dinner. The matter has never appeared as an active organizing principle in the manner that the test-beds, or even the first proposal's rejection. Similarly, in my interviews with NSF Geo Officers, no mention was made of this issue.

[37]Also interesting is a rise in IT PIs from 5 to 9 – this despite criticisms of too great an IT focus in the first proposal.

Table 1: Comparison of disciplines represented in first and second GEON proposals. PIs 'excluded' from the second proposal are underlined

GEON 2000 Proposal (Not-Funded)

IT Research (5 Listed)
1. Database systems
2. Visualization
3. Information mediation
4. Cluster and Grid computing
5. Digital Libraries for geospatial data, Metadata modeling

Earth Science Research (19 Listed)
1. Active tectonics
2. Rock properties
3. Petrology, Structure, Tectonics
4. Hydrology
5. Geobiology
6. Stratigraphy
7. Stratigraphy, Isotope geochemistry
8. Geophysics
9. Geophysics
10. Geodynamics
11. Paleobiology
12. Hydrology
13. Geodynamics
14. Structure, Tectonics
15. Geophysics, Information Systems
16., Petrogenesis, Geochronology, Tectonics
17. Stratigraphy
18. Geochemistry
19. Hydrology

Education (1)
1. Digital Libraries for Education

GEON 2001 Proposal (Funded)

IT Research (9 Listed)
1. Database systems
2. Augmented Reality and Visualization
3. Meta-models for Geoscience
4., Remote Visualization for Geosciences
5. Geo-Visualization, Data Models
6.Semantic Data Integration
7.Data Grids
8. Cluster and Grid computing
9. Database mediation

Earth Science (ES) Research (10 Listed)
1. Active tectonics. Landform analysis, fault distribution, and kinematics
2. Petrology. Metamorphic history, pressure, temp., time databases, structure maps
3. Geobiology. Fossil occurrence and biostratigraphy database
4. Geophysics. Potential fields, remote sensing, lithospheric seismology databases
5. Geodynamics and numerical modeling
6. Geodynamics. GPS vector database, strain modeling, active tectonics
7. Structure, Tectonics, geologic maps
8. Geophysics, Information Systems, and broad geoscience data sets
9. Stratigraphy, stratigraphic databases
10. Petrogenesis, Geochronology, Tectonics. Database on igneous rocks, geochronology, spatial distribution of igneous rocks and faults

Education and Outreach (3 Listed, 2 overlap)
1-3 Digital Libraries for Education

The question of the exclusion of individuals, disciplinary specialization or perhaps even of types of data is clearly one of significant concern in the development of an umbrella infrastructure for *the earth sciences*. It speaks to the distribution of resources, the development of tools and the access to informational resources. GEON participants regularly discuss, even debate, such issues amongst themselves. Particular participants speaking in the name of a disciplinary specialization often claim an uneven distribution of developmental efforts in GEON. These actors are able to problematize technical trajectories by virtue of their presence, their voice, and attempt to reframe discussions (see Ch. 4.2). However, the methods of this study do not easily lend themselves to the exploration of subjects not actively discussed by participants. Those PIs from the first proposal have not been part of GEON's daily, weekly, bi-annual or annual meetings. They do not appear in video-conferences, list-serves or emails. They are not part of GEON's activity as it has been defined in this study. It is with such matters that we can see the limitations of tracing actor's categories and practices. The benefits of an analysis based on following the actor are circumscribed by the movements of those actors; it is an effective method for outlining a controversy in discussion, an exclusion in action, and consequences as they are framed by participants, but to the extent that actors are successful in an 'exclusion' (and then never speak of it again) there is no further data to collect[38].

--

It would perhaps be too strong to claim the generation of an 'organizational push' to cyberinfrastructure. With the 'funding push' (Section 2.1) we have seen multiple modes of creating incentive: a large pot of funding, distributed unevenly to encourage collaboration, and with a topical focus on IT. What has been described in this section is closer to a diminution of the institutionalized barriers deterring multidisciplinarity (Rossini and Porter 1984; Butler 1998; Gershon 2000): building (more) lines of communication and organizational forms connecting directorates, collecting a competent base of evaluators of diverse technical proposals, generating a skill-set around large long-

[38] One possible analytic strategy is well described and executed in 'social worlds' analysis(Fujimura 1988; Casper and Clarke 1998; Clarke 2005). Here it understood that to be silenced or excluded is highly consequential to the outcome of an endeavor (Bowker 1988; Star 1991; Bowker and Star 1999). To the extent that one can identify a social world, it is possible to note participation and exclusion. Exclusions can themselves be traced and consequences can be generated based on this data. Following this approach, it could have been to actively seek out for interviews those PIs left out of the second GEON proposal.

term funding &c. Also in this section a model for enabling ‘scientists to build their own infrastructure,’ has been described, which, as per CI, is able to function within the confines of the NSF's mandate for basic research. The primary vehicle for this is a notion of ‘balance.’ First, balance between research topics, and goals, in CS and domain. Also, balance in the number of PIs in their respective fields and research institutions.

Defining GEON as balanced between CS and Geo was an iterative process, across multiple proposals, through review and feedback. In its second round GEON was in competition with LEAD. As with SCEC, LEAD is also ‘narrowly defined’ in terms of particular hazard prediction sciences, in collaboration with a supercomputer center (NCSA) and with geoscience lead institutions. However, in this second competition GEON was awarded the ITR. In section 2.1 we have seen that there was some competition across the earth and atmospheric divisions, however the relations are primarily collegial. In the next section we trace the generation of a mounting ‘pull’ for CI.

Conclusion

In this chapter I have argued against the a priori definition of ‘what is multidisciplinarity’ and instead have followed actors as they engage in the activities they define as multidisciplinary. We have seen this structured into the ITR program, which encourages cross directorate collaborative funding in the NSF, in the organizational practices of enacting ITR across the directorates, and in putting together a proposal for ITR which balances IT and domain research. In each of these three scales -- ‘institution, organization and technical’ – we have seen an activity of multidisciplinarity. In three distinct approaches disciplinary differences are characterized and then bridged. In practice this has included an organizational learning in the form of acquired logistical and mediation skills; a material production of 'tools' to organize this activity; and, in language, it has included a framing of infrastructure projects as a balance between disciplines.

From ethnomethodology we could call this the '*topicalization* of multidisciplinarity'(Garfinkel and Sacks 1970; Lynch 1985; Francis and Hester 2004). . I have argued that, in particular, Weingart's definition of multidisciplinarity and a 'hypothesis test' as to its reality cuts out the broad and consequential activities of actors 'doing multidisciplinarity'. To be clear, my opposition to Weingart's argument is not intended to lend support to Gibbons et al. or other claimants on revolutions in science. To argue that 'multi-', 'inter-', 'cross-', or trans-' disciplinarity are phenomena in the world is not necessarily to lend support to a vision of an existing Mode 2 science. Rather, it is a respecification of the question: given actors' use of these categories (at the NSF, in the practice of science, in policy circles and in funding solicitations), what can we learn about their operations in practice and its preservation in material arrangements? Where and by whom is multidisciplinarity done? How is it conceived in thought and enacted in practice, organization and technology?

This reframing does not cut out the 'litmus test' question ("are we really transdisciplinary ?"); such research is part of the active field which NSF participants regularly engage as they go about their practices of multidisciplinary. As we have seen, the results in *studies of multidisciplinarity* come to be folded back into the *practice of multidisciplinarity* and inform action. In this chapter we have focused in particular on the work of Cummings and Kiesler, but many such studies come to inform the activities at the NSF. Studies problematizing multidisciplinarity have come to define new goals, new strategies of organization, and new criteria for evaluation. The litmus tests of multidisciplinarity are at play in the practices of multidisciplinarity.

By tracing activity at the NSF, by both members and non-members of that organization, it has been possible to observe a finer granularity of work at the scale of the institutions of science. Multidisciplinarity is an active category of talk, and is enacted in funding solicitations, review and awarding. It appears as funding incentives to work across directorate boundaries, in organizational practices for arranging collaborations, and in negotiations over the disciplinary criteria of basic research. In the NSF, GEON was orchestrated through multiple techniques which formally ranged the entire organization, particular directorates, and also non-NSF members. However, this only half of the institutional scale. The NSF is by no means the only important institutional actor within the geosciences. In the next chapter I turn to the 'communities of geoscience' as represented by institutions such as the United States Geological Survey, SCEC and eventually by GEON and Geoinformatics.

A subtext of the next chapter still relies on the understanding of multidisciplinarity as developed in this chapter. However, I shift in focus to the work of geoscientists to generate 'pull' for CI. This is the *enactment of a need* within the geoscience community and a niche for a particular technical configuration of information technology which has come to be cyberinfrastructure. They do so by drawing on the problem-set developed in CI, within ITR and KDI, and 'in house' within the geosciences.

These concerns must be enacted locally by making them endogenous to ‘Geoinformatics’.

– Chapter III –
Building the ‘Pull’ to Cyberinfrastructure: Geoinformatics –

Making a Public Issue of Cyberinfrastructure

> A significant need exists in many disciplines for long-term, distributed, and stable data and metadata repositories that institutionalize community data holdings.
>
> --Atkins Report

> You’ll never change everyone out there, but I think GEON is making believers of a significant new fraction of the community.
>
> -- Geo NSF Officer

Why would a division of the NSF with an explicit mandate to produce scientific knowledge about the earth invest in information infrastructure *research*? We have already seen a vision of cyberinfrastructure in which (at least) two disciplinary sciences continue their specific investigative work while also generating an infrastructure for the domain. I have called this strategy for achieving funding within the National Science Foundation (NSF) ‘twice basic research and infrastructure to boot’. In this strategy the NSF mandate to support basic research is achieved and coupled to a plan for long term infrastructure development for the sciences. But why would geoscientists engage in collaborations across IT and domain in the first place?

When I pose this question within GEON it is usually met with incredulity. They respond that earth scientists have a need for greater access to computational resources; that they lack visualization technologies to effectively manipulate data about the earth;

that the geosciences are woefully behind in their adoption of information technologies; that the legacy archives of data about the earth are threatened by poor archiving, or; that multidisciplinary collaboration requires the integration of heterogeneous datasets. The GEON project is premised upon an accepted problematization of the current informational state of earth sciences: we have a need for greater tools and resources, we are behind the technological curve, and we have underinvested in the preservation of our data. This is often called the 'pull' to information technology, the 'demand' by geoscience for science centered applications, or, most colloquially, the 'need' of geoscientists for computational resources. There has come to be a comfortable recognition that these are a *public issue* for earth sciences.

There is a more subtle question that GEON participants find more difficult to cast with such certainty: why amongst the possible plans for informationally transforming the geosciences is cyberinfrastructure (CI) the chosen strategy? CI is a particular strategy in the generation of computational resources for the sciences. The programmatic statements of CI envision a single informational base for the sciences more *generally*, and in GEON for the earth sciences *specifically*. They cast the development of these resources as pioneered by geoscientists themselves, as part of a program in basic research, but also in close collaboration with computer scientists. They envision interoperable resources, integrated data, and preserved archives. CI is a highly ambitious, resource intensive and long-term model of informatization. Even the most positive calls to CI admit it to be an experimental and untested approach.

It is not that CI is the only conceivable configuration. Other models for computational development are available. As we saw in the last chapter participants and

GEO NSF officers alike are keenly aware that the GEON project ‘could have been otherwise’. The first GEON proposal was rejected in favour of a more narrowly focused information technology research (ITR) project, SCEC; in its second year GEON was funded in close competition with a similarly tailored ITR project, LEAD. In both rounds of ITR funding solicitation many proposals were reviewed which sought to generate alternate configurations of IT/domain collaborations. These included large scale endeavours such as SCEC and LEAD, but also many medium and small ITR projects. While ITR monies had to spent in a manner which framed activity as contributing to computer science research, cyberinfrastructure was only one outcome of this NSF funding line: “when people tell me that geoscience money was spent on computer science I roll my eyes, but when they tell me we could have funded hundreds of smaller research projects I at least have to listen,” (GEO NSF Officer). Within the NSF, and in accounts of the geoscience community at large, an ambivalence remains as to the CI vision of a single informational base for the collective earth sciences.

The ‘pull’ to CI is a matter of contemporary work in the earth sciences. This chapter traces at the level of action a building rationale to cyberinfrastructure within the earth sciences. I argue that this is a process of making a *public issue* out of cyberinfrastructure (Mills 1959). The most important vehicle for this transformation is Geoinformatics, an emergent organization (Taylor and Van Every 2000) within the geosciences to bring together and make visible informational resources in the earth sciences, in particular to the NSF and the institutions of geoscience. In this chapter I am not problematizing the existence of ‘pull’, ‘demand’ or ‘need’ for cyberinfrastructure, rather I am *respecifying* pull as an achievement in action (Garfinkel 1991). GEON was born out of the work in Geoinformatics as the *private troubles* of individual geoscientists were cast as the *public issues* of the earth science community. In turn, GEON has come to be a participant in Geoinformatics and an actor in generating a greater pull to cyberinfrastructure.

The Atkins Report and other documents of CI make large and sweeping claims about the dangers of data loss, the benefits of knowledge integration, and the revolutionarily potential for emerging tools and computing resources. However, these appeals are pitched at a high level of generality, and for most practicing geoscientists, at a low level of familiarity. The Report is cast in broad gestures which refer to 'the conduct of science' and an equally broad software and hardware application layer 'tailored to domain needs.' It contains scattered references to informational projects specific to the geosciences such as the Earth Simulator[1] and the Network for Earthquake Engineering Simulation (NEES), but these are few and far between. The majority of the report is rhetorically wide-ranging, written in a tone of generality intended to address physics and biology as much as geology. It does not immediately address the particular research concerns of actual domain scientists.

In this chapter I argue that it is not enough to point to benefits of an informatic revolution, it is not enough to preach the dangers of failing to adopt computing technologies, these must be 'brought home' to the contemporary and everyday efforts of 'on the ground' practicing earth science.

---

[1] Earth Simulator is primarily Japanese spearheaded project "targeted at analysis of global environmental problems through simulation of geophysical, climate, and weather-related phenomena." (Atkins 2003). EarthSim in the text serves primarily as call for national competitiveness and an example demonstrating how the US is falling behind in computing and scientific endeavors.

The materials for enacting a vision cyberinfrastructure are emerging from within the earth sciences. At the NSF, the institutors of earth science and amongst practicing geoscientists lack of access to information technologies is increasingly coming to be seen as a source of problems in the conduct of research. Difficulties in easily securing computational resources, the balkanization of data, and threats to the legacy archive are the material for casual conversations at conferences. These are the raw materials for the generation of a pull to cyberinfrastructure. These private problems of geoscientists, told in stories and everyday conversations, must be framed as the public issues of earth science and cyberinfrastructure must be enacted as the solution.

Enacting the pull to Cyberinfrastructure is co-extensive with a transition identified by sociologist C. Wright Mills from a *private trouble* to a *public issue*. For Mills a trouble is the experience of an individual. These are variously considered as unfortunate contingencies or blamed on irresponsible action. In contrast, issues are collective phenomena. Phenomena framed as issues are seen to affect many individuals as they are swept along in changes which could not be predicted and for which no responsibility can be laid. Mills uses examples such as ‘being without a job’. In normal circumstances, in the US, joblessness is framed as the private personal trouble of an individual and their family. Such a trouble is to be resolved by action on their part such as retraining, looking for a job and finding work. However, in the face of an acknowledged crisis, such as a recession or environmental disaster, private troubles can be re-framed as a public issue. ‘Being without a job’ becomes ‘unemployment’. Such issues are a matter of national or international concern to be resolved by action of the state, charity, non-governmental organizations or foreign aid. In speaking of a transition from problems to issues I am speaking of an institutionalization: the mechanisms by which a problem is made relevant to a public and placed under the jurisdiction of a community, public administration or the state (Donzelot 1979[1977]).

As a personal trouble a lack of computing resources are to be resolved by the conventional mechanisms that have been established for the conduct of science: finding resources, securing a grant, allying with an informational institution. For instance, if a specific set of computing or visualization resources are sought for a scientific research question the solution is to acquire a grant from a relevant institution: the NSF, a state-based equivalent, or a private fund. Another solution is perhaps to contact a supercomputer center and secure some 'compute cycles'. Such activity in the US is still 'institutionalized' (Mukerji 1989) but the solution sought is the responsibility of an individual (or small group) i.e. NSF grants, in the past have most often supported particular scientific research endeavours.

To articulate computing resources for the sciences as a matter of infrastructure is to make them a public issue. Cyberinfrastructure is an argument for a different configuration of individual and institutional action. In no vision of CI is the case made lightly: embarking on a quest for universal informatics is a heavy investment in computational resources without clear and immediate 'science returns'. How the concerns of cyberinfrastructure are made relevant to a community of practicing geologists and the institutions of earth science *is also* the production of institutional concerns, geoscientific interest, and technical particulars. This chapter traces the practical work of reframing individual private troubles as public issue within the broader geosciences and CI as the solution to this problem. The primary vehicle for this has been Geoinformatics.

CI is an infrastructuring movement (Kling 1991; Kling and Iacono 1995; Kling 2000). In order to be made relevant to work in the earth sciences a 'local wing' of supporters have had to be mobilized within the earth sciences. The vision of CI is to operate 'from below,' that is, these projects should be spearheaded from within the domain community. In the earth sciences this has crystallized around Geoinformatics:

> Within the Division of Earth Sciences of the National Science Foundation, the effort to create the cyberinfrastructure that we need is referred to as *Geoinformatics* (www.geoinformatics.org) and arose out of a series of meeting and workshops as well as our community's energetic response to Information Technology Research (ITR) opportunities at NSF. (Keller 2003)

Geoinformatics has come to be synonymous with cyberinfrastructure; however, it did not begin as such. 'Geoinformatics,' in just less than ten years, has moved from ideas and ideals, become the title of an institutionalized funding line in the NSF, a regular subsection of the Geological Society of America (GSA), and an annual conference. Each of these activities has been intimately tied to the work of GEON participants, particularly the IT and Geo PIs. Over time Geoinformatics has become almost synonymous with GEON. What remains, what is still contested is the full vision of CI: that GEON would become part of the singular national cyberinfrastructure, the clearinghouse for interoperated data about the earth.

The first section of this chapter provides a brief outline of the first year of Geoinformatics. By drawing on the narratives from NSF officers and GEON participants

I outline three accounts of trajectories for Geoinformatics and GEON. The second section of this chapter traces the practical in situ activities of earth scientists articulating the concerns of CI as the public issues of earth science. The discursive tactics within Geoinformatics have not been dissimilar to those of cyberinfrastructure: problematizing disciplinary difference as hindering scientific research. However, for Geoinformatics these must articulated relative to the lived problems in the earth sciences such as posing a geoscience left behind in a computer revolution; archives of legacy data soon to be lost; and a need to integrate practice, knowledge and data for scientific advance. The third section of this chapter traces the work of participants in their efforts to position GEON as cyberinfrastructure, as the informational platform for the earth sciences, and as synonymous to Geoinformatics.

3.1 – Enacting the Histories of Geoinformatics

For participants in GEON, for GEO officers at the NSF, and for a subsection of practicing earth scientists Geoinformatics has been a key term defining the trajectory of contemporary research. In this section I will trace three actors' accounts of the rise of Geoinformatics. In each account, in these histories, an implicit future trajectory for GEON is enacted. The purposes of GEON and its future role *for* the earth sciences remains an open question *in* the earth sciences. These histories enact these differing trajectories for GEON. They are not the empirically researched histories of detail oriented scholars, but locally composed narratives built up from personal experience and reflection. They serve as sense-making tools for future action. In ethnomethodology these practices are generally referred to as generating *accountability* (Garfinkel 1967), but here I will more tidily refer to them as 'enacting a history, present and future'.

Enactment theory focuses on action and its consequences but takes interpretive practices as part and parcel of these. Enactment, in its initial formulations, specifically referred to the methods by which organizational actors come to constitute the 'outside environment'(Weick 1977). Weick notes that an organization reflexively discovers "what it is up to" in retrospect, making sense of its own actions and thus investing "their settings with meaning" (Weick 1997: 272). Enactment as a methodological concept, brings attention to the practical methods by which organizations and institutions are built *at the same time* as the environment which "impinges on them" (Weick 1997: 267) and also the methods by which these environments are known. These are "the raw materials from which a sense of the situation is eventually built" (Weick 1997: 267).

An environment must come to be known, it must be made tractable and visible to members engaged in practical action. Such methods can range from everyday talk to more formal methods such as, for example, a survey, use of economic and demographic statistics, or tracking hits on a website. These methods are practical and organizational; they constitute an environment and make up the activity of the organization. For example, marketing research relies on such methods as 'focus groups' to gain understandings of target populations. The focus group is the source of knowledge about 'the environment' but it is also a work of practical organizing to execute this method. In constituting their environment as a knowable object, actors are also *organizing*.

Geoinformatics is such an example. It has been an occasion for discussing 'large-scale' issues of IT development in the geosciences. In enacting Geoinformatics actors have been simultaneously defining its own ‘internal’ role and generating methods for evaluating geoscience community needs, resources and its particular problem set relative to interoperability. As a methodological proscription mirrored in ethnomethodology and actor-network theory, Weick argues that “the last thing we want to do is define away their solutions to sensemaking by imposing for them the logical but empirically empty distinction between internal and external worlds” (Weick 1977:273-4). In this chapter the ‘external worlds’ are the geosciences (as ‘a community’, as ‘practicing scientists’) and in this section I focus on the history of Geoinformatics as an enacted category.

A history is actively generated by participants in GEON and Geoinformatics(Chalmers 2000), its enactment is part of sense-making(Weick, Sutcliffe et al. 2005): understanding a contemporary environment and planning for a future. In this section I draw on three actors’ histories, as well as documentary evidence they provided, to pose the histories of Geoinformatics. These histories come be resources in practical action as the trajectories of Geoinformatics GEON and cyberinfrastructure come to be linked.

Each of these three histories draws forth information technologies as a public issue for the earth sciences, but they do so differently. I have chosen to elaborate these three narratives because they are commonly laminated descriptions of the rise, purposes and future directions of GEON. In the first narrative GEON is a final manifestation of Geoinformatics, and both are part of larger CI efforts. This history of Geoinformatics contains a future in which GEON is to be the clearing house for geoscience data. In the second narrative the emphasis is on GEON as one of multiple geoinformatic projects. Work on the informational aspects of earth sciences precedes GEON, and multiple projects remain concurrent. Here Geoinformatics is an assembly of existing projects brought together to meet a funding pull in the form of ITR. In the third narrative Geoinformatics is inspired by bioinformatics and other parallel informational efforts in the sciences. Here the geosciences are on a larger trajectory of scientific change in the face of information technologies.

A Confluence of GEON and Geoinformatics

The first history of Geoinformatics is explicitly described as 'history'. At the Geoinformatics 2006 conference GEON PI Krishna Sinha presented an introductory keynote speech entitled a "Brief History of Geoinformatics". This history enacts a confluence of two trajectories: GEON and Geoinformatics. Here GEON emerges from early Geoinformatic efforts and then comes to be the ideal model for its future.

Sinha's presentation traced the Geoinformatics' lineage back to workshops on Plate Tectonics in 1999. Summary reports for this workshop to the NSF included calls for "creating an initiative that would establish a Geoscience Data Facility." In Sinha's accounts these are the first seeds of GEON. He reports that based on these recommendations, and "follow-ups with NSF," the first proposal for Geoinformatics were funded and then conducted in three workshops during 2000.

These nascent Geoinformatics efforts were quite small, if pressing. Participants received funding from the NSF for three workshops in 2000, with explicit goals to formulate and begin acting upon plans to propose an ITR. The first cobbled together for April 2000 only a month following the above report, the next in May and the third in August. NSF officers and Geoinformatics participants report bureaucratic juggling efforts to fund and organize these meetings: “Thanks for solving all the logical problems associated with the proposed workshops. On behalf of the geologic community, I am delighted that the Foundation is supportive of the community’s need to go digital,” (Email, geoscientist to GEO NSF).

Following on the heels of this meeting, the second Geoinformatics 2000 workshop led to the submission of the ‘Geoinformatics White Paper’. In an interview with an NSF officer I asked of the white paper, “This is the early draft of GEON?”:

> No, this is pre-GEON. But [..] if you look at this outline, that's what GEON is. So GEON became that. But this is Geoinformatics, now to start. They wanted to start this committee on databases. This is sort of one of the first write ups of Geoinformatics. And if you read through this, you can see it's really describing GEON.

In this account GEON is not yet Geoinformatics but it is distinctly difficult to disentangle from it. Many of the initial participants, authors of this white paper, and PIs in GEON’s first and second proposals overlap.

The discursive strategy of the whitepaper follows the formula of the logic of interoperability: characterization of disciplinary difference at the level of data, problematization in the face of scientific advance, and offering a mediated integration as solution:

> The rock record which preserves nearly 4.5 billion years of history has been meticulously gathered through observations over the centuries, and highlight the scientific problems associated with studies of biodiversity and climate change, planetary processes, and the 4-D architecture and evolution of continents. As the complexities of these processes are only recently being recognized through the application of new technologies, it

> is evident that an enormous gain in understanding can be realized only if multidisciplinary data are evaluated numerically, and integrated geospatially through the utilization of Information Technology. (Geoinformatics Whitepaper)

More subtle, but for GEON a crucial detail, the second Geoinformatics workshop participants came to the conclusion of a need for broader partnerships, for greater involvement of the "information technology community" both for the integration efforts but also for a successful ITR proposal:

> Discussions with Peter Arzberger of SDSC, Alan Gaines (NSF) and Geoinformatics Steering Committee resulted in invitation to SDSC to attend third workshop (Sinha Slides Geoinformatics 2006)

Thus the SDSC made its first appearance at Geoinformatics at its third workshop in 2000. In Ch. 2.1 we saw the generation of a particular 'funding push' in the organization of the ITRs. While each directorate of then NSF received a portion of ITR monies, and there was no requirement to collaborate across the Foundation, the computer science directorate (CISE) received almost half of the total annual funding allocations. This heavily encouraged the other directorates to partner with CISE in the cooperative funding of, in particular, large-ITRs. I have argued that this weighted funding push permitted the introduction of computer science agendas into domain science research ventures. At the time, the supercomputer centers (under the PACI program) were supported by CISE.

In attendance at the third Geoinformatics workshop was Sid Karin, then director of the SDSC. It is here that Sinha's account poses the confluence of multiple institutional trajectories: the SDSC and its funding division CISE, Geoinformatics and its impetus in the GEO directorate, and the NSF's current priority area: ITR. Only a few months later in December the first GEON pre-proposal[2] was submitted. Finally, along with several geoscience 'small' and 'group' ITR projects discussed at Geoinformatics, a full proposal was submitted for GEON in 2001.

Sinha's historical account moves directly from the GEON proposal in 2001 to its third All-Hands meeting in 2005[3]. Since its first year GEON has had an 'All-Hands Meeting' as its primary site for internal organizing (see Ch. 4.1 and Appendix B). However, over time this meeting has 'morphed'. In 2005 this meeting became an open conference, primarily focused on GEON but including other informationally oriented projects in the earth sciences. Finally, in 2006 the All-Hands meeting became Geoinformatics: "the GEON All-Hands meeting has morphed into Geoinformatics" (NSF Officer).

The conference was co-hosted by GEON and the United States Geological Survey (USGS), and drew in a much higher number and broader range of earth and information scientists than previous GEON meetings. It was held May 11-12, 2006, at the USGS Headquarters located in Reston, Virginia. Two panel series were run in parallel, always with 'geoinformatic' undertones but leaning towards either CS or geoscience concerns e.g. 'Concurrent Session 1: Natural Hazards' and 'Concurrent Session 2: Web Services and IT tools'. Those present included participants of many of the geo CI projects, such as CHRONOS and SCEC, representatives from the institutions of geoscience, USGS and

---

[2] Large ITRs require a short pre-proposal, and then only with approval will a full proposal be considered. GEON's first pre-proposal was approved.

[3] Between these are several Geoinformatics events and meetings which are not included in his account. Most notably the 'Building the Geoinformatics System' workshop in 2003 – the empirical site of the next section in this chapter – is not mentioned by Sinha. This is perhaps because GEON was not a significant actor at that meeting, although in many senses it was the elephant in the room. See next section.

NASA and of course a generous number of GEON and San Diego Supercomputer Center (SDSC) participants.

This history of Geoinformatics concludes at this event, the point at which this account is presented to a public audience. Thus, this keynote speech, presented at the opening of the conference, frames the history of Geoinformatics, and therefore that entire conference, as a birth, confluence and then equivalence of GEON and Geoinformatics.

In this history Geoinformatics and GEON come to be indistinguishable, but more so they are enacted as inevitably so. Geoinformatics, originally articulated as 'a database for the geosciences,' has over time borne the fruit which is the GEON project. Here the trajectory of Geoinformatics leads to a vision of earth science informational endeavours which must come together as cyberinfrastructure, a single unified platform. In the future posed in this history this has not yet come to pass. However it is the ultimate goal of Geoinformatics, and GEON comes to be the vehicle for making a reality of such a vision.

Geoinformatics as the Confluence of Multiple Existing Trajectories

A second enacted history of Geoinformatics poses it as the confluence of existing informational endeavours within the earth sciences as they were 'pushed' by emerging technology initiatives such as EarthScope and ITR. Here rather than moving towards single unified information infrastructure, Geoinformatics resides at an intersection of existing projects which came together to meet emerging opportunities.

This is a narrative told to me by a Geo NSF officer who closely participated in securing funds for the year 2000 Geoinformatics workshops which Sinha described in his presentation. This officer described his simultaneous involvement in multiple earth science projects with heavy informational components, including projects in cross-directorate wide funding programs.

For example, in 1999 NSF had what was called "an opportunity fund", a cross-directorate program which allowed officers to compete for relatively large amount of money to seed new projects. He describes how NSF program officers, later crucial in funding GEON, were at the time becoming familiar with the emerging mantra's of information integration, in this case tied to the Paleobiological Database[4]:

> [Chuck] wanted to start a paleobiological or a paleontology database. In 1999 NSF still had what was called "an opportunity fund" which was there and allowed program officers to compete for that money if you needed a relatively large amount of money to begin something. When I asked [Chuck] to come up with a request for the opportunity fund, we bounced back and forth on emails. I eventually put in a request for the paleobiological database for the Earth Sciences (NSF Officer)

---

4 Even in colloquial history, this is by no means 'the first' attempt to integrate data within the earth sciences. Rather, when pressed this NSF officer reframed to this endeavor as 'the first' in a new wave of data integration efforts, and recalls long-standing efforts such as Long Term Ecological Research (LTER see also (Baker, Benson et al. 2000; Baker and Bowker 2001; Baker, Jackson et al. 2005)), Unidata, International Geosphere-Biosphere Programme (IGBP see also (Kwa 2005)) or as this officer notes below, Incorporated Research Institutions for Seismology (IRIS):

> Ribes: But you feel that this was the first one in an attempt to create such a project?
>
> Zed: This was the first one in EAR [Earth Science Division of NSF]. Actually, that's not quite true of course. There were databases, many databases that were already in existence. Probably the best known one is the one in IRIS. The Data and Management System, DMS, is located up in Seattle. […]
>
> Ribes: Was this the first attempt to integrate?
>
> Zed: Well, this was the first new one. This was the one I was brought on. IRIS had been around for a while, so this kind of piqued my interest as well. But IRIS was just getting funded year by year by year.

The request was formulated both in the language and concerns of paleobiology, but also the emerging problem set of data and knowledge integration. Again, this is not computer science research on either 'infrastructure' (a word that does not appear) or information integration. Rather, concerns are primarily framed in terms of the domain science's contemporary research goals, and as a problem for paleobiology; below the opportunity poses loss of data and poor archiving as the problem to be solved:

> In the case of the fossil record, there has been 200 years of description of extinct species, as well as the characterization of their paleoecological and geological contexts. However, these data are dispersed throughout the literature. As yet we have no systematic description of the fossil record (NSF Opportunity Fund Request July 19, 1999)

The opportunity fund was NSF wide, spearheaded by a single directorate but necessarily involving partnerships. Potential partners listed in this proposal were the bio-science directorate, the Division of International Programs (IIP) and USGS.

Notably, CISE was not outlined as a potential partner in this early data integration effort. However, drawing on a language which parallels the Knowledge and Distributed Intelligence request for proposals (KDI, see Ch. 2.2.) and prefigures those of CI. The goals of this two million dollar project were specifically directed at data, tools and computational resource development for the paleobiological earth sciences:

> 1) the unification of our paleontological data into electronic database; 2) the development of tools for analyzing those data; and 3) the development of the means for linking these

> data to other relevant biological and geological databases. (NSF Opportunity Fund Request July 19, 1999)[5]

The germ for what comes to be a discursive formula in arguing for information infrastructure is already present: i) over time we have invested human and financial resources in the production of scientific data, ii) this data remains valuable, and could be re-used in addressing the contemporary questions of the field, *but* iii) this data poorly organized to the point where we do not even know what we have, it cannot be worked with in conjunction with other data, and due to poor archiving we are at the risk of its loss; iv) the solution to this is producing and supporting an integrated database, with computational tools to support its use, v) all of this will require substantial community investment and cultural transformation as well as interdisciplinary collaboration.

This general formula of argumentation would come to be one of the central platforms of Geoinformatics, GEON, and as we have already seen, cyberinfrastructure. In making CI a public issue for the geosciences this style of argumentation would be repeated again and again. It is a problematization of extant informational resources coupled to a solution in the form of CI, or more specifically, GEON.

This officer recounts a sense of urgency surrounding the initial year 2000 Geoinformatics meetings, and describes the scrambling at the NSF to support the workshops. Funding itself was cobbled together through a variety of work-arounds:

> It was kind of late in the year, and the first meeting was already scheduled and coming up in April as I recall. So what I did is I pulled – and I don't know whether you want

[5] This project was not funded in this round, however it did receive funding in 2000 under the NSF's Biocomplexity priority area. This is now 'The Paleobiology Database' (httt://paleodb.org). These participant in this data project became one of GEON's first partners in integration.

> to hear it this way, but I kind of threw my weight around a little bit. At AGU we had a little office to run, or to be the secretariat for, the Earth System History Program. So I sent them a request for them to send a request […] as a supplement to their ongoing program. They had their first meeting scheduled for April and this was already in March. And to run a proposal through would be very, very hard. So we had to find an ongoing grant […] I knew I had an ongoing grant here with AGU, *so I simply asked AGU to send me a proposal for a supplement*. All it required was a letter; there was no review, no further review was required.

NSF officers, Geoinformatics organizers and participants alike speak of an urgent need to mobilize in the face of oncoming Earthscope program and the approaching deadline for the immense funding 'push' of large-ITR looming in December. Thus the first of the Geoinformatics workshops focused on the need to "Establish a consortium of universities to develop a community based proposal for the Information Technology (IT) initiative."

The Informatics Gold Rush Makes a Home in Geoscience

The funding of these workshops by the NSF did not occur outside of other informationally oriented endeavours. At the time database federation projects and 'informatics' as a term were itself gaining prominence at the NSF, and in particular at the National Institutes of Health under the rubric 'bioinformatics'. Science reporter Ken Howard dubbed this a 'gold rush' as industry, science funding institutions, information technologists and scientists collected efforts and resources under the rubric of bioinformatics (Howard 2000). Nothing like this has occurred in the earth sciences, but one NSF officer recounts "looking around" and seeing informatics efforts burgeoning in every discipline and feeling that a similar initiative was needed for the earth sciences.

This is an actor's narrative of isomorphism: the means by which the field of geology came to model itself on the activities of its sister disciplines.

Within his own field this officer reports receiving requests for informational sources from multiple disciplinary teams: "all of a sudden information meant something new to me, it used to be 'stuff to know' and it was around 2000 that I started thinking about it as data and computers. We got requests from everyone for information technology monies in geology," He describes the "integrated database in paleobotany" in the previous narrative, and writing the requests for funding this. He also mentions the written reports from Sinha's 1999 Plate Tectonics workshops s funded out of his office claiming that these researchers were under-supported, particularly in terms of information management. He gave me a copy of these reports, which call for a general geologic database effort:

> Scientific breakthroughs in Geological Sciences over the last century have been firmly rooted in integration of data sets across disciplines. Such multidisciplinary thinking gave us the globally unifying theory of Plate Tectonics over forty years ago, and more recently the science of Global Climate Change. In these pioneering breakthroughs, the most significant reason for the achievements can be related to the ability of the community to integrate data […] Future breakthroughs in such integrated efforts will require seamless integration of vast amounts of observations. Clearly the traditional mode of visual assessment of limited databases must be considered to be the first order barrier towards fundamental breakthroughs. […] a general purpose database for geology has yet to be achieved, and is considered by us to be of fundamental importance in real time modeling of multi-dimensional information. (March 2000, Report to the NSF Tectonics Program. Geology and the Digital record: Workshops for developing a national Geological Database)

Just as with the paleobotanical database we can see the enactment of an emerging logic of interoperability in this report: identification of disciplinary heterogeneity in the form of data standards, problematization of this in the face of scientific advance ('visual assessment of limited data'), and posing a solution through integration (here in the form of a 'general purpose database for geology').

This NSF officer recounts a barrage of such criticisms from within his own community, describing this as "a cusp for geology and IT":

> Right about then is when all this business of cyberinfrastructure got started. Of course, they didn't call it that then -- I don't know what it was. But that's when Atkins started putting together all his meetings that led up to the Report.

From 2000 until late 2002 various committees were convened between Atkins' team working on what would come to be the CI report (but at the time was simply an evaluation of the PACI supercomputer centers program relative to scientific disciplines) and domain science communities. GEO NSF officers participated in a variety of these, including environmental, ocean, atmospheric and earth science consultation meetings.

In conjunction with efforts internal to the earth sciences, within the NSF he reports participating in cross-directorate programs linked to other informatic endeavours:

> At the time also, just to give you another insight, I was on the LEE committee - Life in Extreme Environments […] It was also a cross-Directorate committee. And on that committee I kept hearing about bioinformatics - bioinformatics this and bioinformatics that – "a database for all the biosciences". So one day it dawned on me […] So I talked to a bunch of people and got various kinds of ideas on what we would need in Geo as an informatics kind of activity and came up with 'Geoinformatics'. (NSF GEO Officer)

This NSF officer describes working within inter-disciplinary settings, and comes to see biology engaged substantively in efforts for bioinformatics. At the time bioinformatics was experiencing its 'gold rush' of research and funding, and receiving considerable attention within scientific circles and the public (Howard 2000).

This narrative is enacted as the meeting point of already established database efforts in geoscience, a discourse of burgeoning informatic efforts in other fields such as biology, and growing calls by prominent geo-scientists for greater investment in information technology. It is here that this officer sees the birth of Geoinformatics.

--

In this section I have briefly described three laminated narratives of the history of Geoinformatics and the ties of this term to GEON. In the first history GEON emerges from nascent Geoinformatics efforts always already as its future clearinghouse, a 'database for geoscience data'. In the second history Geoinformatics is the confluence of multiple already existing data integration efforts in the earth sciences. It emerges at this intersection as geoscientists organize to meet opportunities for large scale applied information technology funding. Finally, the third narrative poses Geoinformatics as deeply influenced by apparent efforts in scientific disciplines such as biology, as well as what appeared to be a new trajectory for the NSF through CI. Within each vernacular history of Geoinformatics there is also a model for the emergence and purposes of GEON.

3.2 – Enacting the Problematic of Interoperability for Geoinformatics

In this section I focus on a particular Geoinformatics meeting in order to trace with finer granularity the enactment of the personal troubles of individual geoscientists as the public issues for the geosciences. While the Atkins Report has framed a general set of problems for the sciences these must be articulated locally, as the lived experience of earth scientists. Transforming personal problems into public issues is a practical matter; it must be enacted locally and through arguments accessible to participants in the field.

Representing Earth Sciences, Agenda Setting for Geoinformatics

The second occasion for Geoinformatics was held at the SDSC entitled 'Building the Geoinformatics System' (BGS) on May 14-15th of 2003. This meeting was considerably larger than the previous iteration of meetings, and with explicit programmatic goals. Participants included representatives from the NSF, as well as IRIS, USGS, the American Geological Institute (AGI), and some of the newly funded CI projects such as SCEC and CHRONOS. Many workshops attendees already knew, or knew of, each other. If there was any remaining doubt the room was also filled with name tags, place-cards and binders collecting statements of organizational purpose and their roles in IT development.

The diversity of participants, disciplines, and informatics projects contributed to a range of technical approaches for integration. That most of these participants had been selected for their ability to represent large portions of the earth science community contributed to the greatly heated tone of the workshop. Two themes emerge distinctly through discussions at this workshop:

> i) how to approach the question of unified technical ‘system’ for geoscience data, and;
>
> ii) social organization, culture and communication as a key feature in producing this system.

How to frame opinion and experience as ‘community concerns’ was the ongoing negotiation amongst present geo-scientists, and the material to do so was ‘brought home’ (or, made endogenous) by means as varied as citation, anecdotes and generalization.

GEON had been funded over seven months ago. However, few, if any, suggestions were made that GEON should be the site of this unified infrastructure. While the workshop was formally co-organized by GEON lead PI Chaitan Baru there remained a palpable tension in the air as to the role of GEON as an umbrella organization for Geoinformatics. This meeting is notably absent from Sinha’s lay-history of Geoinformatics (ibid.), perhaps because at BGS GEON was clearly still a small player. Only a month past its first all-hands meeting, GEON was not *officially* represented at all (i.e. on the list of attendees), although a handful of PIs from that project were present.

The Atkins report had been published just a few months earlier in February of 2003. Cyberinfrastructure, and its particular vision of the collaboration of domain, information technology and social science, was making its first appearance[6]. It appeared in the title of the workshop and in repeated references and quotes of the Report by

[6] Atkins vision of IT/domain/social collaborations was represented by my presence at this meeting a participant observer diligently taking notes on my laptop and drinking free coffee. I participated in the various ‘activities’ organized at this event, see below. The explicit references to social science in the Atkins Report had sensitized participants to the participation of social science in CI. At this meeing social science’s star was already on the rise, if only as a token. At one point, upon learning my role as ‘GEON social scientist’ a prominent earth scientist and participant of Geoinformatics turned to me, offered his card and said “If you want to see a real project, give me a ring.”

participants. With the publication of the Atkins Report cyberinfrastructure had begun to seem a likely future trajectory for the NSF. The sort of participants invited to this high level meeting were well aware of the shifting winds at the NSF, the ‘funding push’ of ITR and increased efforts to promote multidisciplinary collaborations and the building of infrastructure.

The explicit goal of this conference was to define a trajectory for the development of information technology resources in the earth sciences. This is the ‘Geoinformatics System.’ The agenda, circulated in advance, explicitly asked participants to define the goals of the community, and approaches to system development. Problems were framed with such oppositions as:

“Coordination vs. management”;

“Distributed vs. centralized”; or

“Database vs. federation”[7];

In the talk of the conference itself the loaded term ‘top-down’ and ‘bottom-up’ became the common parlance as shorthands for the agenda items above.

As participants discussed the approaches to a Geoinformatics System it was impossible to disentangle the roles for institutions such as NSF and USGS, organizational concerns and technical approaches. For example, in speaking of a ‘centralized Geoinformatic system’ participants would address models of data storage and distribution (“I don’t think we should try put everything in one place, people like to keep their own data…”) along with discussion of the institution that would orchestrate such an event

[7] The agenda, and extensive programmatic statements are available at http://www.kgs.ku.edu/Geoinfo2/Cyber/ (Accessed July 9, 2006).

(continuing directly from the last quote "…besides you guys at the NSF [waving at NSF officers] don't like to force anyone to do anything with their data"). At stake amongst these participants was more than a dataset or a particular scientific research question, it is the organization of science and its institutional support.

--

In these discussions the NSF (or for that matter USGS, NASA, IRIS, GSA &c.) was rarely discussed as an institution 'out there'. Rather, since all of us could 'point at the NSF' through program managers in the room, it was understood that NSF policy would be heavily informed by the happenings of the conference. The overall tone in the single room of this workshop could be described as earnest. Attendees conveyed a sense that what would come to be referred to as the 'community opinion' was being set here by the many prominent scientists and leaders of the various geoscience research institutions.

In what Taylor has called the autonomous theory of organization (Taylor 1995) actors are actively engaged in making declarative statements in the name of 'their community'. In this model of *organizing* 'the community' or 'an organization' are ontological entities only to the extent that individuals can effectively mobilize them in situ, as their representatives or spokespersons[8]. The hope in this particular conference was of effecting the outcome 'geoscience community opinion':

> Marshalling of collective energies creates an agency whose power supersedes the individual; such an "agency," however, has no capacity to act on its own and can only find expression in the voice of some individual; when an individual is recognized by the community as giving legitimate expression to the collective agency, then that person's voice has a force that no single individual can

[8] Bruno Latour has described these as the spokesperson's of the sciences (1987).

> aspire to – it is literally 'power-ful' […] the recognition of legitimacy depends, however, on the existence of a shared opinion, or view of the world […] that is generally understood to be the basis for action (Taylor and Van Every 2000:87 ft.17)

Here NSF officers are one sort of community representatives. Not in a simple sense of a hierarchical relation, since "authenticity" (Taylor and Van Every 2000:87 ft.17) must be preserved, but in their special capacity to stand in for the community[9]. Even more so than as 'representatives' they have come to this meeting in order to understand 'community opinion' for plans to construct a Geoinformatic System. 'Authenticity' for NSF officers is not firmly grasped in an iron glove. Rather, they express a sense of responsibility to that community[10].

How is the opinion of the community to be known? It is not a visible entity; the multiplicity of American geo-scientists are not at the conference; the community does not reside in the rooms of the SDSC. Nor was community opinion determined by casting votes, through carefully sampled surveys or other 'formal' modes of representation. Rather, it was established through talk, in the utterances of the assembled representatives of the geoscience community. In this room, with the assembled potentates of geoscience present, the institutions of earth science negotiated, argued, orated and debated with an earnest hope to speak 'community opinion'. In doing so they enact the logic of interoperability.

---

[9] Taylor and Van Every (Taylor and Cooren 1997; Taylor and Van Every 2000) trace this notion of representation back to Durkheim (1915).

[10] In an interview with an NSF officer, not at this conference: "Yes, we [officers] are here at the NSF. Yes we sit in the tower. But no, we are not here to decide top-down about the system. We're here to listen, and then make it happen".

Articulating Troubles as Issues for the Geosciences

We should not be misled by the eminence of participants at this workshop. In representing the various institutions of science individual participants are still engaged in the utterly ordinary, mundane activity of speaking; or as Garfinkel liked to say, 'immortal ordinary society'. This is not intended to deflate the significance of the meeting, or minimize the eminence of participants. It is to say that *articulating* the problematic of interoperability for geoscience is work subject to the banalities of everyday conversation and interaction. The organizing of the workshop itself and the stated goals remained sufficiently vague as to permit a very broad framing of topic. At Building the Geoinformatics System workshop narrowing to a specific set of troubles and issues was framed within the organizing of the workshop itself. How to communicate, how 'to speak for the community', was learned as the workshop itself was enacted. What are the 'appropriate issues for geoscience' is at stake in these discussions. The question in this section is not how is cyberinfrastructure relevant to geoscience, but how is it made to be so.

The Building the Geoinformatics System workshop was conducted in a single room with chairs and tables arranged in a large circle. This single assembly would divide in four break-out groups with concurrent sessions, followed by summary reports back to the entire group.

Volumes could be written on the process of this meeting alone. A professional, non-academic, 'facilitator' was hired to organize and lead discussion. Drawing on the language of management and public relations the style of the facilitator raised many an eyebrow in this academic crowd. For example, the assignment to a break-out groups was chosen at random by letters (A,B,C,D) written on the back of each participant's agenda. Participants were encouraged to 'get to know each other' by providing a brief introductory biography at the beginning of each session, but were not able to chose, by interest or experience, the break-out session topic they would attend[11].

Break out groups would convene, discuss, and then provide summaries to the entire workshop. Such activity in organizing was not initially successful. Of the four iterations of break out groups, the first were distinctly incoherent and generated little topical discussion from the group as a whole. Summary reports from these break-out groups were very general and often grand. However, following these the discussion would be opened to the assembled workshop as a whole. Through group discussion participants are able to articulate more specific group concerns. Thus the general trajectory of the discussion was from general to specific. Organizing is a learned skill, and in the face of new organizational forms (more typical of business settings) participants of this meeting required a few iterations before a routine was established.

For example, one breakout group was assigned the task of defining 'The goals of Geoinformatics.' I was not privy to the detailed discussion within the breakout group. Following an allotted discussion time, the group returned and reported a 'summary of findings.' Below I will describe the summary, followed by an elaboration of the discussion as it unfolded following this.

The summary to the assembly began sweepingly: "The goal of geo-informatics should be a data-model of the earth, so those who participate are those that contribute to the job." Sketching a large triangle on an overhead projector this geoscientist noted that it

[11] I have not seen this model of workshop again.

stood for 'depths of knowledge'' and wrote in large print "Large-scale Modeling". He then stated that "users should be able to enter at any level of the system". This summary met with baffled glances from the assembled participants. Meeting with little group response and having used only one minute of the allotted five minute presentation time, the discussant quickly conferred with the break-out group. The triangle was divided into a three storied pyramid identifying the components necessary for data modeling: 'people,' 'knowledge', 'tools'. Still, the triangle was insufficient to spur discussion. Finally a participant NSF officer ('Al'), not from the break-out group, captured the conversational turn and referred to the projected overhead where 'large-scale modelling' had been written:

> Al: We've had some serious difficulties sustaining large-scale modelling efforts. Often we [the NSF] fund these, and they go on for a while, but then when the funding runs out, so does the project.
>
> Bob: Such as?
>
> Al: Uhm, well… Digital Earth wasn't one of ours, but it's a good example.
>
> Bob: Why was Digital Earth a failure?
>
> Chuck: They were never given any resources!
>
> Dick: They had no *drivers*, what were their *goals* in terms of *science*?
>
> Evelyn: The people who worked on it *had* no science questions.
>
> Fred: Mostly no funding… But everything, social, economic … it was a project for the global and continuous archiving of knowledge … *what was it?!* They never really had a plan; it died a slow death over three or four years.

The conversation never reached the scale of the initial task to the break-out group 'the goal of Geoinformatics.' The initial summary was not found to be particularly insightful for the rest of the group, or at least it did not incite immediate dialogue. However, as the interaction continued an increasingly lively discussion evolved over the question of the success and failure in large-scale informational endeavours.

In the excerpt above the Digital Earth project became an exemplar as participants characterized causes for its failure: funding, lack of drivers, goals or science questions, or very generally ‘social and economic’ and a lack of coherence. We will see these particular causal explanations again and again, at all scales of activity in GEON and CI[12].

In this discussion an endogenous comparisons (see Ch. 1.1) is generated between this Geoinformatic endeavour and other large science projects such as the Human Genome Project (HGP) (Weller 1996) and Digital Earth (DE). The material for what is to be considered in building such a large-scale system is generated endogenously through such on-the-fly evaluations and comparisons. The discussion sparked by the label ‘large-scale modelling’ continued for approximately fifteen minutes, despite protests from the time conscious facilitator.

Working from these discussions participants distilled and then wrote “*variables”* to be considered in the construction of a Geoinformatics system: on the overhead they drew up lists of these key features, reproduced here verbatim:

funding
drivers
organization
collaboration
communication

---

[12] The notion of ‘science driver’ is particualry salient in GEON, as is discussed more extensively in Ch. 4.3. ‘Science questions’ has already been discussed as GEON came to be defined as twice basic research (Ch. 2.3).

science questions

social

This is what Weick has called a 'cause map.' In the enactment of an environment, actors must devise methods for understanding the 'external world' and formulate models of its operation: "What we need to be sensitive to then […] is how raw data are utilized (bracketed and separated), labelled, and then transformed into a "network of causal sequences" (p.277)" (Weick 1977; Taylor and Van Every 2000 quoting). These are not the logically derived 'causes' of philosophy, or experimentally isolated 'causes' of physics[13], rather they are endogenously enacted causes in commonsensical talk. These variables, as the participants called them, were outlined for the purpose of collectively thinking through building the Geoinformatics system.

--

By the second day the participants had become more familiar with the adopted format for the meeting. Breakout groups became more focused and summaries more cogent. I will follow a single iteration as, within the breakout group, personal problems of earth scientists were articulated as issues for earth science. As a participant[14] in this workshop I was assigned to, for example, a breakout group entitled "Our opportunities and the emerging crisis." To guide our discussion we were given only two hand-written guiding notes: "What should Geoinformatic technologies look like?" and "In 1980 we had 80,000 students, now we have 16 thousand. How can these revitalize geoscience?"[15]

---

[13] Although we should not distinguish these a priori. Their distinction too comes in their practical achievement (Garfinkel, Lynch et al. 1981).

[14] Though perhaps not an eminent one.

[15] This claim was neither supported by any particular reference, nor challenged by any of the break-out group participants. While this note did not spur any particular discussion, there is a general sense amongst the geoscientists I have spoke with that the field has failed to capture the attention of contemporary

Despite the alarmist topic, participants responded with quiet sobriety. Alarmist topics and grand questions assigned to the breakout groups had become the norm and, ignoring the assigned topic participants had set themselves to tackling issues they considered to be of relevance. This break-out group's discussion remained considerably calmer than the charged atmosphere of the general discussion. What the note had framed as an 'emerging crisis' was quickly redirected into discussion of:

- i- concerns about the redundancy and overlap of existing technology initiates in the earth sciences: "are we producing the same tools, or even the same data, over and over again, wasting our money?" and;
- ii- data sharing, or as one participant initially formulated the topic: "the question of sharing".

The first topic was merely mentioned, raised in discussion, but received no further attention. The second, continued for the remaining time of the breakout. Speaking of the data he had collected in a small ITR grant, a geoscientist in the breakout group noted:

> As a PI [principal investigator] I am necessitated to include my data in a system that is somewhat interoperable, public, and accessible. But the problem was of course that everybody and his cousin had a database. I didn't have any place that was my own database, so I would have loved to contribute my data to a standing database, but there was no way I could begin a database on my own.

Another participant requested clarification of "why are you *necessitated* to include data?" and the speaker noted that this is "because of this freedom of information business". This

---

students. I have also heard the occasional suggestion that this is because geoscience has fallen behind relative to the trajectory of technical advance.

is a reference to the US Freedom of Information Act (FOI). A discussion ensued formulating the relevance of this utterance to data, and scientific practice more generally.

The US Freedom of Information Act secures citizen access the records of American government. A recent amendment places data produced with public funds under its jurisdiction[16]. The breadth of the amendment remains untested in court, but it certainly includes databases and may also include scientific notebooks and possibly even emails of researchers. The amendment resulted in significant controversy within scientific circles (McGinley 1999). Claims have been made that, rather than as a matter of public access to government, the amendment will facilitate industry attempts to undermine scientific research; for example, by making available to scrutiny the details – data, notebooks &c – of environmental and health studies which support climate change or link cigarette smoke to cancer (Hilts 1999).

The participant who raised the topic of the FOI first mentioned these controversies, and then noted that as the dust has begun to settle around the legislation, scientists have come to ask questions as in his extract above. 'How do we preserve our data, who is going to pay for it, and where should we put it?' Computer scientists in GEON have noted how, as the amendment has trickled down into the institutions of science, it has added significant impetus for data archiving and integration efforts by making scientists themselves responsible for preservation. The discussion which ensued quickly traversed through the entire range of topics which in social informatics, science studies and information science have come to be known as the 'problem of data-sharing':

---

[16] This is, in fact, a one-sentence amendment pushed through as part of a very large Omnibus Appropriations Act for FY1999, sponsored by Alabama Senator Richard Shelby: 'Public Law 105-277'.

> Fred: There is a great fear of making data public though! The first thing you have to think about is people's careers. Databases can either be protected, semipublic or public. And I've found that a lot of suspicions disappeared after people saw how many people were worried about their own careers.
>
> George: This is about reward structure, we need to make some kind of rewards structure. Like, if people that are digital are able to publish their data much faster, and then everyone sees that they will be quick to take it up
>
> Fred: There is some cultural change involved in making something a reward. Will people actually acknowledge data as a reward? Not just as making a dataset available, that's never done anything for anyone. Data contributions should be considered something like publication.

As an analyst I can parse this conversation into a set of concerns about data sharing. This is a well formulated problem in academic literatures. Sharing data (and mostly, not sharing) has been studied from the perspective of its effect on career trajectories (Ceci and Walker 1983; Campbell, Clarridge et al. 20002i); psychology and values of personal re-use of data before publication (Sterling 1988); the reward structure for data publication and making data accessible (Van House, Butler et al. 1998); failures to acknowledge or cite shared-data use (Sieber and Trumbo 1995); and, 'cultural' transformations in accreditation of citations (Ceci 1988).

In the excerpt above and the extended discussion of data sharing in my breakout group, participants generated an in situ cause map of the problem of data sharing. Mirroring the academic literature they explicitly used the terms 'reward structure,' 'careers,' 'cultural change', 'private' and 'public'. Of course, the discussion was informal and did not draw explicitly from scholarly articles. There were no references in this discussion. Many of the participants are long-time advocates for data-sharing and may be familiar with the literatures, this said, these problems have become part of the vernacular of information efforts in geology. How to share data, and the barriers to this, are regularly discussed. Over the years in GEON I have heard such discussions at every major meeting. The problematic of data sharing is enacted regularly at such events.

--

The ‘problems of data sharing’ writ large must be introduced to Geoinformatics through discussion and by connections to the particular disciplinary matrices of the geosciences. While the Atkins Report or articles in prominent journals such as *Science* feature problematization of data sharing, this does not simply *diffuse* out to the sciences. Rather, they must be locally enacted and made an endogenous public issue (Mills 1959; Millerand and Ribes forthcoming). A private trouble, such as the experience of being unable to contribute data to an existing archive or having the resources to initiate ones own, must be rendered as public issues, general beyond personal experience. In the case above the legal and policy decisions within congress are made relevant, in talk, to geoscientific practice and then linked to a more general set of laminated concerns: career trajectories and rewards, citation and so on.

--

Continuing in the discussion participants related various anecdotes which enacted the “theme of cultural change,” as one participant put it. The vignette below captures a complex endogenous cause map that draws on the topics raised thus far in discussion; why data sharing is difficult, and a model for transformation which includes cultural and community change, the psychology of data-sharing, scientific practice and the particularities of data structure. In less than a minute of speaking the social, the technical, psychology, and practice are traversed and linked through narrative.

The vignette is an endogenous comparison, generated in situ: the speaker is comparing the cultural dynamics of his own field, paleontology, with seismology. Seismology is characterized as comfortable and familiar with data-sharing, while paleontology is problematized as resistant, and lacking in vision. Contained within the story are also causal explanations for how each field has come to have a relationship to data-sharing, and an explicit theory for engendering change:

> He comes from this field of science which is used to community data-sharing and focused investments in a single place to benefit everybody else. I think he's taken advantage of that. Their psychology changed in the mid-'80s, that group of scientists – seismologists, when they went to a single facility [SCEC]. […] *So it takes community thinking to change before the cyberinfrastructure can change that community*. The paleontologists are all used to looking at their fossils in their laboratory with their microscope, and they're happy to do that. But when I [a paleontologist] come along and say I want to develop this cyberinfrastructure for paleontology, they go, "wow. If we fund you, we can't fund 20 small grants like me. Plus, I don't know that I want to put my data into your database. That's time consuming. How am I going to get money to do that?" It's a community mind-set, plus the data are very different. Seismology data are simple compared to some other types of data. So the challenge to the earth scientists is this huge range of data types, plus psychological issues within those communities. (emphasis added)

The range of explanatory resources brought to bear in this vignette is quite broad. To understand how the problematic of data sharing is made an endogenous public issue for geoscience we not go no further than this explicitly articulated cause map, enacted in everyday conversational talk: Why is seismology accustomed to data-sharing? Twenty-five years ago they ‘went to a single facility’’; Why are paleontologists not so accustomed? They spend all their time conducting individual lab work; How do paleontologists resist the construction of cyberinfrastructure? They see the investment as best spent on individual basic research projects; Should CI be imposed upon them, ‘if we build it will they come’? No, the community mindset must be changed, only then will CI change them; Is this all about psychology and community mindset then? No, the data of seismology is more simple than the data of paleontology, and their respective institutional arrangements also play a role.

This cause map brings together social, technical and institutional explanations; they are intertwined effortlessly in such casual banter. They are not based on empirical research in the sense imparted by social science, nor would not stand up to any sustained inspection; but this has not been the framing of the breakout groups. Rather, the purpose was initially to discuss ‘the challenges and possibilities for geoscience,’ and as the conversation unfolded it was become focused on articulating a coherent set of understanding of the problematic of data-sharing in the geosciences.

Mills argued that the role of a sociologist was to produce sophisticated linkages between troubles and issues: “Do not allow public issues as they are officially formulated, or troubles as they are privately felt, to determine the problems that you take up for study,” (Mills 1959:226). He argued that the complex formulation of a problem requires understanding biography and history, or in other words, that the personal narratives of troubles should be interwoven with the abstracted generalizations of public issues. A national issue of unemployment cannot be addressed without grasping the particular mechanisms and experiences of individual troubles. For Mills a sophisticated modelling of a problem draws connective strings between the troubles of individuals, thus transforming them into an issue.

In this section, instead of intervening ‘as a sociologist’ by transforming troubles into issues, as Mills suggests, I followed the work of actors as they have generated these linkages themselves. In the breakout group stories become a vehicle for linking personal troubles with acknowledged public issues; public issues were articulated through accounts of personal troubles.

The final minutes of the discussion in the breakout group continued as rounds of personal stories, or accounts of the troubles of close colleagues. For example, stories followed of a geoscientist who had failed to achieve tenure at MIT despite having dedicated enormous time to creating a valuable public data repository; how in New Zealand an excellent fossil databases has been maintained but two-thirds of this are inaccessible abroad because “nationalism” prevents sharing to an “international community of scholars”; a “top-notch” geographic information systems (GIS) scholar who could secure a job in a geology department because his degree was in computer science.

In such discussions it is not necessary to point to scholarly studies of data-sharing or cultural transformation. Instead the above claims rely on a ‘common sense’ logic of familiarity and shared experience. To the extent that the narratives make references to familiar individuals, recognized institutions, accepted lay-histories of collaboration or cases of scientific knowledge production, the utterances of participants hold true in the discussion and serve as substantive materials in supporting an enactment of public issues.

It should be noted that while there was a professional sociologist in this group, who had identified himself during the introductory biography, the terrain of these discussions was not deferred to me. Cultural transformation, psychological makeup of the community, reward structures in science, institutional history in the geosciences are not taken to be the unique competence of a social scientists. These are features known through experience, learned as a feature of membership (Lave and Wegner 1991). Mills claims that an intimate knowledge of biography and history is needed in order to create sophisticated links between private troubles and public issues; the two must be made to cohere, the lived experience of troubles must be made to align to the abstracted formulation of a 'community issue'. It is just such a knowledge of biography and history that these eminent geoscientists were able to articulate together with a problematization of them as 'issue'. Members in this discussion felt a practical competence to evaluate the data-sharing difficulties within their field based on their long-term familiarity and experience *in* that field. Competence here was defined as the *unique adequacy requirements* of an experienced geo-scientific researcher and/or participation in an institutional field of geoscience and/or first hand experience with data sharing. Broad as the requirements were, I had none of these (Garfinkel and Wieder 1992). It was by these criteria that abstract concerns of data-sharing are enacted for geoscience, and made a public issue for Geoinformatics.

Enacting the Problematic of Interoperability as an Issue for Geoinformatics

The Building the Geoinformatics System workshop as a whole was not organized as a storytelling venue. The breakout group I traced in detail above organized a moment in which stories facilitated a relatively detailed articulation of troubles and their enactment as issues. A further step remained in translating private troubles of geoscientists to public issues for geoscience: the detailed talk of the breakout groups was coupled with reporting back to the entire assembly.

In summarizing the results of this break-out team, the format for these summaries did not permit recounting detailed narratives. As we saw with the first summary I described ("so those who participate are those that contribute to the job") translating the discussion in the breakout groups is a form of cognitive work with varying success. By the second day of the workshop, participants had collectively developed a loose set of conventions for the summaries, and a competence in communicating breakout group "findings".

The representative from the breakout group in which I participated formulated summaries of the discussion into three themes: 'redundancy and overlap', 'cultural change' and 'reward structure'. These summary encapsulations are the methods by which the issues of CI came to be linked to the troubles of geoscience. While the breakout groups were conversational, and informal, summaries to the group came to be presented in a few minutes (ostensively five), and in bullet point format. Here detailed linkages between personal troubles and public issues in the form of narrative stories had to be foregone in favour of explicit programmatic statements. In these summaries these come to be formulated almost the level of generality found in the Atkins Report, with only a grazing contact to geoscience. I will briefly treat, in turn, the three themes summarized from the discussion above.

“Redundancy and overlap” came to refer to a concern that without some “top down” supervision information technologies may be developed and applied again and again in the earth sciences – wasting valuable resources. As an “exemplar” he referred to SCEC, the seismology center understood to have changed the field in the 1980’s. Close top down supervision, by the NSF for example, could ensure that the funds, time and resources of earth science would not be squandered in the repetitive development of ‘one off’ project specific information technology solutions.

Similarly redundancy and overlap are laminated concerns within CI. From the Atkins Report there are “real dangers of disappointing results and wasted investment for a variety of reasons including […] excessively redundant activities between science fields or between science fields and industry,” (Atkins 2003:4). Avoiding redundancy is one of the primary stated reasons for building the national cyberinfrastructure. Resources can be saved if specific disciplinary needs, e.g. in computation or storage, are channelled through such a general informational resource.

“Cultural change” came to refer to the entrenched geoscience communities and the means by which to foster change within these, to “make data sharing a staple of geology’s culture”. Here culture is cast as persistent, with a momentum to be problematized relative to technological change, data sharing and goals of building community resources. It is the culture of, for example, paleontology which hinders the uptake of novel techniques for data sharing. Similarly, the success of cultural change within seismology must be explained relative to the ‘simpler’ configuration of its data, and institutional transformations in its past.

Similarly, cultural change is a laminated concern within CI. The Atkins Report warns of the possibility of underestimating the difficultly of technology uptake, the "lack of appreciation of social/cultural barriers," (Atkins 2003:4). The participation of social scientists is framed as one solution, but the general framework in the report is of mediation: maintaining diversity across the disciplines of science while translating across these at the informational layer (see Ch. 5). This particular model was not articulated by participants at Geoinformatics 2003: culture remained the competency of those within the field, and solutions to the problem of culture were primarily addressed at the level of reward structure. However, articulations within Geoinformatics have begun to resonate with those in CI; culture, 'the social' and social science are potentially relevant actors in the future trajectories of Geoinformatics.

"Reward structure" came to refer to the means by which scientific work is evaluated, and the consequences of such structure on participant's engagement with, for example, data sharing[17]. This topic generated a great deal of commentary from the participants at this workshop. How can writing metadata be made to count towards tenure? Will graduate students that work on such topics be considered geologists? Speaking in the name of USGS one participant noted:

> I think we all know there is a lot of work that goes on in our organizations that has to be done by a scientist but never gets counted as science. And I think we all know that the people who do that work don't get any kudos. Actually they don't get very far at all.

To the extent that reward structures and career trajectories do not reflect the current activities of practicing scientists and that such tasks as metadata writing cannot be

[17] Vernacular uses of 'reward structure' is discussed more thoroughly in Ch. 4.2.

delegated to an administrative technical staff, then extant structures and trajectories can be problematized relative to the progress of science. They become public issues for the advance of geoscience.

Similarly, reward structure is a laminated concern in CI. The Atkins Report is acutely focused on means for all parties to conduct basic research while also building infrastructure. In Ch. 2 we have already seen infrastructure as emerging at the intersection of two trajectories of basic research. ITR provided an excellent occasion to enact such a vision: any proposal for infrastructure building in this funding line had to demonstrate components of 'high risk,' 'non-incremental' basic research. Along with domain and computer scientists, social scientists too are cast as 'interested' in basic research: "Building, operating, and using advanced cyberinfrastructure must be done in a systemic context that exploits mutual self-interest and synergy among computer and information, and social science research communities who see it as an *object of research,"* (Atkins 2003:7, emphasis in original). The model of CI building in which participation is 'mutually self interested' is consistent across the scales, described as 'twice basic research and infrastructure to boot' in Ch. 2, the two-tier approach in Ch. 4 and at a finer granularity in Ch. 5 as the routine for ontology building. Both in the writings of the Report and in this Geoinformatics workshop scientists are cast as rational actors seeking to maximize, in this case, science findings and their careers.

--

At this Geoinformatics workshop the Atkins report was occasionally referenced and the term 'cyberinfrastructure' made intermittent appearances. Each mention of the concerns in CI had to be endogenously enacted by situating such concerns relative to geoscience practice, data, organizations or institutions. What is the community opinion about how the NSF should organize its funding of Geoinformatics? Geoscience education is losing students, how can IT make us more exciting? Our archive is in peril, how do we preserve our heterogeneous datasets?

In framing the discussions in 'breakout groups' issues were articulated relative to first hand experience and personal *troubles* of participants within the field. In framing the discussion as 'summary reports' participants articulated geoscience concerns relative to abstracted *issues* at the scale of institutional action. In both cases a more general body of concerns which I have called the problematic of interoperability were enacted as issues for Geoinformatics, this is the generation a 'demand', 'need' for cyberinfrastructure within the earth sciences. It is only in such a local enactment, in the language and concerns of earth science, that we can speak of a pull to cyberinfrastructure.

3.3 – When Push comes to Pull: A Not-Quite-Yet Inevitable GEON

GEON had a very small profile at Building the Geoinformatics System. In November 2002 the GEON large-ITR had been awarded. While this generated considerable discussion from within the earth science community it did not yet mean that GEON would be the informational 'platform' for the earth sciences and by no means did it make GEON synonymous with the Geoinformatics system. Three years after its funding, coming full circle from proposal to participant, in 2005 GEON was able to visibly position itself as a central actor relative to each Geoinformatics endeavour. By 2006 the GEON All-Hands meeting had become an annual Geoinformatics meeting. In the intervening years between its initial funding and Geoinformatics 2006 GEON PIs had been busy making headway in the forums of earth science.

Between the 2003 and 2006 Geoinformatics meetings, both computer and earth scientists in GEON were building various kinds of *couplings* with institutions, databases and individual researchers in the earth sciences (Orton and Weick 1990). These are the practical processes for linking GEON to those repositories holding the data of the geoscience community. Couplings range from exchanges of technical strategy, promises for future data sharing and in some cases registration of data to GEON ontologies.

It is a laminated phrase that ‘GEON has no data.’ It is not an instrumentation project producing data about the earth. Rather, it is envisioned as a ‘clearinghouse’ of the various extant data producing institutions and repositories in geology. With each coupling GEON adds partner institutions and domain datasets to its inventory of accomplishments. The outcome is an extended listing of GEON participants, greater visibility within the ‘geoscience community’ and a framing of GEON as a forthcoming institution in its own right. With the emergence of a new line of funding for Geoinformatics, GEON come closer to becoming the informational platform for the earth sciences. Below I treat, in turn, GEON’s loosely coupled organizational relations, the ambivalence of its public face in view of this growth, and the emergence of a new funding line for Geoinformatics.

Loose Couplings as the Growth of GEON

In its 2001 proposal GEON listed 15 participating institutions. These included the SDSC and USGS, DLESE (Digital Library for Earth System Education) along with the geoscience PI universities[18]. In 2006 GEON listed 22 additional 'affiliated researchers', and 3 new 'agency partners'. These have grown to include NASA, a 'computing node' at Berkeley's Lawrence Livermore Labs, the Kansas Geological Survey and the Canadian Geological Survey. These links and partnerships were achieved primarily through local site visits, by having representatives come to the SDSC and attend local workgroup meetings to discuss shared technical futures, or through GEON visits to their institutions. For example three GEON PIs visited NASA's Goddard Center in the Fall of 2003, establishing possibilities for collaboration, data sharing and shared representational conventions.

GEON PIs have also been conducting negotiations outside of the US. This effort is often referred to as iGEON. During a summary presentation of GEON's accomplishments to the NSF during a 2004 site visit, international connections were reported with Canada ("WMS Server at GSC, Vancouver, BC"), China ("Computational Geodynamics Lab will host a GEON PoP node for iGEON in China"), and Russia ("Discussion with scientists from Russian Academy on data integration and use of Grid computing for geodynamics codes"), as well as Australia's sister model for an earth science information infrastructure: AEON , the "Earth and Ocean Network". In 2005 an iGEON workshop was held in India, and another is planned for 2006 in Beijing.

These relationships can be characterized primarily as *loose couplings* (Weick 1976). Rather than the standardized process and technical interconnections across organizational boundaries, loose couplings come in the form of alliances, promises of

---

[18] Arizona State University,; Bryn Mawr College; DLESE (Digital Library for Earth System Education); Energy and Geoscience Institute (EGI), University of Utah; Penn State University; Rice University; San Diego State University; UNAVCO, Inc.; University of Arizona; University of Idaho; University of Missouri; University of Texas at El Paso; University of Utah; Virginia Tech.

collaboration and gestures of good will. The importance of such ‘weak ties’ cannot be overemphasized (Granovetter 1973; Granovetter 1983). In GEON they have simultaneously serve to mark a growth of the formal organization and enabled a focused allocation of resources.

In enactment theory ‘loose coupling’ suggests a disjuncture between technical interdependencies, ties across organizations and formalized identity. They permit actors to collectively organize across organizations while maintaining local adaptations, decisional autonomy and a distinct identity. In GEON such couplings have come as exchanges of scholarly papers, *demonstrations* of technology, *promises* of future collaboration and data exchange, *possibilities* for establishing ‘GEON nodes’ (see Ch. 4.1), and *plans* for registration of institutional data repositories.

These are not the ‘deep technical ties’ of the functional data interoperability which some claim as the explicit mandate of GEON: “People ask me when are you going to know that GEON is a success? It is a success when Joe Blow at X-university can sit *at* the university, and he can use GEON tools that he doesn’t know the inner workings of, and can get his project done” (Leonard Johnson, NSF Program Manager for GEON at Keynote Speech, All-Hands 2005). The investment in time for the development of integration technologies such as ontologies or web services is quite extensive, spanning years, and involving technical developers, training for domain participants, and outreach to a broader community (see Ch. 5). Few of GEON’s partnerships have resulted in exchanges of data, even fewer of these have resulted in accessible interoperable data at ‘x-university for Joe Blow’. However, all of GEON’s partnerships do appear on the website, in publications, and in the formal annual reports to the NSF.

Loose couplings have often been used in organizational literature to imply a 'light critical tone', suggesting a veiled 'decoupling' rather than a loose one. For example, Meyer and Rowan have argue that organizations may develop "myths" which make it appear as though an organization is complying with law, policy or 'cultural norms' while their practical activities remain unaffected (Meyer and Rowan 1977). There is certainly a set of critics of GEON that adopt this interpretation of its various partnerships (see next subsection). However, *within* GEON building partnerships and promises of data exchange are understood as a long-term developmental strategy which still meets today's requirements for marked progress.

By linking to the institutions of geoscience GEON is both 'growing' *and* securing data for a future (planned) point of technical development. It is in this manner that the quick growth of partnerships are articulated by GEON actors. As one computer scientists noted at a weekly workgroup meeting just following a visit to NASA: "We start setting the groundwork now with NASA, so that when we're ready we can just get the data."

While GEON's funds are nominally large, particularly from the perspective of an earth science community accustomed to smaller projects, within GEON there has been a growing sense of strained financial resources, and limited human resources. Thus a rationale for loose couplings is that by securing promises today these 'new partners' can serve as surrogates to the growth of GEON today and begin negotiations for a time when 'as an infrastructure' GEON can support data integration efforts and eventually access for its users. Embassies must be established across formal organizational boundaries to prepare a pathway for future relations of trade.

Thus it should not be misconstrued that increasing GEON's partnerships in the form of loose couplings is singularly motivated by 'Machiavellian' expansionist interests. Partnerships with geological institutions are seen to be a vehicle for GEON in its everyday goals for federating data and addressing primary 'science questions'. The data of geoscience is in the hands of individuals, research teams and institutions. In order to secure access to this data, GEON must foster partnerships:

> The idea behind the visit is to obtain access to data beyond maps. For many of the GEON goals, real analytical databases need to be accessed; I want to accelerate the process. Rock, Stratigraphy, paleontology, ore deposits, geochronology, geochemistry, structure (and maybe even geophysics) are the kinds of information we need, and I would like to encourage the USGS to share whatever they can at all three levels of GEON access. Similar things have to be done with NASA, so there is a lot of work ahead of us, beyond just gathering maps for GEON. I am also hoping that […] we can gather all the geologic state maps for the eastern US (even though many of the State Surveys are not wanting to share) and have them be registered through GEON. Its should be a very useful trip. (Geo PI emails the ICC[19], March 13, 2005)

Outreach to the community is motivated by goals of integration, and addressing 'science questions.' Outreach often comes in the form of *promises* to register data with GEON. These relationships of promise later come to appear as 'partnerships' in GEON's repertoire, linked on the web portal, and are listed as accomplishments in reports to the NSF. They stand as markers of growth in public forums.

Equally significant, these partnerships come to be mobilized in generating a "community acceptance" of integration efforts for ontologies and other technologies of

---

[19] Internal Coordinating Committee, see Ch. 4.1 or Appendix B.

interoperability. In developing such technologies the geo PIs have come to see themselves as insufficient to stand in for the details of technical geologic knowledge. Securing endorsements can put some institutional weight behind GEON's knowledge representation efforts (see 'community outreach' in Ch. 5.2). Even such technically minded partnerships are loose couplings. In its first three years GEON's ontologies were not prepared for large scale registration efforts and few services were available.

An institutional endorsement, by participation in an ontology workshop, by including links to development efforts on a website, does not equate to a full exchange of data or reliance on a service. A promise of future registration is not equivalent to an enactment of that registration, however, such agreements to a future course of action can serve to plan the activities of tomorrow, and serve as surrogates of success today.

Ambivalence and the Public Face of GEON

For those outside the project looking in, the growth of GEON as loose couplings has been received with some ambivalence. The modes of evaluation for GEON and its success rely on competing models of 'what is GEON' or even 'what is cyberinfrastructure'. For those with ideas of GEON as a platform for the earth sciences the array of partnerships can mark growth. For those looking for 'deliverables' in the form of science findings or functional tools the same institutional interconnections can mark a misdirection of efforts. To the extent that GEON is enacted as an umbrella infrastructure project with partnership and community building goals then loose couplings are successful growth. But to the extent that GEON is enacted as a technological project with clearly defined functionalities a partnership on paper is merely 'myth' – loose couplings can signal misdirection, a lack of focus or appear actively disingenuous. These definitions of CI and of characterizing the needs of the earth science community are a site of negotiation in the enactment of GEON itself.

--

While GEON, as a large ITR project, received considerable attention from informationally oriented earth scientists or the upper administrative scientists within the institutions, in its early years and even today the 'average practicing scientist' may not be particularly cognizant. For example, although GEON was not formally invited to Geoinformatics 2003, participants of that workshop were well aware of the project and evaluative discussion were prominent; meanwhile at the Geological Society of America (GSA) in the same year, my attempts at casual discussion with geoscientists was usually met with complete unfamiliarity.

GEON participants have engaged in various outreach projects to increase project visibility amongst earth scientists. For instance, such efforts have included booths and posters at the major geoscience conferences of geoscience such as the Geological Society of America (GSA) and the American Geophysical Union (AGU). These events served to tie together GEON, Geoinformatics and cyberinfrastructure in an earth science community setting.

At GSA 2003 the outcomes of GEON research and integration tools were prominently paraded at the Pardee Keynote Symposium: "Geoinformatics and the Role of Cyberinfrastructure in Geosciences Research". As with most professional scholarly bodies the keynote symposiums at the GSA are moments in which focussed topical research is set aside to convene as academics under a common banner: geoscientists. That such an eminent venue would be dedicated to Geoinformatics, with prominent representatives of the NSF (such as assistant director for the geosciences Margaret Leinen) presenting alongside GEON PIs, speaks volumes to the convened geoscience community. Individually, geoscientists may be relatively unaware or indifferent to such endeavours. They may actively disapprove of 'earth science money going to computer science research,' or of 'students not learning field techniques and only knowing dips and strikes.' But to the extent that such symposia are able stand in for community opinion or at least set an agenda for discussion, GEON has been able to effectively place itself in prominent public view.

Similarly, in publications ties have been drawn between GEON, Geoinformatics and major instrumentation projects such as EarthScope:

> The creation of GEON is a first step in developing the critical cyberinfrastructure necessary to achieve the vision of *Geoinformatics* and facilitate other research initiatives, in particular *EarthScope*. GEON is working closely with organizations such IRIS, the U. S. Geological Survey, SCEC, and UNAVCO as well as other IT efforts within the Earth Science community. In particular, the U.S. Geological Survey has joined as a major partner (Keller 2003)[20].

---

[20] Whether GEON is, or is not, intended to facilitate research initiatives such as Earthscope is not the topic of this section. I am interested, rather, in the means by which such relations are enacted: how GEON is becoming, but perhaps not yet, the clearinghouse for EarthScope data or synonymous with Geoinformatics. For example, whether 'GEON is information infrastructure for EarthScope' remains a hope of GEON participants but highly contested outside this circle. GEON can certainly be described as having established a loose coupling with EarthScope. For example participants of each project attend each others meetings, conduct occasional site visits, and publications from both projects make mentions of each other. However, as a single example, when I asked an NSF officer directly about the link between GEON and EarthScope he responded with significant decisiveness against such a union. He describes how upon securing the Major Research Equipment grant (MRE - see appendix A) geoscientists had flocked to Earthscope for funding, resulting in a heterogeneous "Christmas tree" of research projects. In response the project had been narrowed to three independently managed subunits:

> We had become a Christmas tree, which was exactly what the leadership at NSF did not want. They wanted a well-

The next year, in 2005, GSA approved a new division of Geoinformatics, a regular subsection of the conference dedicated to matters of "computation, information and geology". This division was organized with support of GEON participants, and with explicitly stated intents to generate further support for GEON. A geo PI writes to the general GEON participants outlining his plan and seeking further support:

> Along similar lines, i.e. generating community support for GEON and similar projects, I decided to approach the Geological Society of America to create a new DIVISION OF GEOINFORMATICS. The proposal has the support of GSA management, and I am looking to gather supporting statements from over 100 GSA members to complete the paper work. If you value this initiative, send me an e mail stating that you are supportive of the new Division and I will be in touch with you. I emphasized to GSA that long term community activity in IT based solutions for our science questions will influence policy makers more than individual projects. (GEO PI to GEON email lists)

This same year GSA published 'Geoinformatics: Data to Knowledge' an edited collection by GEON PI Krishna Sinha. Finally, in 2006 GEON and USGS partnered to organize an annual conference for Geoinformatics; for all intents and purposes that was GEON's All-Hands meeting. By 2006 Geoinformatics was simultaneously an annual conference, a division of the annual GSA convention and a publication of that group. Each of these initiatives had been spearheaded by GEON participants, and each

---

> focused project. We were having quite a time with the 3 components as it was, explaining and trying to justify these as a single project. […] We wanted to manage this as 3 separate pieces because it would be a lot easier to manage, […]. So if anybody thinks we were going to start bringing in GEON under Earthscope, then over my dead body.

prominently featured the GEON project. In conference and in publication Geoinformatics was becoming intimately linked to GEON.

An NSF officer comments, for him GEON has already become the clearing-house for any questions about IT in the earth sciences. He describes part of his responsibilities at the NSF as visits to individual geoscience departments across the nation. In his slides he would include updates regarding the progress of Geoinformatics, but if there were detailed questions about the information technology he would direct them to GEON:

> Geoinformatics, this, for example, […] People would talk to me about it and I would say 'get in touch with the SDSC'. Just look up GEON on Google and you'll get there.

For this officer GEON has become the default referent for general questions about IT in the earth sciences, a contact point for developing technical interdependencies, and a basic equivalent to Geoinformatics.

But GEON is not yet the inevitable platform for the earth sciences. Debates continue in multiple forums, often in a language paralleling those articulated at Building the Geoinformatics System. For example at the GSA Pardee Symposium on Cyberinfrastructure one geologist turned to me in the middle of a presentation[21] "as a community we just don't want to go *top-down*! Lots of other disciplines might be able to do that, but lots of us are field researchers, and our data is diverse!" This appeal to 'field-research' is a laminated shorthand for the autonomy and individualist nature of his version of geology, it is an implied a preference for 'bottom-up' approaches.

[21] Presumably seeing me as a geology graduate student?

Partnerships with the various institutions of earth science, including international alliances, contribute to the development of further partnerships. These loose couplings, are the materials in generating an accountability for GEON. One prominent account has been the effective growth of GEON as an emerging clearinghouse for geoscience data. This growth in partnerships is usually framed in a positive light as a marker of success in the project. However, another prominent account draws on these partnerships to express doubts about them, casting loose couplings as myths. The same sets of relationships have served both to point to stagnation or as markers of a success in the making. For example, in speaking of “all these partnerships” a geo NSF officer’s stated:

> They agree to do everything they are asked, but its not clear what they are actually *doing*. […] Some PI, some geoscientist, comes up and says I want you [GEON] to integrate this for me. Fine, but how do we know that's the best thing to build now, the correct path to be on? So I think GEON suffers a little bit there in terms of getting a bit thin in some areas and defocused.

What is at stake in these discussions is the definition of the ‘pull’ in the geosciences and the issues GEON is meant to address. This debate dovetails two questions: what are the purposes of the GEON project? And what are the needs of the geoscience community? As GEON has grown – in partners, in name – GEON has become increasingly synonymous with Geoinformatics. In contrast to 2003, by Geoinformatics 2006 GEON is able to host this meeting in conjunction with the pre-eminent geoscience institution USGS. From one definition of a geoscience community need (earth science is heterogeneous and divided, this precludes multidisciplinary collaboration) GEON is succeeding in achieving its outlined goals. From a competing definition of community need (earth scientists need novel functionalities in the conduct of research) such partnerships appear to be a misdirection of efforts, and a squandering of limited resources.

A Geoinformatics Goldmine?

In 2005 EAR/GEO announced a new funding line for 'Geoinformatics' under its division of Instrumentation and Facilities (IF). With this funding line, what began as the generation of a pull for informationally oriented projects, through workshops and conferences, become a 'push' as a potential source of future funding. The request for proposals (RFP) was worded very strongly in support of data integration efforts. It does not mention GEON, however, it does encourage applicants to form links with existing infrastructure projects in the geosciences, and in particular "proposed information technology platform(s)". For GEON participants this RFP became an occasion to shift past loose couplings that had characterized its earlier years and had become a source of criticism. They sought leverage their position as the most likely information technology platform for the earth sciences by offering themselves as a means to match the criteria of the RFP: interoperability with extant earth sciences data and IT projects.

GEON participants were quick to jump on the opportunity and this email, directed to the PIs, appeared only weeks following the Geoinformatics RFP:

> TO: GEON PI List-Serve
> Date: May 9 2005
> Subject: Geoinformatics call for proposals
>
> I think it would be good to encourage people submitting proposals to this program to reference and link to GEON. The proposal text mentions the following:

> "EAR/IF expects that Geoinformatics proposals will: demonstrate an awareness of existing geoscience information technology infrastructure and developments, where appropriate, present plans for integration and compatibility of proposed information technology platform(s) within the network of existing geoscience information technology infrastructure"
>
> We will post this call on our website….and perhaps mention that those interested can contact one of us in GEON… is that OK?
>
> PS We need to include this call for proposals in our new News section on the front page [of GEON's website].

The strategy here matches those outlined above: encouraging partnerships across projects, increasing GEON's public profile through posting on its website and making Geoinformatics synonymous with their own activities. GEON is well positioned to make itself available as a support structure for other geoscience information projects, assisting in securing funding by offering possibilities of a technical coupling. By offering GEON IT services as a solution to the integration problem the project becomes a 'push' in conjunction with the Geoinformatics funding line.

In visiting the NSF I was fortunate enough to interview program officers in the midst of evaluating the proposals of the first Geoinformatics funding call. As I entered an office to discuss GEON, one EAR representative pointed to a stack of yellow folders sitting behind him. These were the 22 proposals for the new Geoinformatics funding line:

> this [RFP] was an experiment for this community and it has really kind of grown and exploded to what you see here with 22 new projects seeking to be like GEON. We didn't predict that, but we thought that was probably going to be the case […] So I think that's what's exploded in the last 5 years. GEON has really been a focus point, a nucleus, a

> catalyst, if you will, because they've reached out. […] Some of these communities, some of these 22, have caught on to that and they've linked up with San Diego or another supercomputer center to help them.

The effort of GEON to align itself with the Geoinformatics funding line was substantively successful, and similarly its effort to define itself as 'the geoscience network.' The proposed projects themselves are each considerably smaller than GEON, described as "contained within one disciple" or perhaps crossing two or three:

> GEON is the only what I would call truly platform activity. Every other proposal we have over here is a single domain, specific project. If we funded all 22, each of those could plug into GEON as sort of a GEON node- for mineralogy, or a GEON node for paleontology, or a GEON node for structural geology. So GEON is unique. There is no other GEON.[22]

Part of the evaluation of these proposals – through peer review, the program officer himself, and the panel which makes the final recommendations – is how convincingly a collaboration has been described with existing geoscience informatics "platform(s)".

--

To have effectively made GEON *a consideration* in the evaluation of future information project in the earth sciences is not equivalent to an inevitable GEON as *the* geosciences network. During my research at the NSF, and in informal discussions with NSF officers throughout the years, there have been many arguments deployed for and against such push for GEON to become such a platform. We have already seen such debates in at Geoinformatics 2003

[22] He is speaking figuratively here, the Geoinformatics funding line is nascent and quite small. He estimates only a small portion of these will come to be funded.

i- ‘top down’ and ‘bottom up’ strategies;

ii- ‘community driven’ projects coming together or single ‘centralized’ databases;

iii- whether the community is technically and collaboratively backwards and must be transformed or whether the infrastructure should support existing arrangements;

Arguments of NSF officers evaluating Geoinformatics paralleled these discussions. For example one officer summarized a lively debate regarding the feasibility of such a single platform, and the NSF’s role in promoting it.

> We've discussed this quite a bit internally is when to simply say "Okay. GEON is *it*. GEON is a platform. You guys all play with GEON. Plug in to GEON. We don't have enough money to fund everybody. Everybody's IT person use an IT person at San Diego." We're not there yet, partly because we're not sure GEON is where we want people plugging in. I view this next proposal [renewal possibilities of GEON following its ITR grant] as a huge test for what's going to happen there. It's conceivable this is not going to review well, and we sunset it. I doubt it, but I don't know.

Today the geosciences are on a trajectory which has come to include cyberinfrastructure in its possible futures. CI has become part of the going concerns in the research, funding and organizing work of practitioners in the earth sciences.

Conclusion

The ‘pull’ to Cyberinfrastructure in the geosciences is not complete. Or more accurately it must be continuously enacted. In doing so a ‘need’ for information technologies must be articulated in line with the goals of CI. In the last chapter we saw how GEON 'could have been otherwise,' or in other words, the same funds could be have been distributed to multiple smaller IT projects or a single 'more focused' large ITR. In this chapter we have seen how even following the awarding of the GEON ITR, a 'platform' and 'clearinghouse' for earth science IT and data, it is by no means inevitably so. Rather, the inevitability of GEON is an achievement in practice.

The histories of GEON are tied to those of Geoinformatics. Actors project multiple trajectories for its future development. In a history in which GEON was born out of Geoinformatics, as its ideal typical manifestation, its future is tied to becoming the Geoinformatic model for the earth science, 'the geosciences network'. If GEON was always already an information project amongst others in geology, then its future is tied to the production of short term resources that benefit the earth sciences. If GEON is a cyberinfrastructure project amongst many emerging such endeavours, then its future is tied to that of cyberinfrastructure or informatics (to BIRN, and SEEK and the new Office of CI).

The public issues of geoscience are at stake in the definition of Geoinformatics. Will the Geoinformatics System be a project of the NSF as it oversees the development of IT solutions and ensures low redundancy and high interoperability, or; will Geoinformatics emerge at the intersection of 'bottom up' efforts in the earth sciences? To an extent GEON has successfully positioned itself as the spokesperson for Geoinformatics through the conferences it has organized, in publications and through a deferral of expertise. But from another perspective, the purpose of GEON was never to become an umbrella and speak for the informational concerns of earth sciences.

The purposes of GEON are at stake in this daily work and along with these its modes of evaluation. Is GEON a community building project, collecting informational projects under a single umbrella? To the extent that participants at the NSF and in GEON are able define the project as such, its work in collecting partnerships, building alliances and providing a venue for community discussion mark the successes of GEON. Is GEON an information technology resource provider? To the extent that the project is defined, and evaluated, in such a manner, the various partnerships appear as disparate linkages.

In each of the three sections of this chapter I have outlined the practical enactment of GEON as a public issue. However each enactment is specific and each is the site of negotiation and debate. GEON is the geosciences network or a single project amongst many. The problematic of interoperability is cast as the solution for the earth sciences or as a 'top down' imposition. GEON is a success in building partnerships or a dubious effort at providing an umbrella of unknown value. In each of these sections what is at stake is the existence of a 'pull to cyberinfrastructure'.

The Atkins Report, the model of funding in ITR, and organizing for CI more generally, articulate revolution as motivated 'from within', as necessarily the work of domain scientists:

> At its core the ACP [Advanced CI Program] involves rethinking the processes and methodologies underlying individual scientific and engineering fields. Domain scientific and engineering researchers must step up and enthusiastically create and pursue a vision.(Atkins 2003:50)

This is what I have described as the generation of a pull to cyberinfrastructure. Not simply a formulation of 'need for IT' within earth science, but the adoption and 'enthusiastic' uptake of CI as *the* model for IT development. This is the embracing of the logic of interoperability which envisions a divided sciences incapable of communicating, collaborating or sharing data across disciplinary boundaries; problematizes these in the face of progress, multidisciplinary research, and interoperable computing resources; and posits a resolution in the form of the development of a single underlying informational substrate: cyberinfrastructure.

--

In Part I of this dissertation I have focussed primarily on the institutional action at the NSF an in the emerging venues of Geoinformatics. In Ch. 2 these have been the models of the funding CI, an in particular ITR. This has included the practical organizing at the NSF to bring together disciplinary work, and defining GEON as twice basic research along with infrastructure development. In this chapter institutional action has been the work of standing-in for the communities of geoscience: articulating its needs, and the work of enacting a ‘pull’ to Cyberinfrastructure. In the next chapter I turn to the question of practical organizing within GEON. I ask, how have the computer and earth science participants planned and implemented the development of a technical infrastructure for earth scientists; how have they come to know their users, and; how have they organized around disciplinary difference?

–Chapter IV –

Organizing for Interoperability

–

Emergent Infrastructure

At the birth of GEON participants had only the vaguest organizational vision, and even less formal organizational expertise. The NSF funding proposal was similarly lacking in specific organizational commitments; at best there is a budgetary distribution between the eleven PIs, an outlining of long-term goals, and two 'test-beds' which would serve as developmental drivers and empirical foci for GEON's IT tools. Much of the initial kick-off meeting was dedicated to introducing available IT technologies to geo-scientists, and sketching out 'science questions' and the necessary domain resources (databases) to answer these questions for the two test-beds.

It was only days before the All-Hands meeting that a project manager, Dogan Seber, was finally hired for GEON – this could be called the first 'formal organizational act'. This was six months into the project. At this time they also hired an administrative assistant, Margaret Banton, dedicated both to GEON and more generally serving as a co-ordinator for Chaitan Baru, a senior member of the SDSC and a lead IT PI within GEON[1]. The program manager, already a PI within GEON, and a geo-scientist, would re-locate to the SDSC. This move has proven crucial, as Seber provides the only 'domain

[1] As organizational, communications and administrative tasks have taken greater prominence within GEON, Banton has ceased being Baru's personal assistant and become a full time administrator for GEON, and recently has shifted out of all secretarial administrative tasks and concentrates exclusively on communications and co-ordination. This signals an underestimation of these sorts of tasks in the GEON proposal and members initial organizing activity.

presence' for GEON at the SDSC. Approximately one month following the first All-Hands meeting (seven months into the project), the GEON workgroup was initiated at the SDSC, a weekly meeting of the entire SDSC team working on GEON. In many senses this appears slow progress, however, GEON is an invention almost from the ground up. Many of the participants had little experience in large-scale projects. Even for those with such experience, building an umbrella infrastructure for the earth sciences provided significant new challenges. GEON has been a continuous process of creative invention, not only technically but in time and resource management, and in enacting collaboration across distance and disciplinary difference. For this work I use the short-hand *organizing*.

This chapter follows the work of GEON participants in enacting an *emergent infrastructure*. This introduction provides a brief overview of the literature on new organizational forms. These do not represent the approach adopted in this study, but rather provide a heuristic understanding of GEON as a 'type' of organization. From the organizational literature GEON and community CI projects more generally are novel organizational *forms,* with a research driven goals and a loose organizational structure dubbed *adhocracy* and seeking to produce a technical artefact 'infrastructure'.

Instead of placing GEON within a general typology I focus on GEON participants' everyday methods for organizing through talk and in practice. GEON actors endogenously develop typologies for understanding their organizing work. They develop a division of roles and meetings for planning and executing the project and for distributing resources. And they devise methods for coming to know the community of earth scientists they seek to serve and methodologies for long-term technical development.

Theoretically this chapter follows the syncretic approach of Taylor and Van Every; it is from their work that have adapted the term *emergent organization* (Taylor and Van Every 2000). The key method in their approach is to follow situated talk and other forms of communication and coordination as the means by which actors constitute organization: "What we ought to be studying is not organization or ideology, because neither has any ontological status independent of communication, but he processes of communication by which we continue to construct both to become the world we live in," (ibid.: x). In substituting the term infrastructure for organization I seek to emphasize members work of constructing a technical artefact -- its increasing materiality over time -- without losing the sense of continuous local enactment (Latour 1991). GEON is an emergent infrastructure in that outcomes remain at stake in the daily deliberations of its participants. While GEON has a proposal, with articulated goals, technical means, funding and a five-year development plan, enacting this into an organization (Weick 1969) and an infrastructure for the earth sciences is the daily work of participants.

The chapter itself is structured into three sections. The first section describes the framework for understanding organizing as talk and practice (Boden 1994), and provides some of the primary categories that GEON participants have negotiated in their work of organizing (Garfinkel 1967). It is in the variety of meetings and the everyday encounters surrounding them that GEON has been shaped as an organization, and from which an architecture for infrastructure has emerged. The second section outlines GEON participants' methods' for coming to know its future users, and provides a detailed vignette of such activity. The primary means for knowing the community of geoscientists has been GEON's earth science PIs themselves who have served as surrogate users, informing and participating in design decisions (Woolgar 1991). The third section outlines the 'two-tier approach,' a methodology within GEON for the long term technological development of infrastructure. In this approach members identify differences across the domain / IT boundary (Gieryn 1999) by characterizing diverging interests of computer and earth scientists. The goal of the two tier method is to dovetail these diverging disciplinary interests and produce committed (Becker 1960) stakeholders in the development of community infrastructure.

--

The liabilities of producing a new organization are well documented in the literature. Stinchcombe notes that "as a general rule, a higher proportion of new organizations fail than old. This is particularly true of new organizational forms, so that if an alternative requires a new organization, it has to be much more beneficial than the old before the flow of benefits compensates for the relative weakness of the new social structure," (Stinchcombe 1965, p.148). This slightly functionalist[2] take on new organizations nevertheless sums up a general consensus in the literature that i- new

---

[2] Van de Ven and Poole (Van de Ven and Poole 1995) have identified four general theories of organizational change and emergence:
i- life cycle theories: embracing a metaphor of organic growth, identify stages of development – change has underlying logic.
ii- teleologic theories – organization is purposeful, adaptive and goal directed. Both these accounts (i and ii) are often functionalist (Merton 1968).
iii- dialectical theories: organization exists in pluralistic world in which all compete for domination, stability and change is the product of wins, losses and alterations in the balance of power
iv – evolutionary theories: change occurs through a continuous cycle of variation, selection and retention. New elements such as rules or routines arise through random change, selection occurs primarily through the competition for scarce resources, retention preserves them through some type of copying or reproduction process.
Notably, she ignores the well developed strains of historicist and interactionist approaches *within* organizational literature, such as enactment theory, ethnomethodology or emergent organization.

organizations must be distinguished from new organizational forms and ii- that there is an immense investment for both, but new organizational forms are quite costly.

Aldrich (Aldrich 1999) has called new organizations *within* a form 'reproducer' while organizations of a new form are 'innovators'[3]. The immediate impulse is to classify GEON as an innovator organization, which is true to a very large extent, however, as I have noted above, the NSF funding was awarded to the SDSC because of their ability to manage projects which – previous to the introduction of the term cyberinfrastructure – shared some of the ambitions of the project. This said, the sheer scale of the endeavour, the heterogeneity of both the IT and domain community, and the novelty of the technologies involved places GEON outside the scope of a reproduction of anything. Furthermore, GEON participants themselves are at a loss for previous models, and generally find themselves creating ad hoc structures, only lightly influenced by previous experience.

Mintzberg has called this an 'adhocracy'[4] in which there is little "formally organized hierarchy, tending to emphasize mutual adjustment and to engage in team projects, to use matrix forms, and to mix centralization and decentralization. Role clarity, sharp divisions of labor, chains of command, and standardization are weak in

---

[3] "Reproducer organizations are defined as those organizations started in an established population whose routines and competencies vary only minimally, if at all, from those of existing organizations. They bring little or no incremental knowledge to the populations they enter, organizing their activities in the same way as their predecessors. Innovative organizations, by contrast, are those organizations started by entrepreneurs whose routines and competencies vary significantly from those of existing organizations." (Aldrich 1999, p. 80).

[4] The term adhocracy is attributed to sociologist Alvin Toffler in his popularly directed social history of modernity 'Future Shock,' he identified this as the quintessential modern form: "[adhocracies] now change their internal shape with a frequency – and sometimes a rashness – that makes the head swim … Vast organizational structures are taken apart, bolted together again in new forms, then rearranged again. Departments and divisions spring up overnight only to vanish in another, and yet another reorganization," (quoted in Mintzberg 1992:255).

adhocracies, but the search for innovation is strong," (Mintzberg 1992). These organizations are understood to arise, and function best, in situations where there is little agreement about goals, means, or structure.

The notion of an adhocracy depathologizes these organizations[5]. Mintzberg argues that in organizations dominated by professionals[6] function best as adhocracies. These are problem-solving, rather than performance, structures:

> high horizontal job specialization based on formal training; a tendency to group the specialists in functional units for housekeeping purposes but to deploy them in small, market-based project teams to do their work' a reliance on the liaison devices to encourage mutual adjustment, the key coordinating mechanism, within and between these teams; and selective decentralization to and within these teams, which are located at various places in the organization and involve various mixtures of line managers and staff and operating experts. (Mintzberg and McHugh 1985)[7]

The adhocracy is not especially informed by classical principles of management, and is particularly indifferent to staples such as the notion of the unity of command since it may

[5] These cases, which fall outside the norm of bureaucratic decision making, and which are rare (at least in the eyes of organizational theory) have at times been called 'organized anarchies' – here problems, solutions, participants and choice opportunities are views as flows that move relatively independently out of the decision arena, known as the 'garbage can'. This model of how decisions are made and goals set is based on contingent relations between participants, resources and ideas available at a particular time, it stretches outside of the usual rational decision making models: "Although the system described seems bizarre and even pathological when compared with the conventional model of rational decision making, it does produce decisions under conditions of high uncertainty," (Scott 1992, p.306, see also Cohen, March, Olsen 1972).

[6] Which he understands as highly informed about their environment, autonomous, and having jurisdiction not only over membership but also the ethical, political and social play around members. When professional organizations become very large (as in a university), they become 'professional bureaucracies,' which are dominated by a differing logic of standardized expertise and accreditation (Mintzberg 1992).

[7] They go on to note how an adhocracy, focused on innovation, cannot allow for standardization for coordination – since the innovation of GEON is a variety of standardizations, this does not seem to apply here. This said, perhaps the adhocracy structure is a threat to the ability of GEON to produce operating standardization..

be necessary to override the chain of authority at times. While drawing on the specialization, expertise and professional knowledge of various fields, the adhocracy must not be dominated by any particular specialization since this is understood to produce standardization rather than innovation. Instead its members must find methods to break through the conventional specializations and differentiations:

> most of the managers do not "manage" in the usual sense – that is give order by direct supervision. Instead, they spend a good deal of their time acting in a liaison and negotiation capacity, coordinating the work laterally among the different teams and between them and the functional units. Many of these managers are, in fact, experts, too, who take their place alongside the others on the project teams. (Mintzberg 1992:256)

Within GEON this 'managerial' structure has already emerged, each sub-division (whether it be a sub-group at the SDSC, or each of the PIs) has a manager/co-coordinator who does not take any official mantle of power, but rather has the special responsibility of maintaining lines of communication. However, these 'managers' have not self-consciously adopted this role, and so they do not systematically adopt such an identity. Rather than managers, per se, they are GEON participants trained as computer or earth scientists: "administrative and operating work blend into a single effort – planning and design cannot be separated, since both require the same specialized skills" (Mintzberg and McHugh 1985).

Mintzberg claims that innovation is the goal and the results of such efforts can never be fully predetermined. This means a more standard understanding of process, where formulation of problem is followed by implementation, is impeded. Any action of planning which separates conception from action – planning from execution, formalization from implantation – hinders the flexibility of the organization to respond creatively.

Mintzberg also point to a series of problems associated with adhocracy:

> i- even dedicated members of adhocracies periodically exhibit low tolerance for its fluidity, confusion and ambiguity
>
> ii – combining the ambiguities with its interdependencies, this form is one of the organizational forms most leading to politicization
>
> iii- the adhocracy is not competent at doing ordinary things
>
> iv- the root of its inefficiency is the high cost of communication
>
> v- there is also the likelihood of unbalanced workloads

We will see below how each of these problems comes to be articulated by participants as the central difficulties in GEON's development (see esp. Ch.4.3). Mintzberg's ideal type characterization and problematization of an adhocracy fit well with the everyday ethnographic experience of GEON. However identifying such a 'type of organization' does not follow with the methodology of this dissertation. I provide Mintzberg's framework for the reader as a heuristic sketch of 'what GEON looks like' from a typological analysis. Furthermore Mintzberg says little about the emergence of adhocracies, the subject remains under-researched. Thus it is difficult to say whether GEON is a typical case: do adhocracies originally begin 'pre-organizationally' and then become adhocratic, or are they designed as such? From the perspective of an interactionist inquiry the question is misconceived. Focusing on practice, talk and material arrangement instead I ask the question, what are the methods by which an organization is constituted and how do members render its structure durable?

4.1 – Organizing the 'Form' and 'Structure' of GEON

In the methodologies which emphasize doing and talk there are divergences on just 'where' to find the practices of interest. Ethnomethodological studies of scientific work have usually been concerned with 'the laboratory', 'the field' or perhaps 'the classroom' as sites of practice. In the next chapter I focus on GEON's efforts to create a 'knowledge lab'[8]. These are sites of unambiguous practical action, where 'something gets done'. The intended results are technologies of integration, "something new about the Rockies" or a representation of geoscience knowledge. However, while GEON is a project with the goal of building a national information infrastructure, the methods to do so are primarily organizational. In this section I focus on 'meetings' as the primary site for organizing in GEON. These are the annual All-Hands and Principal Investigator meetings, the weekly workgroup meetings at the SDSC, and the sporadic topical workshops. Each meeting frames a different topic and purpose, but together they traverse the organizational, technical and institutional scales.

Meetings as the 'Organizational Forms' of GEON

Meetings have been the common center point for the diverse activities in GEON. Participants identify over a dozen types of recurrent meetings: annually, bi-annually, monthly, weekly or daily. It is in these meetings that the bulk of planning, coordination, distribution and delegation occur. Rather than seeing these gatherings as epiphenomenal

[8] I use laboratory in the sense imparted by lab studies (see Ch. 1). In this formulation the lab is the place, the practical methods and the equipment which render the material world knowable. Similarly with knowledge representation a 'knowledge lab' is the site and practical methods by which the 'knowledge' of the geosciences is made inspectable and manipulable through language and by inscription.

to a technology project it is through the talk, the activity, the demonstrations, the declarations, at meetings that an organization becomes an inspectable object. GEON is constituted again and again through the summative statements of its activities, purposes and projected actions. Those larger conferences which span all participants simultaneously serve to characterize disciplinary difference and to formulate solutions to these divergences. Contrariwise, those smaller meetings which span only selected participants are occasions for topical focus, directed technical activity or planning.

For example at All-Hands meetings (AHM) participants from the various geo and computer sciences, educators, and institutional representatives (NSF, USGS, NASA, Earthscope) give short presentations of the technical and administrative content of their recent work. The goal is to quickly summarize, in an accessible manner, the work of GEON's specialized enclaves for all participants creating a shared organizational map (Taylor and Van Every 2000). Similarly, the initial Kick-off meeting was organized specifically to introduce the planned technologies of GEON to earth scientist, and in turn, for geoscientists to present their specializations and research goals. In contrast to these inclusive and topically broad meetings, the number of participants in the internal coordinating committee (ICC) is kept intentionally small; these meetings are composed of key GEON spokespersons (Latour 1987). The small number of highly authorized participants, the brief agendas and the 'business-like tone' of meetings make possible quick decision-making schedules without encountering the minutiae of GEON's diverse constituency. Meetings can serve to summarize diverse activity, rendering the organization knowable to participants, and they can serve to direct specific activity, rendering the organization manipulable in situ.

In her studies of organizational work ethnomethodologist Deirdre Boden has placed meetings at the center for understanding the enactment an environment; an organizational past; the collective understanding of an organization's current state and; the strategies of future plans:

> Meetings are the essence of organizational life. They are the locus of management in action. [..] Meetings are also ritual affairs, tribal or corporate gatherings at which the faithful reaffirm solidarity and warring factions engage in verbal skirmishes. When in doubt, call a meeting. When one meeting isn't enough, schedule another. However formal or informal they may be – whether the synthetic crispness of the executive boardroom or the laid-back egalitarianism of a student council – meetings are "where the action is," even if "nothing ever seems to happen" in them. They are the encapsulated enactment of the organization itself, with all its goals, agendas, coalitions, anarchies, and occasional battles and assassinations" (Boden 1991:229).

In GEON battles and assassinations are rare. The polite atmosphere of meetings is more likely to be tinged with technology inspired confusion and a collective insecurity of identity and purpose. For example, in the next section I trace a joint editing of the GEON webpage at a PI meeting, this becomes an occasion for discussing GEON's shifting purpose and identity. 'What is GEON' is negotiated by participants over time, across disciplinary differences across geo and computer science and within the disciplinary umbrella of 'earth sciences'.

Boden's vision of 'meetings within meetings' is apt for GEON. Larger, 'more important' annual meetings such as AHM are planned in the smaller environments of weekly meetings. This includes agenda setting, order of speakers, run-throughs of technology demonstrations and rehearsals of keynote speeches. In turn, the long lists of 'action items' built-up at an AHM spawn workshops, workgroups and mailing lists. Following from one meeting to another a 'practice oriented' ethnographer might be tempted to ask "where is the actual work?" but within organizations it's meetings all the way down. We shall see in the next chapter that a great deal of work in the 'knowledge lab' of ontology building was also conducted at meetings, workshops and over video conferences.

In the chart below is a brief elaboration of the primary meetings which have shaped GEON as an organization. For a more detailed description of these meetings please see Appendix B. The names of these meetings – All-Hands, concept space -- are what Boden has called *laminations* (Boden 1994, see also Ch. 1.3). Laminations are collected meanings that overflow their locally negotiated definitions in interaction. Ethnomethodological method has the limitation of requiring the analyst to demonstrate the situated establishment of meaning through negotiation amongst members. Meaning is imparted partially through the subtle conversational contestation of categorical boundaries. However Boden's notion of lamination acknowledges that the same members may come to impart meaning over time, that a phrase can become 'stock' through layered uses and shared experience. This is the case with GEON's meetings.

For example, All-Hands have come to be the annual turning points for GEON, one each year marking a significant milestone: having moved past the introductory phase of the kick-off meetings the first All-Hands was the beginning of 'real work'; the second was conducted just previous to the key NSF site visit and evaluation, allowing participants a rehearsal before their public evaluation (Goffman 1974[1986]); the third became a conference, marking GEON's public opening, and; in the fourth became synonymous with Geoinformatics. Each of these events have gained layered significance – or become laminated – for GEON participants, connoting a sense of progress, growth and tracking a trajectory of success. While laminations can be traced in action by the long-term ethnographer here instead I provide brief elaborations of these seminal meetings to guide the reader in understanding GEON's organizing These categories are the continuously enacted 'form' of GEON as an organization.

The Meetings of GEON

Kick-Off -- The first meeting following the funding of the GEON project. This meeting was held at the SDSC on November 17-20th, 2002. The kick-off was an assembly which primarily served to introduce the IT team and their planned technologies to the geo-scientists. A great deal of time was spent in presentations and Q&A about these novel and emerging technologies. In turn, geo-scientists presented their *science questions* and some initial descriptions of the kinds of data and integration necessary to achieve their goals.

All-Hands (AHM) -- The AHMs were intended to bring together all GEON PIs, their research retinues, and the growing collection of partner institutions. These meetings provided an occasion for the distributed research teams to 'see across' GEON diverse sub-projects. The explicit goals were for each PI to provide summary presentations of their local work, collecting the necessary resources to govern GEON and in turn forming coordination plans for future action. Over time these meetings have morphed into a public conference and eventually into Geoinformatics – see Ch. 3.3.

Principal Investigator (PI meets) -- These meetings included the GEON PIs, but also a smaller selection of their research teams and representatives of significant regular partners such as DLESE. As with the AHM these became opportunities to collect the work of GEON's distributed teams. An extensive excerpt of this meeting is explored in section 2 of this chapter.

Internal Coordinating Committee (ICC) -- A regular distributed meeting held monthly using video conferring technology. The group was small, usually involving only the lead PI, the project manager, the two leads for the *test-*

beds (see below) and GEON's administrative coordinator. The meetings were primarily administrative, including setting the agendas for AHM and PI meets, discussing new GEON partnerships, or planning for upcoming conferences, posters and demos. The small group of selected participants and usually tight agenda gave these meetings a very decisive feel, particularly in contrast to AHM and PI meets, which ranged broadly in topic, purpose and style. In the terms of enactment, this meeting offered a relatively clean vantage point for observing GEON act *as* an organization.

Workgroups -- The workgroup meetings are the weekly meetings for activity in GEON that is centered at the SDSC. These come closest to representing the 'everyday' site for managing GEON. These meetings bring together the top level administrative managers of GEON (lead PI, project manager, administrative coordinator), along with the central team of IT experts. This meeting has become the organizational nexus for GEON. Over the years various efforts were made to open this meeting to the (non-SDSC) geoscientists: call-in telephone lines, webcasting and online archived meetings. Following a major local (SDSC) re-organization of GEON in January 2004, the single weekly meeting divided into a changing configuration of daily meetings. The particular meetings have shifted over the years, but the initial divisions were: portal, knowledge representation, visualization, systems, GEMS, GIS/Mapping and workflows.

Workshops -- GEON workshops are focused and topically specific, they are events rather than regularly held meetings, and often involve many participants from outside GEON's core group. For example, GEON has had workshops on visualization and ontology, and also a larger 'Cyberinfrastructure Summer Institute for Geoscientists' for basic introduction, framing science questions, and training with emerging tools.

Concept Space -- The ontology development workshops. These meetings were site for 'knowledge capture,' and formalization in the development of knowledge mediation technologies. The name for these meetings has shifted over time, from concept space meeting, to ontology workshops, to community based ontology development. These meetings are discussed extensively in Ch. 5.2

The 'Structure' of GEON as Sensemaking and Framing

It is difficult to capture the heterogeneity of practical action and talk of the various GEON meetings. For example, for an ethnographer, the PI meets provided opportunities for topically framed informal discussion with GEON's key figures. This included shared flights to El Paso, breakfasts, lunches and dinners, and tours of the local rock formations. The same opportunities were open to all GEON participants, and became crucial moments for exchange and arrangement. The ‘formal’ moments of meetings (i.e. between breakfast and the morning break 9:15-10:25) blended seamlessly with informal work-inspired discussions. As Boden notes, the actions in and around meetings serve to “inform, amuse, update, gossip, review, reassess, reason, instruct, revise, argue, debate, contest, and actually *constitute* the moments, myths and, through time, the very *structuring* of organization” (Boden 1994:8).

It is this full range of these activities which the concept of *sensemaking* attempts to capture. Sensemaking describes the methods by which members come to see circumstances as meaningful, as ‘an event’ or an organizational ‘decision’. The methodology here is to unpack the notion of and event or a decision, and instead take it as an outcome of collective practical reasoning:

> The operative image of organization is one in which organization emerges through sensemaking, not one in which organization precedes sensemaking or one in which sensemaking is produced by organization (Weick, Sutcliffe et al. 2005:6)

Within enactment theory sensemaking is the practical activities of *talk, writing and editin*g, with a particular focus on organizational forms (such as meetings, agendas, reports or memos). The everyday flow of organizational activity is made explicit through talk and, particularly, articulated in those categories which are organizationally salient. In turn, talk comes to inform writing, or other representations (such as an organizational chart). These must be made to follow emergent or established forms of inscription (Yates

1989), again a form of articulation into salient representational categories. Then, finally, reading, editing and further writing places text in a dialogic form accessible to multiple participants and again performed through negotiated talk. Taylor and Van Every have neatly encapsulated the relation between the talk/writing of sensemaking and the action of organizing: "sensemaking is a way station on the road to a consensually constructed, coordinated system of action" (Taylor and Van Every 2000, p. 275). At that way station, circumstances are "turned into a situation that is comprehended explicitly in words and that serves as a springboard to action" (p. 40)[9]. The outcome 'decision' is then distinguished ('black-boxed') from the methods of its achievement, and an organization can be said to act.

Meetings are sites for reducing complexity, bracketing out the multiple concerns and goals in the activity of organizing. Part of sensemaking is creating a bounded activity, this is often called framing. For example, GEON is a project that is simultaneously geoscience and computer science, it is community building and infrastructure building, it is visualization and knowledge representation. By defining a meeting by subject or specialization it makes possible a focus on particular aspects of development. A simple example are GEON's topical workshops, such as visualization, where it is possible to leave aside other aspects of technical development (such as systems) or pressing concerns of the moment (such as organizing for an upcoming All-hands meeting). A more subtle example are the workgroup meetings, primarily composed of information technologists, where it is possible to speak of technical development or administrative work substantially divorced from geoscience concerns.

[9] Quoted in (Weick, Sutcliffe et al. 2005).

Bruno Latour (1996) has also described this as framing, but unlike sensemaking in enactment theory has emphasized the role of material arrangement. Latour contrasts 'literal framing' with 'framing as a metaphor.' He argues that framing is understood metaphorically by interactionists[10] where it is seen to be conducted through language and continuous in-situ negotiation. In this formulation, he notes, any particular interaction is open to a continuous interruption of topic, process or method. Latour's framing involves material objects – or non-humans actants – such as buildings, "partitions, hideaways, fire-doors" (Latour 1996). Material arrangement helps to hold particular frames in place; they offer greater resistance than the constant negotiation of metaphorical frames. With such resources in hand it requires less effort for agents to bracket out the noise of "the rest of their history as well as their other partners." For Latour metaphorical framing (as per Goffman (1974[1986])) is still negotiated within interaction, but he notes that we must also acknowledge the work of non-humans in producing bounded exchanges. These permit interaction to be less complex. Without the aid of buildings, conference rooms and agendas, interaction would remain more open to 'interference'. Thus at a 'GEON visualization workshop' it is possible to excuse (or exclude, depending on the perspective) the interference of current developments in the field of knowledge representation, research on terrane movement in geophysics, and even whether GEON will receive funding in the next cycle.

As with ethnomethodology, sensemaking tends to emphasize negotiation and talk rather than object-based framing. Charts, diagrams and other inscriptions have a tendency

[10] By which he means certain symbolic interactionists, 'pure' ethnomethodologists and other scholars of the micro.

to be reduced to interpretive practice. Similarly, instruments and tools can disappear from the sensemaking process altogether (c.f. Hutchins 1995). With 'literal framing' a focus on material engagement becomes critical, and ceases to a be a process of 'social construction.' The malleable frames of social construction are accompanied by material arrangements of a relative obduracy.

Michel Callon complements this notion of framing with one of *overflowing*. He notes that framing must be considered an expensive undertaking, requiring a great deal of effort and time. Drawing on the theatre metaphor, as per Goffman, Callon emphasizes the physical and practical aspects of framing, as per Latour:

> Without the theatre building and its physical devices; without years of training and hours of rehearsal put in by the actors; without the habitual mindset of the audience and carefully written dramas which deliberately limit the range of preprogrammed interactions, the framing of a stage performance would be quite simply inconceivable. (Callon 1998)

The notion of an overflow points the (regular) failure of framing in thoroughly isolating interaction from interference. The topic of a workshop can be overruled in the face of immediate concerns and agendas are regularly revised before and during meetings. A presentation on knowledge representation languages can shift to geoscientific content and then the particularities of data practice. Framing, whether through physical or symbolic devices, is expensive and subject to planned or unintended overflows. It is, however, in many senses the primary business of GEON to produce increasingly "frozen" frames in the form of an architecture and a long-term development plan.

--

We have already discussed meetings as framing devices, with their topics, agendas, buildings, conference rooms, and power points. In addition GEON is assisted by various structuring laminations. GEON is in the process of infrastructure building; in Taylor and van Every's terms it is an *emergent organization*. Commenting on Latour and material arrangement, they note that an emergent organization will not yet have collected the **"**frozen**"** quality of "yesterday's organizing" (Taylor and Van Every 2000:277), but must be understood as processually engaged in achieving this. This is particularly true of GEON which is organizing with an explicit focus on the development of automation.

Below I describe some of the structuring laminations that are becoming increasingly materially embedded in technical architecture, scientific focus and long-term development strategy. These are the GEON architecture, the test-beds and the two-tier method. Lamination primarily refers to interactionally built meaning over time. This describes persistent categories used in meeting talk, in proposals, or even in a software architecture document. Over time, though, analysis must shift from 'pure' laminated meanings and trace as they become artifacts in practice: GEON as an emerging technological platform. As with Latour's frames they provide resourceful 'cuts' of GEON for participants and become increasingly resistant to re-arrangement as they are embedded in network and software architecture, in routinized organizational form and developed software tools. And as with Callon's overflowing these frames require work, may not cohere with laminated meanings, may co-exist with overlapping notions, or contradict the organizational forms described above.

GEON Architecture and the 'Software Architecture Document'

GEON participants, and particularly GEON IT participants, have often relied on a tripartite division of the technical aspects in GEON: i) systems ii) visualization iii) knowledge representation. Each of these is considered a fundamental component of GEON's future infrastructure services for the geosciences, but also a site of research within IT. Briefly, systems refers to hardware and network, the grid, and a broad set of 'other' concerns such as security. Visualization are the tools for representing data to the human eye, including mapping. Knowledge representation is nominally the primary purpose of GEON, and includes data integration, searching, and the various knowledge technologies such as ontologies and workflows. The divisions do not map neatly onto, for example, the daily workgroups at the SDSC, however these divisions have served GEON participants as a loose coordinating mechanism. These divisions also inform this study, as I have primarily followed the work conducted as knowledge representation, and more specifically with a focus on ontology development. I will treat each division in turn, and knowledge representation will be treated in greater detail in Ch. 5.

On first glance systems refers to GEON's hardware: its computing, networking and storage capacity. However grid computing is also included within systems. A particular meaning for grid computing is notoriously difficult to pin down, even within those working the field. The word appears with scores of meanings in the Atkins Report (as a synonym for collaboratory, for computational cycles, for physical infrastructure). In the GEON proposal 'grid' refers to links between heterogeneous and distributed forms of computing:

> in GEON we are required to deal with an extremely heterogeneous computing, storage, and networking environment […] GEON will be a pioneering project in

> "democratizing" grid technologies since the GEON Grid must eventually reach a large population of users that includes scientists, government policymakers, engineers, and educators (GEON Proposal)[11].

GEON's physical systems are often called nodes. Nodes are the distributed hardware components of computation, storage and networking. Each of the geo PIs have a node installed at their home institution. Nodes may come in various configurations, emphasizing storage or computational capacity. The nodes themselves are called 'stacks' (because they are appear as vertically stacked computer components). One goal in GEON is the standardization of these stacks: "you should be able to roll them off the truck, plug them in, install the software, connect to the net, and go." In the growth of GEON the establishment of a strong partnership with an institution has often been coupled to the establishment of an on-site node.

Systems has also included a variety of work outside the confines of hardware or the grid, in particular security (Bhatia, Chandra et al. 2005). Security itself is a broad category including concerns for the preservation of the archive (backups), access to system resources (computing and data), privacy (of data and workspaces), protection from viruses or other tampering. Other aspects of security, such as ensuring the quality of data, have usually come under the jurisdiction of knowledge representation or are considered extrinsic to the architecture and in the hand of domain scientists.

[11] See also metaphors of the grid as a regulator of diverse resources such as visualizing tools, access to data and distributing computing cycles (Buyya 2002) and as a 'virtual organization (VO)' designed to facilitate equitable sharing of heterogeneous resources "In defining a Grid architecture, we start from the perspective that effective VO operation requires that we be able to establish sharing relationships among any potential participants. Interoperability is thus the central issue to be addressed. In a networked environment, interoperability means common protocols. Hence, our Grid architecture is first and foremost a protocol architecture, with protocols defining the basic mechanisms by which VO users and resources negotiate, establish, manage, and exploit sharing relationships." (Foster, Kesselman et al. 2001:205)

By all accounts visualization has remained the most neglected aspect in work of GEON's tripartite architectural division. This is partially due to personnel issues, such as the departure of the lead for visualization at the SDSC. The category itself is nebulous, at times referring to mapping and GIS technologies, and at others for more abstract sorts of data or metadata visualization. The latter have often come under the jurisdiction of knowledge representation.

Knowledge representation is the technical facet of GEON concerned with data interoperability. In GEON the particular focus in with 'semantic integration,' which broadly refers to the tying together of data and tools by shared meaning. Semantically integrated resources can be discovered and used by domain scientists by drawing on the familiar terminologies and concepts of their native field. An extensive walk through of a 'user scenario' demonstrating an approach to semantic integration is included in Ch. 5.1. The categories of resources which come under knowledge representation contain i) ontologies or concept maps; ii) workflows; iii) 'smart searches' including concept navigation and iv) data integration, particularly semantic data integration. These four categories overlap considerably in terms of shared software platforms, the experts engaged in producing the tools, and the understandings of knowledge representation (c.f. Berkley, Bowers et al. 2005).

--

In January 2004, at the weekly workgroup meeting the lead GEON PI presented his 'version 0.1' of the GEON software architecture. This is a document which came to be the focus of a great deal of collective effort by both CS and geo participants. For example, shortly after its introduction, a two day workshop was organized in San Diego for collectively discussing the architecture document. Beginning as a highly generalized description of the workgroup divisions and their responsibilities, this document has become a highly detailed plan -- often changing in-situ -- of the implementation and larger architecture of all technical components of the central GEON architecture[12]. The document includes user scenarios, a timeline for development, a plan for the physical nodes and driving applications. The architecture document and user scenarios are discussed in section 4.2, and science drivers are discussed as part of the two-tier method in 4.3.

--

Other 'cuts' (aside from systems/visualization/knowledge representation) to the GEON architecture have been proposed and occasionally enacted. For example, the divisions: public/portal/grid. GEON exists 'publicly' to anyone who wishes to visit the geongrid.org website, or its many subsections distributed across the PI institutions. The portal (which requires a login and password) offers access higher-end resources such as computing tools, draft ontologies, data registration services and so on. The grid, here meaning GEON's distributed computing cycles, is limited to a select group, or as it was described at the time of the use of this description "grid access is vetted" (Geo PI). To some extent these categories are real because they are held-together (or framed (Latour

[12] In one presentation at the SDSC I argued that the sub-division of the GEON workgroup was not sufficiently counterbalanced by new lines of communication. The result would be a 'centrifugal' tendency, where co-ordination would become difficult. The GEON workgroup itself is a hybrid mixture of highly technically specialized talk (which excludes most IT participants) and general organizational discussions; it does not provide a good forum for strong system-level standardization efforts necessary to produce a platform of GEON's scale. As a good countervailing, or 'centripetal' tendency I argued that the Software Architecture served as a good boundary object for co-ordinating technical efforts across the spectrum. There does not seem to be an equivalent co-ordinating mechanism at any level beyond SDSC technical interaction, although perhaps the Portal will serve in this role as the various databases and services are brought on-board.

1996)) by software i.e. an increasing effort is required to gain access across up the divisions.

The Test-Beds

The test-beds are the two U.S. geographic areas of empirical geoscience focus: the ‘Rocky Mountain’ and ‘Mid-Atlantic’. The test-beds were included as of the second GEON proposal. Actors’ accounts link their addition the proposal for the purpose of meeting the requirements for basic research in the earth sciences (see Ch. 2.3). They have been the continuous focus of development since, and in early 2004 the test-beds were renamed DYSCERN (Dynamics, Structure, and Cenozoic Evolution of the Rocky Mountains) and CREATOR (Crustal Evolution Anatomy of an Orogen).

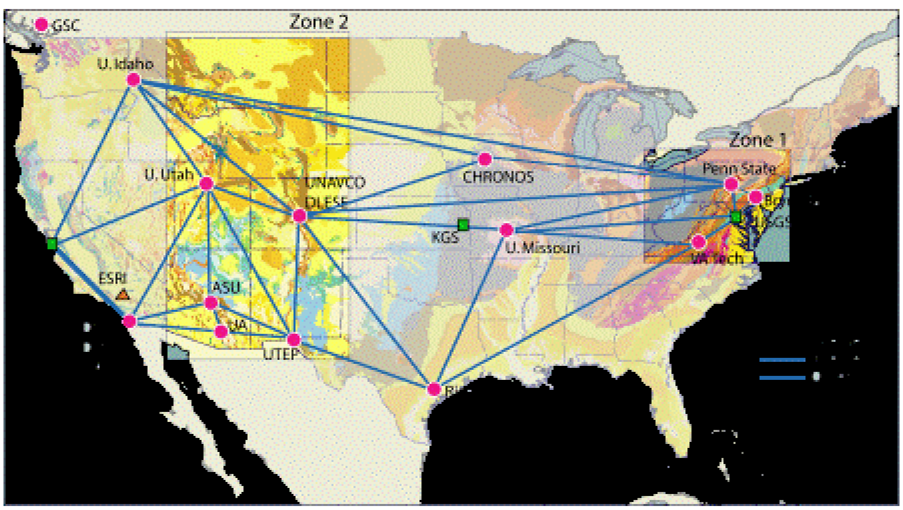

Figure 10: An representation of the GEON network, and the two test-beds (GEON Proposal).

In annual reports or during the NSF site visit, the test beds serve to organize the communication of findings and ongoing efforts in the earth sciences. The test beds have also served as organizing principles for GEON’s work, and access to resources. Decisions about what does or does not count as a relevant test-bed development of IT resources have been delegated to the two lead earth-science PIs. Thus, a routine has emerged in which these two geoscience PIs serve as gatekeepers for access for IT resources. Requests for development efforts (whether for computing resources, ontology workshops or visualization tool development) on the part of geosciences participants ostensively run through one of the two test-bed representatives.

The vocabulary of the 'test bed' is common within CS and IT circles, it is used several times in the Atkins report, and for the Digital Libraries projects (Borgman 2000), and for CI projects such as SEEK and BIRN (see Appendix A.).

The Two-Tier Method

I will follow this lamination more closely below, however, to begin, the 'two-tier method' is a short-hand used primarily by lead PIs. It refers to multiple practical strategies, but they have in common the goal of bridging what have been framed as GEON's central difficulties: IT vs. domain research; short vs. long-term development; science vs. infrastructure.

Very briefly, the first tier refers to the development of information technologies based on 'science drivers', actual research questions of contemporary geoscientists. The second tier is the development of those tools in the first tier using the most contemporary IT available i.e. with consideration for extensibility, usability, proper documentation, metadata and so on. In this formulation geoscientists 'learn something new about the Rockies' *and* are building infrastructure; meanwhile, information technologists have a 'context' for development *and* are working on cutting-edge computer science research.

Organizing the Scales of GEON

It is from the various meetings and structuring laminations in GEON that it has become possible to speak of scale. Scale is an organizing principle of this dissertation but is also reflects the kinds of sensemaking resources brought to bear in situated activity of meetings[13]. In a single morning of presentations a keynote speaker from the NSF will outline the GEO directorate's intended trajectory towards CI, and then detailed work of three technicians for the last three months will be summarized into a ten minute power-point presentation, and a sociologist will outline the formation of pidgin languages across IT and domain. Following synopsis comes action. Later that afternoon, technical efforts will be redirected from research to a functional demo, a draft proposal to the NSF is co-authored for supplementary funds, and GEON's new webpage is collectively edited. During synopsis and action, several PIs, administrators and an ethnographer are capturing in notes, recordings, and during the later years in archived webcasts. Lists of 'action items' are later distributed collating the reports, decisions, and responsibilities. Today, webcasts, meeting minutes, recordings and an ethnographer's notes still maintain traces of these meetings.

What is small and what is large here? The technical design, programming and testing of various SDSC members over months is shown to us in ten minutes on a projector screen the size of a wall. So too are the NSF plans for CI which dedicate ten percent of a 700 million dollar budget to Geoinformatics. Technical developments come

[13] Scale and 'organizational,' 'institutional' and 'technical' are not actor's categories, as such. The term scale is used in a related meaning, grounded in computer science, referring to the size of a computing system: 'scaling-up' or 'scalability' referring respectively to growth in users, computational capacity or geographic scope, and to designed capacity to increase these. Technical is also an actor's category, again mirroring the use in this dissertation, but usually referring to the knowledge of geosciences or the development of information technologies.

to be the material for securing funding from Geoinformatics. Technical and institutional scales come to be tied in the 'flatland' of situated meeting interactions.

Commentators have often framed ethnomethodology as the microfoundations of macrostructure (Collins 1981; Giddens 1984), but this mischaracterizes the treatment of scale in ethnomethodology. More accurately, Taylor and van Every have wrapped together ethnomethodology, ANT, enactment theory together within their own approach and called it 'flatland thinking.'[14] These approaches are contrasted with 'established organization and management theory' which take organizations to be entities outside and apart from the practice and communication in its everyday:

> Established management theory treats organization as structured entity, of which the best-known manifestation may well be the notorious organization chart […] This is accomplished by according […] a privileged ontological status to organization, which is implied to have an existence autonomous of communication, in that it serves as the container for the latter (Taylor and Van Every 2000: 141-2)

Flatland thinking rejects reified notions of micro-macro – "organization-as-entity/individuals-as-component" (Taylor and Van Every 2000: 143) – and focuses on the everyday *achievement* of scale. We should not take organization to be a scale outside of the constitutive work of participants in its everyday enactment. Organization is not a 'meso-scale' phenomena with it own emergent properties (c.f. Vaughan 1999), rather it is *performed as such* in member's accounts which seek to i) summarize environment or organization, ii) speak for the organization, or iii) or put into action an organization (Boden 1994). "[T]he purpose of this methodological stance [...] is not to legitimize one

[14] See also Latour in his extremely cheeky treatment of the same notion (Latour 1995).

level of structure at the expense of the others, but rather to examine social practices whereby structure is made to happen, made to appear i.e., accomplished by and for members of society" (Hilbert 1990:795). These accounts are formulated in talk, or inscription, and are the modes by which the organization is constituted:

> organization must be found at a single level – a flatland – which is invariably situated, circumstantial, and locally realized in a finite space, involving real people. (Taylor and Van Every 2000: 143)

Scale then becomes part of the repertoire in sense-making. "To work with the idea of sensemaking is to appreciate that smallness does not equate with insignificance. Small structures and short moments can have large consequences." (Weick, Sutcliffe et al. 2005:414). Institutional demands and technical developments are worked through and then together in the framed interactional 'single-level' of meetings. It is made possible by ordinary reasoning, and the summative, statistical and calculative tools (Callon and Muniesa 2005) brought to bear on the task: 'community opinion' is known by its spokespersons (Latour 1987); 'geoscience knowledge' is captured in concept maps and ontologies; 'GEON's growth' is measured by website hits and its 'size' through an increasing list of partners.

Social scientists can also participate in organizing the scales (Callon and Latour 1981). Notably, I have often summarized my findings about 'the geoscience community,' 'relations between IT and domain' or 'emerging routines in ontology development'. Also I have referenced various findings of published studies in infrastructure development, social informatics and multidisciplinary collaborations and brought them to bear on the 'larger picture' of GEON activity. Other GEON participants have also brought to bear studies of the geoscience community, collaboration, or technology deployment (e.g. Ch. 2.2 on the KDI study). These have all been the methods by which in everyday GEON activity the institutional dynamics of the NSF and the difficulties of technology enactment can be known and made tractable in practical organizing.

In meetings, GEON's activities in the technical can be linked to institutional demands; both can be known and made tractable through the spokespersons and representations brought to bear in situated action. In the next section we will more closely inspect how GEON's geoscientists come to stand-in for their community, and particularly as archetypes of 'geoscience users'.

4.2 – Of Infrastructure without Users

Arguably the most important outcome of GEON is to be the organizing site for a new collective of geo-scientists regularly drawing on its computing resources to conduct scientific research through mapping and visualization, collaboration, data integration, processing and so on. However, as an emergent infrastructure GEON begins with only planned uses and no users. The question then becomes how to build an infrastructure without knowing the final set of its uses? The intended uses are defined in advance: visualization, computing and data integration. The intended users are also clear: the geoscience community. But outside these definitions uses and users begin as empty categories. Visualization of what datasets and with what forms of representation? Data integration of whose data and to conduct what research questions?

The geoscience community, as such, cannot speak for itself; it requires representation of its extant practices, emerging needs and future goals. This section outlines GEON's strategy in its first years for coming to know its intended uses and users. At GEON's formal inception information technologists found an already available reserve of representatives: the geoscience PIs and their research teams. The PIs came to be surrogate users, speaking for the geoscience community as representatives, spokespersons and intended users.

A familiar dynamic was (re)produced. Geoscience PIs would stand in for the diversity of data forms, research practices and future uses, and information technologists would integrate a common infrastructural base across these serving the earth sciences.

As we have seen the Atkins Report defines the primary interest of CI as supporting 'the conduct of science'. This is also 'outside the range of the report'. The conduct of science is a specific matter and an empirical question. CI is not built on informational blank slate, but rather on top of existing heterogeneous datasets, specific research traditions and established channels of communication. While CI is intended to transform it must do so by supporting existing practices; the philosophy of technological change is of *transparent* transitions across *usable* systems. Users come to use CI systems that resemble those they are familiar with, that support their current research and then are *migrated* to the latest technologies[15]. But if the conduct of science is specific, and CI must support these, how are they to be known?

--

In the Atkins Report the question of how extant practices and resources will come to be known – how the CI will support both existing computing arrangements *and* a

[15] The strategy of supporting existing practices is represented in the two tier approach, but is best exemplified by knowledge representation. The functioning of a knowledge representation system is premised upon drawing on the user knowledge base, language and even work practice. See next chapter.

future interoperability – is provided in silhouette only: with broad brush strokes the Report encourages collaborations of domain, CS and social science. No specific methods or even roles for IT, social science and domain are outlined. However, this is not because uses and users are considered a solved problem within CI circles. The problem of uses and users is an active site of research, debate, and continuous innovation of method in both the academy, and in the software industry: "Improvement, of course, is crucial to an industry as vast and unpredictable as software – with most systems delivered late, over-budget and failing to meet what are seen as user needs" (Mackay, Carne et al. 2000:99).

The question of uses and users has been addresses with a startling variety of conceptualizations and methods. Three ideal types of approaches presented below fall on a rough spectrum from isolating users from uses, seeking users to inform uses and isolating uses from users: i- the user can be taken as the object of research, as in human-computer interface, here particular uses are bracketed out in favour of understanding 'users in general'; ii- the user can be the source of data for the development of uses (or functionalities) as in requirements testing, or as a contributor in the design process as in participatory design and; iii- the user may be considered a problem, refusing novel functionalities or as 'resisting' imposed change as in studies of technology uptake[16]. Even

[16] Within computer systems design see Friedman (1989), for requirements engineering Jirotka and Goguen (1994), for studies of resistance see (Hirschheim and Newman 1988; Miles and Thomas 1995). There are multiple scholarly communities engaged in the question of design, users and uses: for CSCW see (Schmidt 2000; Prentice 2005) for participatory design see (Schuler and Namioka 1993). Rather than contributing to the debates STS scholars have taken the various methods as object. For a review of theoretical approaches see (Oudshoorn and Pinch 2003), for studies of requirements testing see (Mackay, Carne et al. 2000), for HCI see (Suchman 1987), for resistance see , for a history of the thinking about the user see (Bardini and Horvath 1995), for users in large community systems (Star and Ruhleder. 1994), on gender as shaping user technology see (Cowan 1987; Martin 1991), on gender as shaped by user technology see (Kammen 2003; Oost 2003). Finally STS scholars have made explicit contributions reconceptualizing users/uses and offering new design approaches (Suchman 1994; Bowker, Star et al. 1997; Dourish and Button 1998; Dourish 2006).

a casual glance at the IT journals with any focus on implementation will divulge an ongoing debate over the framing and methods to approach uses and users. Rather than eliding the problem, the Report's silence on particular methods for knowing uses and users should be read to reflect this ongoing debate. Atkins leaves open the question of a particular investigative approach and instead encourages a syncretic combination of methodologies found in social science, IT and CS.

In addition to gesturing to these methods the Atkins Report contains another, implicit, solution to the problem of users and uses which heavily mirrors what comes to be GEON's own approach. We have already outlined this in the form of a 'bottom up approach.' CI projects are to be spearheading by domain researchers. ITR came to be an opportunity for enacting this model (see Ch. 2.1):

> Only domain scientists and engineers can revolutionize their own fields. At its core the ACP involves rethinking the processes and methodologies underlying individual scientific and engineering fields. Domain scientific and engineering researchers must step up and enthusiastically create and pursue a vision. (Atkins 2003:50)

Domain scientists should take leadership positions in the planning and implementation of CI for their communities. This is a model of the scientist as a knowledgeable and able representative of: domain needs, existing resources, established research and communication practices, and aware of the cultural properties of that community. However, in the Report, the domain scientist and a knowledge of the contemporary domain configuration is insufficient: "Experience has shown that simply automating existing methodologies and practices is not the most effective use of technology; it is necessary to fundamentally rethink how research is conducted in light of new technological capabilities," (Atkins 2003:15). Alone, the domain scientist is immersed in a field of extant methods, research questions and technical configurations. The domain scientists must be also be coupled with information technologists: "The substantial and ongoing involvement of information technology specialists is required to ensure that innovative new uses of technologies are identified" (Atkins 2003:50). This is the IT&CS expert as innovator, as able to envision changes in domain method and everyday practice and as able to model similarities across disciplinary difference. Together computer and domain scientists are able to bridge existing research practices and novel technologies, eliding resistance and unusable software outputs, and reconfiguring the scientific user.

--

Woolgar, using a modified semiotic approach, takes the IT designer as a *writer of users* who by programming software produces "a machine that encourages only specific forms of access and use," (Woolgar 1991, p.89). His empirical ethnographic focus is on expert communities of designers and computer programmers and how these groups inform the 'configuration of the user'. This notion describes how in managing the process of software design these programmers will create models of the user, imputing interests, goals, tasks and methods. Models are produced by various 'testing' procedures such as surveys or experiments conducted on samples of users (Beynon-Davies, MacKay et al. 1997) and resulting in reductionistic statistical representations of 'the user'. These models come to inform the programming of software products "defining the identity of putative users, and setting constraints upon their likely future actions" (Woolgar 1991:59). Representations of the user come to inform the design of software, its programming, and those features made available to the user. As these user models are encoded into the software actual user's choices in practice are curtailed. This is the configuring of the user.

Woolgar's analytic approach has been extended by Mackay et al. who explore novel approaches in software engineering for producing knowledge about users such as requirements testing. Woolgar's study implicitly follows the 'waterfall model' of software design, popular in CS at the time. The waterfall model advocates an orderly progress in design and implementation of software: "stages are executed sequentially, as opposed to iteratively, with each step being completed – for example, by 'signing off' – before the next commences" (Mackay, Carne et al. 2000: footnote 20 p 754). Within IT circles the waterfall model has come under general disfavor and is now thought never to have reflected actual practice. A formal organizational model of contiguous and chronological steps was always coupled to practices that crisscrossed the linear 'waterfalls'.

In MacKay's approach this linear model (designer-configures-user) is complicated by iterations of designer/'user' interactions, where roles for 'users' is expanded to include a company hiring the programmer ('the boss') or a participant in design and future client ('market'). Rather than simply configuring users through software output, designers too are configured. Designers are constrained by the demands of their clients and by their organizational settings. As with Woolgar, configuration is partially conducted through the reductionistic survey representations of users, but MacKay also includes the organizationally specific methods of requirements testing and, most importantly, programmers' accountability to those who hire them. As business clients, future users configure programmers by evaluating work in progress, rejecting particular software design or requiring updates following project completion. Mackay describes this as an actor-network which collects a market, users (or consumers), designers, and the designer's organization.

Even MacKay's iterations of configuration appear too formal to describe much of GEON activity. Instead I will again rely on the sense achieved in lamination. The models of geoscience users are built up over years of characterizations and are constrained by uses which must be made to meet definitions of computer and earth science basic research. The GEON proposal contains a loosely described set of uses for GEON (or in actors' terms 'functionalities') as defined in the architecture: computing and networking systems, knowledge representation and visualization. In vague outlines these come to configure GEON's infrastructure development goals and the driving basic research questions of geo and computer science.

The more detailed representations of the user are often crafted very casually by geoscientists in meeting talk, through feedback on the continuous stream of software demos, and by computer scientists as they observe interactions of diverse geoscientists. Models of the user are put forth in the characterizations of the geoscience community: "we are fieldworkers, we're not used to standardization from above" Geo PI; or endogenous comparisons: "we don't all go out on boats together, like the oceanographers, or sit in a room with ticking tapes telling us the temperature, like the atmospheric guys, we all do our own thing, and we all go out there [into the field]" Geo graduate student. In AHM or PI meets geoscientist may come to debate these characterizations amongst themselves, while in the disciplinarily defined confines of a concept space meeting they may stand uncontested (see Ch. 5).

The range of ways in which geoscience PIs come to stand in for a geoscience community is as diverse as the avenues for interaction within GEON. They occur at the variety of meetings (e.g. IT focused workgroups, knowledge representation focused concept space meetings, or collective AHM) and in their interstices through email, phone or video conferences.

At times, when geo-scientists and IT experts are closely collaborating geo-PIs will *actively* speak for the community. In these cases they are able to represent:

i- *the communities' technical requirements:* such as computing resources, interoperability of datasets or standards for representation: "we need to develop GIS layers that match up with the ones that USGS is using" Geo/GIS expert;

ii- *the communities' informational competence*: such as the ability to use particular software suites, web services or even data collection tools "remember, I represent the low-tech end of the geological community, and there's a lot of us" Geo PI;

iii- *the communities' resistance to technical change*: such as making data publicly accessible, or unwillingness to take-up a software suite that is not user friendly "you can do the most with ArcInfo, but it has such an enormous learning curve that not very many people ended up using it" GEO PI[17], or ;

[17] ArcInfo is a command line interface GIS program. The implicit contrast is with ArcGIS, which is has a graphical interface.

iv- c*ommunities' demarcation of science or science-contributions*: such as what counts as geoscience, and what may be considered useful but distinctly not science: "all this metadata sure will do us a lot of good, but we're not going to get any credit for having done it." Geo/Visualization Member.

The notion of lamination speaks to the interactionally developed local meaning of these user characterizations, while also pointing to their persistence as they are layered over time. Even if always in a situated manner characterizations such as 'geoscience as fieldwork' or 'metadata as not science' come to accrue greater weight *as topic* as they are deployed repeatedly. Boundaries may be debated or contested in action ("data is nothing without metadata and we need to make the NSF and everyone else knows that" GEO PI) but in the temporally shared interactional space of GEON the debate becomes increasingly loaded with laminations. Models of the user are built up over time.

The notion of configuring speaks to the encoding of user models into *a software program*. Particular software tools will be developed rooted in these models, configuring the user. However with infrastructure we must broaden configuring from a 'single piece of software' to the choices of which software, what functionalities, and what linkages between these. GEON, and CI more generally, are not single-use-applications; the plan is to provide many tools and resources to a broad swath of the geoscience community. While characterizations of the geoscience community do inform single software tools, decisions are also being made about what software to support or what data to integrate. With information infrastructure a future user community is configured not only by programming a particular software tool, but also by what software tools are supported and whose data is integrated. Two examples are detailed below: the generic mapping tool (GMT) visualization suite, and legacy data in metamorphic petrology.

--

In GEON geoscientists have pushed for integration with GMT, an established data visualization and mapping tool used within the geosciences (GMT Accessed Aug 2, 2006). GMT is freeware, open-source, and popular within many of the geoscience communities. However it is not interoperable with GIS, a major technological thrust of GEON's technical initiative. Interoperable here means many things, but in short, GMT produces static image such as maps or contour drawings which are not interactive. No further information can be elicited from the image once it has been rendered (as with a jpeg image[18]). In contrast GIS' layered structure allows for re-sizing, re-rendering and linking to other datasets. A click of the mouse on a GIS rendered map could provide further data on rock type, age or history of seismic activity.

However, the suitability of GIS has often come into question within GEON. Firstly, for its technical capacities relative to the geosciences and, secondly, because GIS software is proprietary and licenses are expensive. GIS is primarily a geographic rather than geological software suite and lacks many features deemed necessary for earth science research (which are available in GMT, a tool long used within the earth sciences). One example of this debate within GEON collected around the focal mechanism 'beachball' function.

[18] GMT has multiple output formats, but the 'standard' is an encapsulated PostScript File (EPS).

The beachball is a standard visual trope in seismology, used to represent a best estimate of the ‘slip’ in an earthquake and the orientation of the fault on which it occurs; this is the focal mechanism. The positioning, size and orientation of the beachball represent the center of the earthquake, the magnitude of the event and the stress orientation of the intersecting planes, respectively. While GMT supported this key function, available GIS tools did not; this is the case with a variety of geological visual tropes.

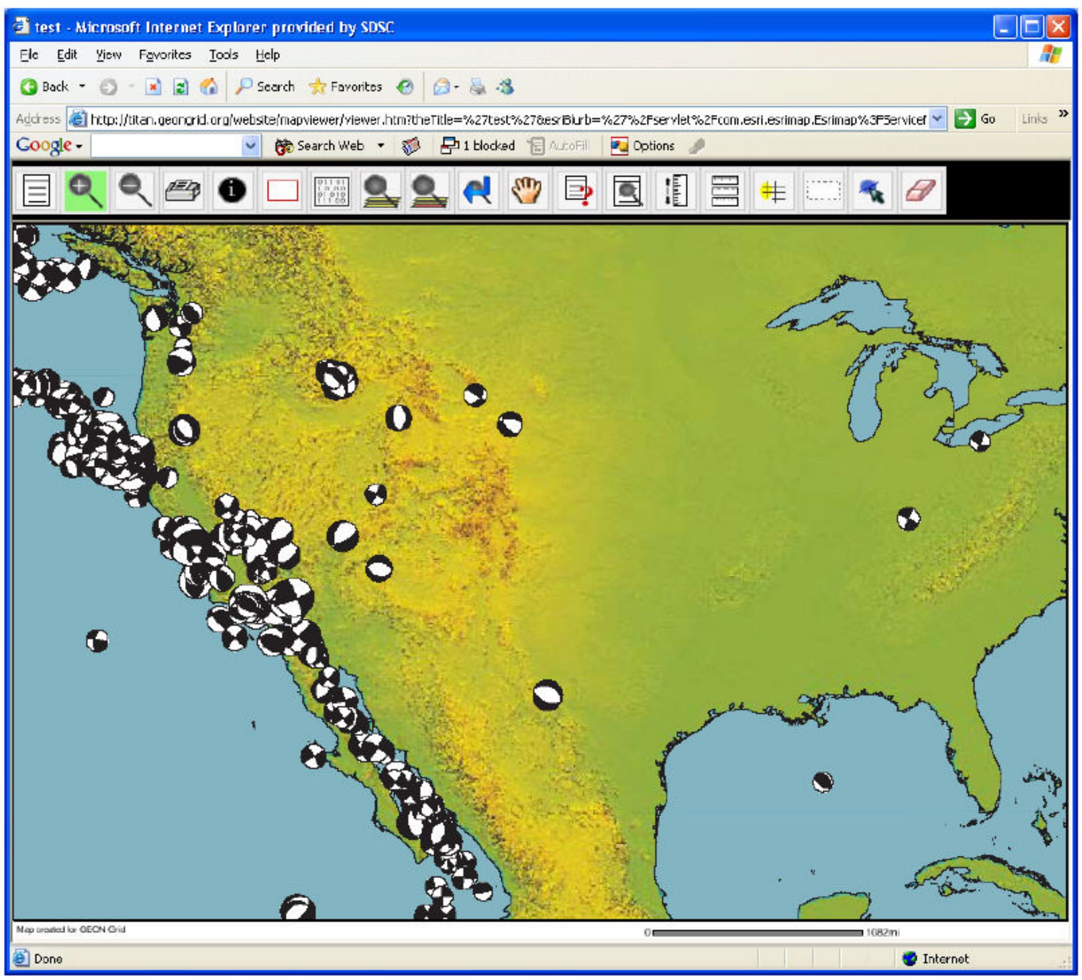

Figure 11: The 'beachball' representation of focal mechanisms.

In a series of debates around GEON's second year participants debated the developmental trajectory around mapping and visualization. Participants could chose to support an existing user base for GMT. For example by using GMT as the default program for its visualizations outputs and providing web services or data integration tools. Alternately, GEON participants could also produce tools for GIS which match the functionalities of GMT such as the beachball. On occasion in these debates Geo PIs declared themselves as unwilling to make a switch as long as GIS did not support already established uses. If GIS could match these functionalities this resistance could be overcome. Meanwhile other PIs characterized the community as unwilling to go through the learning curve to switch the more complex GIS programs. IT programmers suggested various solutions to this, including designing interfaces that mirrored GMT's, or creating integration tools that facilitated the use of GMT with GIS software suites. This is the final approach adopted by GEON, GMT and GIS would be linked through a *workflow* (see appendix A) which facilitated communication across the software platforms. This solution fits neatly with the computer science goals of IT-PIs, but as a high-end development project has involved months of programming, and remains in the beta stage.

The result is a rear-guard effort on GEON's part not only to provide data interoperability with GIS suites, but also to ensure GIS is able to match the services of the 'outdated' GMT visualization package. Over time, and through various technical approaches, GEON has come to support the beachball visualization by combining both software suites through integration. A workflow permits the linkage between GMT's beachball visualizing and GIS' layers function:

> This workflow demonstrates how, using ArcIMS [GIS] services, data sources of various formats can be queried uniformly and later be assembled together to a map. Heterogeneities include structural and representational differences as well as semantic differences (Jaeger, Memon et al. 2005)

. The tool has been years in the making, and is yet not considered usable by a broader geoscience constituency. However the tool is 'infrastructure' in the sense that it is not task specific, and can be adjusted to fit different data profiles. In the long term it will become part of the workflow suite developed as a part of GEON.

The investment in time and expert SDSC resources required for producing this tool forestalls any simple consideration that GEON can simply 'integrate' in all cases. Choices must be made as to what software and visualizations to support. Integration is a costly solution; the choices over what to integrate have downstream consequences for users who must rely on available resources from 'the geosciences network' or invest in their own costly solutions. Users are configured by what software solutions are chosen within GEON *for* the geoscience community.

Above *what* is chosen to be integrated, the particular integration solutions also have downstream consequences. These downstream consequences are characterized within debates by GEON participants themselves. For example, in combining GMT and GIS tools, a complex linkage between open-source (free) and proprietary tools have been (literally) framed into GEON's architecture.

GIS software is substantially proprietary. The most significant GIS software provider is ESRI, a company with close ties to GEON (for example, co-sponsoring Geoinformatics '06). ESRI produces such popular software as ArcGIS, ArcInfo, and AutoGIS. With the extremely limited user base of contemporary GEON (i.e. a few geo-

PIs and their teams, most of whom already own licenses) ESRI has agreed to waive licensing requirements. But as the user base expands will the free licenses also be extended? This has been the occasional site of debate: whether GEON should rely on proprietary tools. Speaking for his community a geo PI notes “we just don’t all have licenses, if GEON is for all of us, then we shouldn’t go that [GIS] way.” If ESRI begins to demand licensing fees as GEON expands will it be possible to find funding support for its constituency or will this be the responsibility of individual researchers? Or, will it be possible to pull back from a heavy dependence on GIS configured in the development cycle? The geoscience user community is configured by the choices of software support in developing an infrastructure.

The problem of technology change and adoption is made endogenous in meetings, a differential model of uses and users is generated in which i- GMT has greater functionality for geoscience than existing uses for GIS, ii- GMT has the momentum of acquired skill sets and the community may be unwilling to learn a new skill set for GIS iii- GMT is free and open source, while GIS is expensive and proprietary. While not spoken in Woolgar’s terms the endogenous implications are the same, the Geo PIs are debating the future configuration of a GEON user community: will they use existing software suites or be migrated over to a new one? Will they use open source or proprietary software? Will integration of proprietary and open-source tools force particular software arrangements on the community? Will a subsection of the community be excluded from GEON by such a choice?

--

Models of the user also configure a continuity with the past of geoscience data. For example, how to manage the masses of undigitized legacy data within the various sub-fields that are GEON's explicit constituencies? A geo-PI with a specialization in metamorphic petrology has repeatedly characterized her field in meetings noting that the data in her field is scattered throughout a multiplicity of print journals, dissertations and filing cabinets. This data may not be available in a structured database, or even in a digital format. There is no mandate or funding within GEON to digitize data, but how can GEON produce advanced IT tools for this specialty when there is no data upon which to operate? GEON PIs have been spending considerable time in securing funds and expertise to make digitization of their data possible, however, the cycle for funding, organizing and then actually digitizing legacy data will be years in the making. Thus this is an impossible 'bootstrapping' operation for GEON within metamorphic petrology. In an endogenous comparison the geo PI characterizes her community relative to geophysics: "they have been working with remote sensing for years, decades! We don't do that, all our data is on paper, all of theirs is in databases!" This PI worries that the result may be an uneven development of integration and computing technologies across the disciplines represented in GEON.

The problem of legacy data is made endogenous in meetings. A differential model of uses and users is generated in which there are i- users with long histories of digital data and database experience and ii- users with traditions of fieldwork, small datasets and little digitization. Future users are *understood by participants* to be configured by the investment of technical resources in disciplinarily specific resources such as digital data and integration tools. The debates in GEON are about future users and the allocation of resources across user communities: which communities' research will be supported by this 'geoscience network'? what are the implications of an uneven disciplinary development?

Woolgar's notion of configuring the user is correct. In designing software a programmer requires models of the user, and these models come to be encoded in the software itself informing preferred user action. MacKay, in complicating Woolgar's linear model, is correct. In designing software a programmer too is configured by user representations such as surveys or by more active interventions by users participating in the design process. MacKay is also correct in pointing out a greater range of actors: not merely designers and users, but also 'a market', 'clients' and the programmers 'organization'. Here I have added three new dimension to the process of configuring.

1- In the close quarters of GEON meetings user models are built up interactionally over time in laminations as geoscientists work to characterize their own communities. They are surrogate users, at times speaking for themselves as individual users and at times speaking for their communities. The range of activities that model the user are as varied as the forms of interaction such as self-characterizations, cross-characterizations and endogenous comparisons. The actors mobilized in configuring are more than designers and users ('user needs') but may also include established usage patterns, legacy data or resistance, to name a few.

2- With large scale infrastructure 'configuring' must be expanded beyond considering a single software program. In addition it is the range of software that is supported. In configuring infrastructure choices are not only about design, but also about what to design, support or integrate.

3- Although the theoretical term ‘configuration’ does not appear as an actor’s category, the stakes of configuring a user base are part of the endogenous methods of actors. That is, members such as geo PIs are conscious of downstream consequences for the implicated users (Casper and Clarke 1998) of their communities. They discuss explicitly, in situ, how expensive or proprietary software may exclude entire disciplines or that underdevelopment of technologies relative to specialties could mean a lack of resources for that specialty in the future. The consequences of software design choices (or choices of what software to support) are at play within actor’s deliberations.

Below I follow more closely the interactions with an emerging object of GEON’s architecture: the webpage and portal. The portal is GEON’s ‘front end,’ what many users will rely upon in order to learn about GEON, its services, and available datasets. A draft of the portal was presented to the GEON team at a PI meeting and a lively discussion ensued topics ranging from the purpose of GEON to the future users of its services. The portal became an ‘occasion for identity’ as information technologists sought ‘feedback’ from geoscientists. By following the dialogue closely we can see the production of the boundary IT/domain as computer scientists seek feedback about their draft portal from geoscientists. We can also see the production of the boundary domain/domain through geoscientist’s efforts to differentiate amongst themselves as diverse users with particular requirements.

An Occasion For Identity: (Re)Designing the GEON Portal

On October 24th 2004 at the University of Idaho in Moscow, in a late morning session of a PI meet, a draft of the GEON website and portal[19] was presented to the team. Two years into the project, there was already a website, but this new version would approach the "mature GEON" with an established physical architecture of nodes and offering tools and services to a user community[20]. The website was to be unveiled at the upcoming Geological Society of America (GSA) meeting in November, but first this draft would be presented by the GEON IT team for the geoscientists to provide "feedback". This discussion would become an occasion to discuss the "identity" and "philosophy" of CI, GEON as an organization and service provider for the geoscience community. Together these participants devised a complex user model which has since informed the design of the GEON architecture and configures its downstream users.

The presentation of the website *by* the IT team *to* geoscientists enacted the boundary IT/domain and enabled an organization of a very informal 'requirements testing' session. As surrogate users, geoscientists would "provide context" to the IT team, informing them about the response of disciplinary scientists, the geoscience community or the average user. Geoscience participants ranged freely in style, speaking for themselves, for GEON as a project, for their specific disciplinary communities, for geosciences more generally or for a generalized 'user'. Information technologists too

[19] I will use the terms portal and website interchangeably, as do the actors. More specific meanings are described in Appendix A, however these tighter uses are often limited to CS&IT only gatherings (see portal and portlet). In this discussion portal also refers the login-required 'inside' of GEON's services.
[20] No particular tool was beyond the 'first tier'(see 4.3) at this point, however demos and draft ontologies were available for inspection and limited use. The portal and website were a framework for future access.

came to speak for the contemporary aesthetics of web page design, and in the name of generalized users.

The *boundaries* in this discussion were (symbolically and physically) framed by the interactions themselves. The outcome is familiar but specific: SDSC IT team / Geo PIs. In over two years of GEON PIs working together this particular boundary had become rote, but as with all laminations it requires a situated enactment. That is, here at this PI meeting it is the 'GEON IT team' that is presenting their output 'draft website' to the 'Geo PIs'. The interaction was structured into the agenda: a five minute walkthrough by a lead IT PI of an IT product would be followed by discussion: "GEON Portal and Homepage – walkthrough and discussion". The boundary IT/domain was complemented with a communicative form, establishing relations across boundaries in the form of feedback and user scenarios.

I have argued that the IT / domain boundary has a nested articulation of finer grained boundaries across the domains. This is the domain / domain boundary. GEON is mandated produce a common informational substrate for the *diverse* geosciences. The domain / domain boundary comes to be enacted in the articulation of disciplinarily specific user models. For example, we will see in this discussion that geomorphology comes to be characterized by particular methods (such as LiDAR), representational forms (such as DEMs) and desired functionalities (such as simulating the removal of flora). Disciplinary difference is articulated through boundary work demarcating geomorphology's methods, representations and uses. As the boundary domain / domain is enacted geoscientists negotiate amongst themselves the means for representing disciplinary specificity and difference across these.

Figure 12. Snapshot of the GEON homepage as it was presented to the PI meeting on an LCD projector.

Establishing the GEON Portal In Situ

Those present at this PI meeting ranged between 16-24 participants, depending on the particular session. Most were GEON PIs, or regular members. Here I will focus on five actors who contributed significantly to this discussion. Below I elaborate a brief description of their training and some of the laminated roles they have developed in GEON over time:

- *Chad* is a lead IT PI from the SDSC. He is also leading this discussion, first presenting a 'walk through' of the webpage and portal, and then lightly facilitating the following discussion.
- *Wendy* is an earth scientist Geo PI. Over the years of GEON meetings she has regularly crafted herself as the representative of 'low tech' geoscience: field research, with few high-end computing tools, little digitized data, low training in informatics skills and limited computing experience. In this discussion she regularly spoke in the name of the 'average low tech geoscientist'.

- *Kris* is a lead Geo PI. Often playing the role of the 'technology champion' by advocating revolutionary visions of CI. In this discussion he primarily spoke in the name of a grand vision for GEON, Geoinformatics and CI.
- *Roman* identifies as sitting at the boundary of IT and geoscience. His expertise is in GIS and knowledge representation. In this discussion he speaks in the name of both a generalized users and the more specific but still generalized geoscience user. Roman also comes to speaks in the name of both by beginning to define a GEON user typology.
- *Mark* also identifies as sitting at the boundary between IT and geoscience. His expertise is in digital libraries for the earth sciences. In this conversation he intervenes several times to recap and re-orient the conversation. He provides summaries of research findings in social science and HCI, encourages 'requirements testing' through close interviews with PIs, and arranges an on-the-fly exercise in the meeting to capture what users might want on a GEON homepage.

In the background of these discussions two participants (the administrative coordinator and an SDSC computer scientist) were taking notes and collecting action items about the comments and suggestions being exchanged. Later these notes were circulated for discussion amongst PIs via email (which in this case did not reignite discussion), and also informed the rewriting of the homepage by IT participants in later workgroup meetings. This analysis will not follow the inscriptions and discussions that emerged from these interactions, however, the outcomes in the form of a new website design are summarized at the end of this section.

Chad began the session with a quick walkthrough of the website. A walkthrough is a verbal description of the website and portal as each link is followed on screen. As participants sat around a large oval conference table the website was projected at the front of the room on a large screen for all to see. Many of the links were unpopulated and Chad's description filled them in. The walkthrough was uninterrupted by discussion. Below is a summary of the links which extend from the homepage based on Chad's description (table 2).

Table 2: Links, and decription of content, from draft GEON website presented at PI meeting.

Homepage Links to:	Content of link
About	**A brief description of the GEON's mission, participants, partners and history.**
Research	**The specific 'basic science' research in GEON. This link leads to further links: IT Research, Rocky Mountain and Mid-Atlantic test bed.**
News	**More news about the development of GEON. Each news item links to greater details about the item.**
Education	**Unpopulated.**
Search	**A text search of the public webpage. (not to be confused with GEONsearch, mentioned in discussion but still unpopulated)**
Portal	**Login/password required. Mostly unpopulated. To contain the majority of GEON resources, such as data registration, access to mapping and visualization tools and 'GEON search' for finding data, resources or integration tools (such as ontologies).**
Quick Links	**More links. Primarily to other CI or Geoinformatics projects such as CHRONOS or BIRN.**
GEON Forums	**Discussion lists, Q&A, and selected archived GEON list-serves.**
Calendar	**Upcoming events relevant to GEON or Geoinformatics**
Tips and Tricks	**Unpopulated.**
Image of the Month	**A rotating image, this links to an archive of previous images.**
Image	**The image itself links to a site offering greater technical details. In this case to Geo PI Ramon Arrowsmith's website.**

Establishing the Uses of the Portal along the IT / Domain Boundary

Following the walkthrough the discussion opened with Chad stating "We're going to spend a full hour on the portal. Providing the SDSC direct input in what the geoscience expects to see there. So that when the next full blown version comes up online, it's acceptable to everybody." No particular (explicit) structure delimited the discussion, "everything is up, everything from the tiniest details to the big things." And with this announcement a flurry of criticisms emerged from present participants.

I will elaborate the topics of discussion, but I will not reorganize these comments for the reader. They emerged as the topically scattered impressions of participants. Then, as an endogenous feature of that conversation, the discussion came to be organized in explicit processes of summation and structuring. In other words, participants engaged in the work of organizing comments and criticisms in situ, and this in turn came to shape the outcome of the interactions. The user models developed which came to inform the final portal design were produced in the situated organizing of conversation.

Commenting on the website Kris began with his impression of the “spirit of the homepage”:

> Kris: I'm not sure this represents GEON. This represents an activity *within* GEON. This is ‘news’ ... and ‘links’ ... and an image. And none of that is *the GEON cyberinfrastructure message*. That we are the people that deal with the cyberinfrastructure. So people come into this site, see ‘home’, ‘research’ and all that but nothing about *the cyberinfrastructure vision* that GEON is supposed to represent.
>
> Chad: what do you think is missing? what do you want extra?
>
> Kris: Well one thing is different that I've tried to do on my websites is saying ‘GEON: Cyberinfrastructure research for the geosciences’ let’s make it exclusive to who we are. The news changes.

The first comment immediately opens the discussion into the identity of GEON: what is it as a message and as a vision? The draft website does not define GEON except by expanding the acronym into ‘The Geosciences Network’. Notably it was not the IT team, in designing the website, that pushed for the use of the term cyberinfrastructure; rather, it is a geo PI. The term has come to mean something to this earth scientist, and it is important to communicate this to a broader community. This larger meaning is not immediately defined in the utterance, but it has been laminated over time.

The identity of GEON – "what are we doing here?" – has been a regular topic of discussion at the AHM and PI meets, particularly in the first two years. Is GEON a prototype, providing experience for a GEON2, or is it a foundation for a next stage in the project? Should GEON demonstrate its value through immediate 'science results' or by developing tools for enabling future results? Is infrastructure intended to support existing practices or transform them? The purposes, plans and methods of GEON's construction are continuously discussed in the abstract at large gatherings.

In the comments above, Kris is aligning GEON's identity with CI, characterizing it as beyond any particular 'activities' in which GEON is currently engaged. Activities are the news items, what GEON participants have been engaged in *as an ongoing project*. Kris' lamination of 'GEON as CI' is not specified in the above utterance. For the moment Chad simply asks what is missing, however it is further enacted in the discussion over time. The 'CI vision' is only one take on GEON and other participants are quick to intervene on its definition.

Immediately on the heels of Kris' comments Mark interjects, speaking in the name of a user. Following directly from Kris' statement above -- "the news changes" -- Mark's comments redefine a question of identity into a question of usage. News items such as "GEON announces Formation of Advisory Board," are about GEON as a project, as an emergent infrastructure. Mark poses a user indifferent to project details, and focused on scientific research. Not 'what is going on in GEON?' but 'what can GEON do for me?':

> Mark: I think, as a user, I wouldn't think about the news part. Almost, really, I'd like it to be clear what I can *do* here. The fact that all that [waving at screen] is 'in the

> portal', and that it's not obvious what's in there ... Those who come with tasks should have them made available.
>
> Wendy: Nobody even knows what 'portal' is!

Mark's model is of a practicing geoscientist user who is not immediately interested in a 'CI vision', or the daily news items of an emerging GEON infrastructure project, but in conducting his or her research. The scientist user is interested in tools or data to assist in an already developed research trajectory. The exchange continues as Mark notes that there are no hints as to what GEON is offering in the form of tools, or particular datasets.

Chad responds that these are in the login-required portal, and also part of the research description. Particular research tasks (such as geomorphology LiDAR research, see below) are supported by IT research in GEON and are described on the draft website as part of each test-bed project. Mark notes "That is so deep though. Not many users will make it that far. And the portal requires a password." 'Depth' here is a measure by how many links a user must follow – in this case four, from the home page->research->test-bed->LiDAR. Depth comes to be defined as something users will be unable or unwilling to follow too far to learn about GEON uses.

Depth-as-hindrance is part of the common sense reasoning of web design[21], but it must be enacted locally and, here comes to be a base argument in the conversation supporting design for scientific users. What may be considered common sense reasoning can be deployed in multiple forms and its particular articulation is often the site of contestation (Garfinkel 1967: esp 38-44, 75). What has remained implicit in the

[21] No citations or studies was required here to support the argument, it was accepted easily by the participants. This said, the finding has widespread acceptance in IT circles and is regular prescription in web-design manuals. For examples of empirical research see (Kiger 1984; Lee and MacGregor 1985; Seppala and Salvendy 1995; Jacko and Salvendy 1996).

discussion so far ('the homepage is important') is articulated in concrete terms: GEON's homepage is important because "people get there first and users are unwilling to follow many links" (Mark); this is even more relevant because "no one knows what GEON has to offer. Why would they follow an unknown trail?" (Mark). A specific user model is outlined in this description: a scientist, interested in continuing their own research, unaware of GEON's capacities, indifferent to project activities, and unwilling or unable to go 'deep' into the system to investigate possible uses.

At the end of the extract, Wendy, speaking for the naïve user, points out that 'portal' is not a familiar term to geoscientists. It is not immediately obvious to the naïve user that the portal will contain GEON's tools, and upon first glance nothing on the webpage indicates this. It is not Wendy herself who is unfamiliar with the purpose of the portal; the notion of the GEON portal has been explained repetitively over the years. It is the access point to the various tools, services and datasets in GEON's repertoire, it requires a login/password and it can be personally configured by individual users ('myGEON'). Wendy here is playing a role as 'low tech geoscientist,' critiquing the use of unfamiliar jargon such as portal. Later in the discussion she reacts to the 'links' section which lead to Geoinformatics and CI projects such as CHRONOS and BIRN: " CI community says nothing to me. So these are links to other cyberinfrastructure sites. But this is very specific. The fancier we are the less its significance." This new user model 'low tech geoscientist' is put off by jargon, disinterested in non-geoscience research or infrastructure more generally.

On the heels of Chad and Mark's discussion on links and depth, Kris once again takes the conversational reigns and steers the direction back to GEON identity, critiquing the draft webpage as a representation of GEON activities, of ongoing project status:

> Kris: That this is not GEON. These are some *activities* within GEON.
>
> Chad: Actually, I don't understand that. Why is this an activity in GEON, and not GEON. What is GEON?
>
> Kris: GEON is much more than news, links and geomorphology --
>
> Chad: -- That! No! Backup backup. So, what we are trying to do is design a webpage that is up to contemporary specs. I think what you are saying and you want to see *everything* in one page.

Chad request for elaboration (or repair (Levinson 1983: esp. Ch.5; Schegloff 1987)) on what is 'activity' prompts Kris to define more specifically. Interrupting Kris' 'what GEON is not' laundry list, Chad shifts the argument to the aesthetics, or 'contemporary specs,' of webpage design. He takes Kris to be arguing for many more kinds of links to be present, and argues in turn that one cannot have 'everything on one page'.

The notion of depth in web design is often complemented with one of breadth. Too much information or too many links on a single page can discourage the user as much as an excessive depth of links. In a similar argument another IT PI later notes that there is "there is a balance between clutter and putting a lot of information there." Pointing at the draft website on screen this PI noted "It seems to me there is still some real estate that could be used better." Studies have shown an 'optimum balance' between depth and breadth (or "real estate") in web design (Kiger 1984; Jacko and Salvendy 1996). Once again, though, the uses of these terms in this interaction remain colloquial.

The appropriate number of links on a website and, how many links deep to go, have been made actively relevant to the discussion through local articulation: the important uses of GEON should be accessible to users with few links and on uncluttered pages. What are ‘few links’ and ‘uncluttered pages’ is only loosely defined, and remains a point for negotiation (primarily) amongst IT participants.

Enacting the boundary IT / domain defines realms of expertise and roles in design. The expertise of the geo PIs is in geoscience and its community. Their role in design is as surrogate users, modelling future uses, research needs, computational competence and so on. The expertise of the SDSC team is in website and system design. Their role is as mediators capturing requirements and faithfully encoding them within future systems.

Establishing Uses along the Domain/Domain Boundary

Thus far the interaction has relied on one laminated boundary in GEON: IT/domain. Relations are established along the boundary in particular ways -- the SDSC team have presented a product to the geoscientists in order to get feedback -- and each group have roles and expertises ascribed to them – IT understands design and capture the domain knowledge and community familiarity expressed by geoscientists. An agenda, a website and the occasioned performance of ‘presenting the website to the geo PIs’ framed two distinct sets of roles for participants based on locally ascribed expertise. The geo PIs will stand in for GEON’s downstream geoscience users, and from this discussion the IT team will capture future uses and encode them within the design of the website and functionalities of the system. The IT team also speak for, or ‘are experts in’, the logistics of web and system design e.g. in the last excerpt Chad forestalled excessive breadth on the homepage because it must ‘be up to contemporary specs’.

However there is another primary boundary in the logic of interoperability and this comes to be locally articulated in the discussion. Following the excerpt above Wendy takes Kris’ comments not as in terms of web design (‘an attempt to include everything’), as does Chad, but as a problematization of the disciplinary specificity of the content displayed on the draft portal. More accurately, Wendy mobilized Kris’ comments to shift the discussion from technical capacity to disciplinary representation.

The image displayed on the draft website is LiDAR data modelling for geomorphology[22] and part of PI Ramon Arrowsmith’s work. Following directly from the previous extract she articulates her criticism, elaborating Kris’ comment and reorienting from Chad’s comment on breadth:

> Wendy: That’s the first page. For people to see. Right now it doesn't say anything about what we're doing. If I knew nothing and I went to this page I would think that GEON is about LiDAR!

Here Wendy is articulating another laminated boundary regularly enacted in GEON: domain/domain. GEON is a basic research project for geoscience and IT, but it is also an umbrella infrastructure project for ‘the geosciences’. GEON’s mission statement, from the proposal, is to ‘enable a more holistic vision of the earth,’ to provide a clearinghouse

[22] LiDAR is ‘Light raDAR’ or Light Distance And Ranging. For geology it offers the generation of very high resolution digital elevation models (DEMs) of land topography. From a high vantage point (helicopter or plane) a laser scans a ‘swath’ of the terrain producing topological data on the territory which is then visualized as a map. A specific functionality developed as part of the GEON project includes data processing tools to remove the tree cover, thereby revealing underlying fault morphology (or ‘base earth DEM’). On the draft webpage the image on the left is the visualization with tree cover and on the right the base earth DEM (Oldow, Xu et al. 2003; Crosby and Arrowsmith 2004; Crosy, Arrowsmith et al. 2004).

for all geoscience data, and tools supporting work across existing disciplinary research trajectories. She has drawn on another aspect of 'GEON's mission' than Kris's CI vision; this is to serve the earth *sciences*. LiDAR, however, is characterized by Wendy as specific to geomorphology rather than as metamorphic petrology, geophysics or paleobotany. Its inclusion on the website is an implicit privileging of their work and an exclusion of the range of disciplinary breadth mandated by the GEON proposal. Her characterization, however, does not go unchallenged.

Chad goes on to note, with some visible frustration, that "the image is rotating!" It is an 'image of the month,' and on another occasion an image originating from another disciplinary specialty will be displayed. "And all the past images can be accessed by clicking on Image of the Month". A brief and slightly heated discussion follows, in which Wendy asserts that she would be immediately put-off by the image, which strikes her as irrelevant to her research: "there is nothing there to even *hint* that GEON might offer me something and what *is* there is *clearly* not for me" (Wendy). Her claim is simultaneously about the identity of GEON (umbrella infrastructure), webpage design (should communicate uses) and a model of the user (specific research interests). Articulating the domain / domain boundary is enacting a cross-disciplinary difference (the methods of geomorphology are not those of metamorphic petrology) and as disciplinary specificity ("the [web]page should speak to me, to my work" Wendy).

After a few tense conversational rounds Roman intervenes with some practical suggestions about the design, and then introduces another laminated concern in GEON:

> Roman: So maybe here we can have this translated into 'mission' right there on the homepage

> Chad: Oh right there on the homepage…
>
> Kris: Yes! I'd like to see a GEON mission statement!
>
> Roman: Instead of hunting around the homepage, because you always go there. Because I hear [Kris] saying that there is still some tension in the geoscience community, that there is money in the earth sciences going to web-design and IT stuff …

This is a tension familiar within GEON (see Ch. 2.3): a community perception that earth science funds may be used for computing research. This topic had been discussed on the first day of the PI meet in a session explicitly focused on 'science results'. A general consensus emerged that the geo PIs had to produce basic geoscience knowledge ("something new about the Rockies") and that these findings should explicitly credit the use of GEON tools and be published in academic journals or appear as conference presentations (such as at GSA). While 'GEON as basic geoscience research' is a common laminated concern, in order to be made relevant to a discussion of the website it required a local articulation. Roman did so by linking the representation of work within GEON with the information communicated on the portal.

The website can be a tool for communicating not only a CI vision, or possible research uses, but also the diversity of PIs empirical research with GEON:

> Wendy: We [metamorphic petrologists] don't have helicopters or lasers. My data doesn't look anything like that! I can't even imagine what GEON can do for me by looking at this page! At least if you explain cyberinfrastructure I'll know that there's something for me. But even that's too fancy. Why don't we emphasize all our research.

Thus far we had seen the occasioned constitution of GEON's *mission as CI* and GEON's *uses for practicing geoscientists* made relevant to the design of the website. Wendy and Kris then broke down the category 'geoscientist' along disciplinary lines, through comparisons of difference and characterizing specificity. LiDAR and geomorphology are scientific content which is too specific and excludes other disciplinary scientists. Even a *rotating* monthly image is too specific for a user unfamiliar with the GEON mission of integration (or the notion *of* integration). Wendy and Kris' utterance formulated GEON as a cross-disciplinary umbrella, and their logic demanded a homepage which faithfully communicated this.

Following Wendy comments Roman identified GEON's empirical geoscience research as a venue for demonstrating the diversity of disciplines. Below Roman initially began by supporting a statement of mission but then concludes by defining two new users and defining new interests for geo PIs. These are the geoscientist user interested in novel research results, and the geo PI who must demonstrate thoughtful spending of funds through domain knowledge production:

> Roman: … That's what's important! Shouldn't we really be supporting the research...? […] Seems to me that instead of news, research should be the lead. You always want to be able to hear scientists that are concerned that money could be a misspent, that are concerned with doing the research. News should be subordinate to research.

Envisioned uses of the website thus far have included capturing and disseminating the GEON vision and enabling use for scientific researchers. Roman is outlining two new sets of interests that differ from enacted user models thus far. The first is a non-GEON geoscientist: a user disinterested in project activities, not necessarily interested in use of tools, but rather in new science findings, and perhaps in evaluating the GEON project by its knowledge outcomes. The second are the Geo PIs, interested in representing GEON as basic research producing research results and demonstrating a cautious spending of geoscience funds. The website has been articulated as venue for disseminating GEON's disciplinarily diverse science results.

We have seen the enactment of the domain/domain boundary[23] serve in articulating a vision of GEON, in generating new models of the user, in marking exclusions and in creating favourable representations of geoscience activity. In these articulations a single disciplinary image on the website cannot serve to stand in for the entirety of the geosciences; as we shall see, this comes to mean that no particular disciplinary ties at all should be represented on the homepage. In these articulations users are configured as interested in disciplinary science results. To see progress in another geoscience discipline is insufficient, GEON must be tied to progress for the diversity of users it is mandated to serve. The design of the website comes to be linked with supporting the scientific uses in many disciplines *and* in generating science results in many disciplines.

---

[23] Note: Interestingly, there is no converse boundary regularly enacted: IT/IT. There are distinct specialties within the IT team e.g. specialists in grid systems do not overlap with knowledge representation, and even within knowledge representation specialists may not overlap across ontologies and workflows. These boundaries can be seen in the distinctions system/knowledge representation/visualization. However I have never seen a parallel boundary work across the IT specialties i.e. grid demanding equal representation with knowledge representation. Actors (GEO and IT PIs alike) have commented on this and accounted for it in various ways such as noting the extended collaborative experience at the SDSC ("they all work together on lots of projects together, in the same place, for years!" Geo PI), or; by essentializing computer science ("they all work with the same stuff, information, zeros and ones" Geo PI). This fits tidily within the double boundary work and relations framing interoperability: domain difference requires bridging with a unified informational substrate.

Making Sense of the Discussion: User Models and Configuring the Designer

In describing the 'user models' above I have *elaborated for the reader* what has remained implicit in member's talk. That is, participants have spoken of 'users' but have not formalized user models; up to this point the discussion flowed impressionistically as unstructured feedback. This came to change as participants began an in situ summation and then reorganization of the conversation. While in the discussion above I have provided elaborations of user models as they emerged, it is in the work of members to summarize the discussion that they themselves come to formalize user models. This sense making activity is the stuff of organizing: through the occasioned talk at this meeting a 'community' of geoscientists come to be represented as users of GEON, evaluators of GEON, participants in GEON, or partners of GEON – and in turn the website is designed to accommodate these models of use.

Above, I have covered approximately 15 minutes of the conversation in fine grain. As the discussion continued participants mentioned other criticisms of the website, devised other potential users, and discussed the public/private boundary marked by login to the portal (see 4.1). Approximately 40 minutes into the discussion the conversation began to visibly lag as fewer comments were made, and Chad began to wrap up the session: "so just to sum up – ". At this point Kris interrupted Chad with his own summation of the discussion thus far, but also shifts into describing a link structure for the website:

> Kris: -- Let me just make on more comment. I think the map should be should be right on top. Right underneath that should be GEON research services. You click on that […] and its takes you in, and opens up all of GEON's goods and services to the community

The map is of the US, with partner sites superimposed and marked with clickable links. Kris had argued that the distributed nature and national coverage of GEON should be apparent to users at first glance. More importantly he had trumped Chad's attempt to summarize the website and instead began providing a structure.

The discussion had opened with an explicit 'free for all,' participants were encouraged to comment at any level of detail as they saw fit. However, forty minutes later, expressing a sense that the discussion is beginning to become circular, Chad's arrested summation is superseded by Kris' structural summation: the front-page should include a map of partner sites, which link directly to the scientific research.

Immediately thereupon Mark begins his own in situ summation and reorganization of the conversion:

> I'm a little worried that the conversation is fluctuating between the blue space on the bar, and 'what is GEON?' while we need to step back and say 'what are the guiding principles of what were trying to do, with the website?' and *then* get down to how its going to look.

Over the last twenty minutes of discussion actors have moved from the website into kinds of concerns they have relative to the entire GEON project: an emphasis on research, tools or GEON's mission, the distribution of resources across a public/private divide, what is accessible to web-searches (Google) and so on. In Mark's summary he parses the preceding discussion into i- the 'functionalities' of the website – what should be accessible and linked; ii- how the website looks; and iii- what will be the lines of a restricted access. These three categories have been built from the discursive materials deployed interactionally thus far.

With a gesture of frustration meant to sweep away the muddle of the conversation Mark geared the direction of the conversation towards "functionality" by suggesting to "get rid of" the image of the web displayed at the front of the room and begin "a list":

> the other thing to do is to get rid of this [waving at image on screen] for a second and put in a... and start a list ... show the key concepts of what you want to do, *let us see it*, and start a list; we're going around and around in a circle.

Mark's emphasis on 'let us see it' provokes sympathetic responses from the room, and several geo PIs note that they are not sure they are being heard, or *what* is being heard by the web designers: "write up everything we've said," "put the notes up on the screen". Mark's call to 'let us see it' captures a sense of informational lapse for the participants. While the website image had become a focal point, organizing comments, criticisms and furnishing an indexical reference, it did not provide a mechanism for displaying 'feedback' to the room: what suggestions had been made and how they were being interpreted by IT participants?

The boundary relations in this discussion are symbolically and physically framed and negotiated *in interaction*. The explicitly stated purpose of this hour is for 'one side' to present this draft of a website in order to generate a feedback discussion from 'the other side'; as the discussion leader notes "providing the SDSC direct input in what geoscience expects to see there." The website enabled a particular enactment of these boundaries and structured a means for communication across them. However, as new information requirements were negotiated by participants the website and the form of ‘requirements testing’ came to be reconfigured.

Boundary relations are characterized by particular informational requirements (Star and Griesemer 1989) which are enacted in situ, they are part of a local sense-making activity. For the first half of the discussion the website was ‘what the IT team is presenting’ and for the geoscientists it was ‘what they are responding to’. Framed in this manner the responses of geoscientists become *data* (a user model) for the IT team in their future modifications of the website. Participants are parsed into those producing knowledge/data about the domain, and those collecting this data – together these constituted the locally established informational requirements and a mode for their exchange. These were the information requirements as initially framed. But just as with the particular enactment of IT and domain relations, informational requirements were also interactionally negotiated.

Following discussions using the draft website the geo PIs came to demand a new informational arrangement. MacKay argues users are not simply configured but also come to configure. Through suggestions, and then through an outright rejection of the on-screen projected website in favour of a new explicitly articulated set of user models, the geo PIs configured the designers by articulating a new set of downstream users.

The geo PIs, as surrogate users, insisted on an explicit acknowledgement of their suggestions by replacing the static image of the website and having these listed on the screen. Consenting, Chad replaced on-screen webpage with a blank text-document file. A new boundary object had quickly emerged, which supported the newly established information requirements of participants. Rather than suggestions voiced to a room, and presumably (but inscrutably) written into notes, the ‘user requirements’ of those PIs in the room were voiced then written and collectively edited for all to see.

Quickly typing up a set of suggestions Chad built up a list of remembered suggestions:

> First page: clickage map, drop downs for research, tools, partners, PI institutions
>
> Quicklinks: forums, tools and tricks, technical resources

Without having the chance to complete his notes Chad was again interrupted. In this discussion even the form of the notation had become a point for debate. To Chad’s emerging list, Mark responded: “no, no no, not just what we want to link, but put a list of *for who* we’re creating these links”. Mark’s ‘for who’ is an invitation to structure the discussion not by links from the webpage, but rather conceptually by explicitly defining user models. Roman takes the cue and begins his synopsis:

> There’s two aspects. One is GEON's own research, we have that, and the other is the GEON service. I bet that loads of users are interested in the GEON service, not in the research... Rocky Mountain is interesting for us; we have this to tie us together. We have to justify why we are doing this. But for most users they don't care. They want to do research.

In a single breath Roman has produced his encapsulation of the discussion thus far, identifying two aspects, but also three types of users: the first aspect are services with i) user’s seeking service and the second aspect is research with ii) PIs interested in communicating their science findings and iii) users seeking these new research outcomes.

Chad, rendered completely silent now began annotating types of users on the fresh text document projected on-screen. This list was edited in vivo by the geo PIs. Over several iterations the new list came to be organized by a set of imagined types of downstream users, which includes the PIs, GEON's partners and 'the community' which is further subdivide into users and contributors. Users are the envisioned geoscientists who are seeking tools to work with their data, or possibly datasets to complement their research. Contributors are those users who wish to register their data to the GEON ontologies thus making their own data interoperable with the GEON tools, and possibly sharing their data with GEON community. Below is the verbatim list as it appeared at the end of this discussion[24]:

> For PIs: Research/Findings, communication amongst us, GEON project status and events (news)
>
> For Partners: Links for visibility on the website, links to their tools
>
> For Community: CI tools and services
>
> > Users: tools (visualization, integration), educational
> >
> > Contributors: tools and instructions for registering data

Participants in this discussion endogenously developed a typology of GEON users, and it was primarily the 'user' geo PIs who did so, rather than the 'designer' information technologists. These were built up using the PIs themselves as surrogate users speaking for themselves, in the name of partner institutions or 'average' geoscientists to name only a few of the imagined downstream users. These models had been developed interactionally through the discussion. PIs had been defined as users interested in project

---

[24] This list made occasional appearances through verbal reference or in small group discussion for the remains of the PI meeting. A substantially elaborated version of this list also made an appearance at the workgroup meeting two weeks later as the new website was discussed. This list maintained the user categories as below.

development details and seeking to communicate 'science findings' to potential GEON evaluators. Users were disinterested in project details but seeking immediate tool access to facilitate ongoing research. As the discussion began these user models remained implicit and emerged as unstructured impressionistic feedback on the website, but as conversation went participants began an endogenous organization of that discussion (Boden and Zimmerman 1991). It is the talk itself, as a process, which generated order in these interactions (Moerman 1988) and it is primarily the geoscientists who developed the user models which would inform the configuration of the user.

The tripartite typology 'PIs/Partners/Community' developed in this discussion informed kinds of future uses of the GEON system[25]. The GEON website has undergone

---

[25] Woolgar's semiotic interpretation of *software as text* poses both a relationship *and* a gap between programmer encodings and user decodings. Particular 'readings' (or uses) of software may differ from those planned by designers. Programmers may encode preferred actions into an interface, but users may come to devise unplanned uses from them. The software tools has positivities, but these must be understood as a feature encountered in the specificity of practice. An interface is an *actant in* but not a *determinant of* downstream behaviour. STS researchers have devised various theoretical accounts that formulate the gap between design intents and uses such as Pinch and Bijker's interpretive flexibility (Pinch and Bijker 1984), Akrich's 'scripts'(Akrich 1992), Latour's 'anti-program'(Latour 1991; Akrich and Latour 1992), or Norman's 'affordances'(Gibson 1977; Norman 1988).

Pinch and Bijker emphasize the interpretive flexibility of a technology in its early stages, and the rise of closure as it becomes established with solidified predominant meanings and uses – these uses remain fixed through 'technological framing' which carries these uses through contexts (Bijker 1995). This perspective is often called the social construction of technology (SCOT). Both Woolgar's configuration perspective and SCOT have been criticized for underemphasizing the possibility of interpretive re-invention: that a technological uses could remain unclosed in design or come to be reopened following a period of time (Mackay and Gillespie 1992). Kline and Pinch have addressed these criticisms with their own studies of reopened technological interpretations e.g. the model T automobile motor as a source of stationary power on the farm (Kline and Pinch 1996). Other studies have emphasized a reverse tendency such that technologies are adopted by users in ways completely unintended by designers, in turn forcing the market oriented designers to re-design in order to ensure capitalization(Martin 1991; Fischer 1992) e.g. Martin and Fischer have both shown how the users uptake of the telephone, although initially focused on business, quickly turned to personal matters (such as gossip or eavesdropping (Kline 2003)) and providers were forced to find new metrics and billing methods to be able to make these new practices fungible.

GEON has not reached the stages of development where its innovations have become the battlegrounds of systematic user feedback, interpretive flexibility, or attempts at closure. Thus my analysis is necessarily based on design, and explicit configuring of the user, rather than studies of practical use. As with all the CI projects, or large scale infrastructure more generally, practical uses will only be possible many years into these ambitious endeavors.

many iterations of discussion and editing since this particular vignette, however, it is clear that this particular 'small' moment largely informed design thereafter. Today GEON's website is more easily understood to mirror the categories which emerged in this discussion rather than those of the draft presented at its beginning. Much more than simply a discussion of the portal, or the 'front end' for GEON, this discussion articulated and then enacted the GEON project itself. Science research, multidisciplinarity, the goals and tools that make up the core of infrastructural goals were shaped as a negotiated interchange amongst domain participants and between domain and IT.

On August 21st, 2006, the homepage's primary text describes GEON as "The Geosciences Network: Building Cyberinfrastructure for the Geosciences"; it includes a mission statement for GEON and a map of the partner institutions which link directly to each PIs webpage. A direct link leads to the 'science research' test-beds, and to the description of 'resources' in the form of data and tools. And still no images or references beyond 'geosciences' associates GEON's homepage to any particular discipline or PI researcher.

Within GEON it has been suggested that later years (perhaps in a 'GEON2') could lead to more formalized approaches such as requirements testing, social networks analysis or surveys of the community. However, for the moment, and in its formative early years, GEON was built upon the representational work of surrogate users: the geoscience PIs and their retinues. As Lee notes in her analysis of the BIRN CI "… the first users of a CI are the PIs themselves" – the intent is to use the tools developed in the CI project to assist the domain researchers. In this manner basic geoscience research will emerge as part of the project. Thus in the short term the PIs are not surrogate users, standing in for a community of geoscientists, but rather they are future users developing tools for their own research. However CI is infrastructure, GEON is general platform for enabling geoscience research and collaboration. As we have seen both IT and Geo participants both have this understanding of infrastructure in mind while designing for GEON and its portal. How is a bridge between immediate scientific research and an infrastructure for the broader earth sciences planned? In planning the GEON project the solution to mediating immediate research and long term infrastructure is dubbed 'the two-tier approach'.

4.3 – In the Interests of Disciplinary Difference: The Two-Tier Approach

A community base of users is in the future of GEON; in its present it is an emerging infrastructure project. One method in imagining a future has been to take the earth science PIs as surrogate users as informing the design of future uses. However, this is only segment of a larger overarching method which in GEON is called the 'two-tier approach'. This is a strategy prefigured in the Atkins Report which sketches a plan of IT /domain relations, but it has received a more methodical attention and articulation within GEON.

Put briefly, in actor's categories the first tier refers to short term results while the second to long term development, extensibility and sustainability. However, the two tiers can also be parsed along boundaries more familiar to readers of this dissertation: in developing CI the first tier enables domain basic science and the second tier IT research and application. Scientists provide the 'context' or the 'driver' for the development of technologies by seeking to answer contemporary research questions. Informed by these science drivers computer scientists 'implement' an architecture using the most up to date technologies that will support growth and sustainability. The two-tier approach is a strategy for overcoming problems well articulated within IT R&D circles by characterizing and then bridging the interests of participants. The result of successfully combining the two tiers is understood to result in infrastructure for the domain.

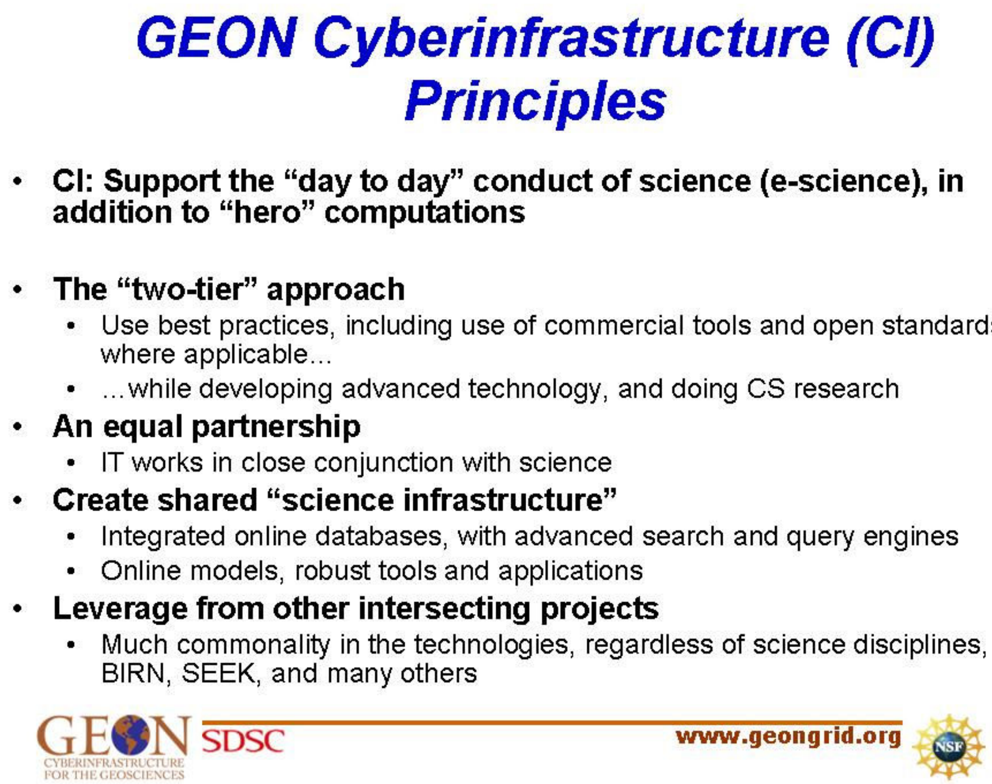

Figure 13: A Slide shown at the GEON AHM descirbing the 'two-tier' approach.

This slide appeared as part of IT PI Chaitan Baru's slide show during the third AHM (2005) entitled "GEON Systems Architecture". A webcast of the presentation and slides are currently archived at http://www.geongrid.org/AM05/presentations.php (Accessed August 8th, 2006). Although it is primarily directed at the IT audience (with a generalized reference to 'science' rather than the more specific earth science), the slide describes the core features of the two-tier approach: the conduct of basic research in the development of accessible technologies, in close partnership with the domain science, and producing infrastructure.

In this section I will trace actor's enactment of the two-tier approach in the practice of GEON. To do so is an act in boundary work and relations. Disciplinary difference, IT and domain, are characterized and then bridged. Individual disciplinarily informed interests are shown by actors to be diverging across the IT / domain boundary. In turn this is problematized relative to the goal of infrastructure building and bridged via the short and long term developmental approach specified in the two tier approach. A *commitment* to the daily practice of infrastructuring is generated for all participants through a 'side bet' investment across the IT / domain boundary (Becker 1960).

Twice Committed to Basic Research and Infrastructure to Boot

The two-tier approach is performed by characterizing a distinct set of interests for geoscientists, another set for computer scientists, and enacting a model of IT / domain relations. It is understood that if IT /domain relations are such that they account for both sets of interests they will lead to participant's investment in long term infrastructure for the geosciences. The goal of this approach is to produce two groups of *committed* "stakeholders" in the collective development of community infrastructure. The two-tier approach is designed to bridge the existing commitments of domain and computer scientists to conduct novel research. Howard Becker has described how two distinct sets of commitments come share a single set of interests in a process he calls making 'side-bets.'

The notion of commitment can assist us in understanding how practical action, individual motivations and collective issues in the development of infrastructure are conceived and managed by GEON actors. Commitment is defined relationally, and as a product. In other words, the two-tier approach seeks to *produce* a new commitment to infrastructuring through the *linking* of two of pre-existing commitments to basic research.

Collecting common sensical uses of the term 'commitment' in sociology, Howard Becker distilled 'being committed' as a matter being in a long term engagement, which is diverse in activity and is a chosen course amongst alternatives. It is long term in the sense that "it persists over some period of time," (Becker 1960: 34) such as a career or a political affiliation. Commitment is not to a single activity or goal, but rather the range of activities as defined by actors "the diverse activities have in common the fact that they are seen by the actor as activities which, whatever their external diversity, serve him [sic] in pursuit of the same goal" (Becker 1960: 34). Finally commitments are to particular trajectories of action out of a span of possibilities. If no other feasible options ('choices') can be conceived then it is not a commitment, there must be "several alternative courses open to him, each having something to commend it, but chooses one which best serves his purposes" (Becker 1960: 34). In the two tier approach commitment is understood to be (i) to a development of infrastructure, which is: a long-term engagement (Star and Ruhleder. 1994), 5 years for the ITR, and a renewal of GEON has loomed on the horizon from its inception;(ii) beyond a single use and single activities (Star and Ruhleder. 1994)– the range of activities in GEON greatly outstrip even the most broad definitions of science, and; (iii) chosen amongst alternatives – as prominent researchers in their respective fields each GEON PI has many more traditional avenues available. These are the produced commitments and they align with the explicit goals of building GEON and CI.

A commitment to infrastructure is sought. In order to produce this, the two-tier approach also comes to characterize two sets of pre-existing and divergent commitments along the IT / domain boundary. In the case of GEON these are basic research in computer and geoscience, each within their respective disciplinarily specific academic reward systems. For instance, finding out 'something new about the Rockies' and publishing this in a refereed geoscience journal, or, developing an ontology and 'demoing' this at a supercomputing conference. This model of disciplinary dynamics is well articulated in the literature as 'career rewards' or 'reward structure' (Gaston 1970; Fujimura 1988; Klein 1990; Fairweather 1993; Sieber and Trumbo 1995; Van House, Butler et al. 1998)[26]. In the academic literature this has involved identifying the ideal typical career trajectories (e.g. academic tenure, salary) and the common reward structures (e.g. publication, teaching) needed to achieve these. The concept has migrated out of sociological and economic literature and into common parlance within CI, but

[26] The research in this vein is varies substantially in method, topic and conclusion. Early sociological studies focused on identifying the rewards required for success in terms of wage, prestige, job-security (tenure) or research funding (Blau 1973), while others focused on the relationships between reward systems and productivity (Reskin 1977). Economic studies have included analyses of market inefficiencies such mismatches between rewards and skill (Acemoglu 1995) or rethinking reward systems as social networks (Montgomery 1991). There is a also an emerging subset of researchers specifically interested in identifying new or emerging rewards structures in, for example, multidisciplinary, collaborative or project based research (Allen and Katz 1995; Roberts, Wermus et al. 2003). In CI circles it has come to be understood that there is a mismatch between career rewards and infrastructure development activity (Kling and Spector 2003; Foster and Gibbons 2005; Kim 2006). Such studies are presented at the scholarly conferences in the orbits of CI, becoming familiar frameworks for participants. In GEON I have never encountered a reference to an academic study on this topic, instead the use of the reward structure concept has become colloquial or common sensical.

maintains this basic analytic structure. In the two-tier approach actors characterize career trajectories and rewards along the boundary domain / IT and in turn the system of career rewards is problematized relative to a commitment for infrastructure development.

Domain scientists are seen to be disinterested in the production and maintenance of community resources. Instead they focus on the development of IT tools which will serve their particular needs in answering a scientific research question[27]. What is traditionally rewarded within a scientific community are 'science results' – broadly understood as new domain knowledge – and not the provision of, for example, database interoperability, visualization tools or computing resources. In short, individual scientists are not usually rewarded within their own community for the production of long-term information infrastructures.

This is a laminated characterization in GEON. Geoscientists acknowledge but often lament this vision of scientific practice. Similarly it is a criticism put forth from IT practitioners. As one IT PI notes in describing the geo PIs:

> In this group I can see some of our guys who just have this interest "Get it ready so I can use it and get something done!" and "When is it going to be ready?" And actually sometimes I think of it as more like a 'typical science' […] The typical science of the 20th century. They just want to publish their next paper and just move on.

[27] This is contrary to Merton's scientific norms of communalism (Merton 1942). Merton described communalism as one of the four cultural prescriptions of science, and characterized it as a community of sharing. Communalism referred in particular to scientific findings, but also technique or method. Later sociological work has characterized these 'norms' as interpretive resources rather than rules: they are the means by which scientists account for their activities rather than what motivate them (Mulkay 1976; Mulkay 1979; Mulkay 1980; Gilbert and Mulkay 1984). Framed in this manner communalism or co-operative work in science becomes what must be explained rather than what is used to explain.

He goes on to describe a second type of geo PI that is engaged in the vision of infrastructure and the development of community tools. But he notes that this enthusiasm is not enough, something else is required to have them "engage deeply, participate daily." An enthusiasm or ethos for infrastructure on the part of scientists "doesn't keep them with us seven days a week." In other words, an understanding of the value of infrastructure is not sufficient to fully motivate a domain scientist to degree necessary for 'daily' work. Another IT PI characterized the geosciences more generally in this fashion, noting that *participating geo PIs* were aware of a commitment to infrastructure, but noting that the *community* maintained a narrow focus on science and particular science contributions:

> Right now I think GEON is struggling with this […] I think the people involved on the project understand that what they are creating is infrastructure. Not doing science. When all the community *wants* is the science. That idea [of infrastructure] is kind of new. People don't get that. Because in the past something has been a scientific problem: "We need to solve this. I need a really huge computer, I'm really going to focus, laser-like on getting my big computer, run my job on the computer and I'm *done*. All this nonsense about creating this cyberinfrastructure so that everybody can use the big computer. I don't care about that." So the idea that you could actually be spending effort on creating a broad based infrastructure I think there is still a little bit new.

Infrastructure here is an idea, a vision or an ideal, but it is also a practice, a commitment and a long term endeavour. In the excerpt above it is an ideal that in the pursuit of science the 'geoscience community' does not share. In the excerpt above this it is commitment to daily work, to the details of technical construction that an individual geo PI does not share. For this, the domain scientist and the community must achieve the sense of making a scientific contribution.

What does and does not count as a scientific contribution is often invoked by scientists themselves, sometimes with a general consensus, but often as a debate over the boundaries of the category itself: 'what counts as new knowledge in geology.' Often these discussion are enacted as a geo PI's concern for their graduate students. GEON's geo PIs themselves are relatively established, tenured, professors with strong track records in the field. Meanwhile the career trajectories of graduate students participating in GEON may appear nebulous as they are engaged with tasks not traditionally rewarded within the geosciences. IT developments are not considered advances – "databases are not science," or "we can't publish all this metadata we're making," (Geo PIs). Even the efforts in the hybrid terrain of Geoinformatics may be contested as geoscience, or may require extensive articulation as knowledge about the earth. A GEO PI describes his coaching to a graduate student working on GEON, presenting at the Geological Society of America "I made him drop all the tech language, even ignore most of what he'd worked on, and instead focus on the geomorphology. If they don't do that they will lose the audience" (GEO PI). With both geoscientists and information technologists the criteria of the domain are understood to reside in community opinion. An established scientist may be able to skirt these, pushing the boundaries of a 'scientific contribution,' but the more vulnerable graduate student or junior faculty member is vulnerable to be excluded as a trained geoscientist.

From an historical perspective it is clear that what counts as a science result is constantly shifting within any given science. For example Galison (Galison 1997) describes how what counts as experimental evidence has shifted in physics as different instruments have been introduced, particularly the computer and digital modeling;

Winkler has shown the rise and fall of images as knowledge within early-modern astronomical communities(Winkler and Van Helden 1992); Epstein has shown the reconfiguration of standards for clinical trials in the face of activism and ethical issues (Epstein 1996). It has not been the norm for participants to generate historically based analyses of change in criteria for scientific contribution; however, they do regularly speak of changing career structure and reward.

Shifting criteria for science results and even an active engagement with changing these criteria is also part of GEON efforts. The various manifestations of Geoinformatics (as a funding line at the NSF, as a division at the Geological Society of America, or as a special issue publication – see Ch. 3.2 and 3.3) have been sites for discussing the re-evaluation of career rewards, or designing new venues for interdisciplinary publication. These can include the publication of data or metadata, credits for the design of tools, or the generation of interdisciplinary units across domain and computer science ("we need a Geoinformatics just like they have a bioinformatics" Geoscientist). However these are considered long-term solutions and will not address the immediate needs of, for example, graduate students working on GEON today.

--

A complementary claim is set forth about information technologists and, particularly, academic computer scientists. They are seen to be disinterested in the practical results of their research or design work. Sitting on one side of the 'brick wall' computer scientists will design programs intended for domain use without much consideration for specific application needs, functionality, or accessibility. Tossing their final product over the brick wall, these computer scientists are able to advance in their own field by publishing their code-innovations in journals, pointing to grants awarded and the product programs themselves which remain in the demo stage. Code, grants and demos are surrogates for successful development regardless of actual outcome. One IT PI expressed this as a concern by a GEO PI that the IT team may fall into a traditional *waterfall model* (see also Ch.4.2):

> in software engineering there used to be, the old word was what they called the 'waterfall model'. […] Our guys here [at SDSC] say "Ah, I built this thing, now you should use it." When they mean "I built this thing. It may or may not work. Why don't you use it and tell me if it is broke and I'll fix it." It is very hard for people to work with this software.

Within CS, the claim goes, little attention is paid to the lifecycle in the domain of the application – has the program received uptake? does it meet the requirements of the client? – and even less consideration is given to providing technical assistance or long-term support for operability. CS is characterized as research: its reward structure is tailored to the development of novel functionalities in theory, publication or programming in the form of proof of concept[28].

Meanwhile software engineering is a technical practice, roughly following one of various design methods, and has few outputs in research or publication venues. The goal of software engineering is the production of stable, robust, 'production quality' applications which will receive some level of support for tailoring, bugs or usability. Software engineering, however, is understood to rely upon established technologies and

[28] As Lawrence notes, 'proof of concept' are "sometimes unstable or highly customized software which often required the people who created it to run it," (Lawrence 2006 (forthcoming)). The purpose here is to show a viable functionality *could* be produced, but such a demonstration does not typically require any robustness, considerations for interface, extensibility, support, documentation and so on. In moving from proof of concept to production quality software often the demo program is completely thrown out in favour of a fresh start from software engineering principles.

methods rather than the development of new functionalities. The application of new software functionalities, developed as theory or proof of concept in CS, is seen as the domain of the SDSC – a hybrid of CS and software engineering.

There is a nebulous terrain between 'pure' computer science and 'mere technicians' and at the SDSC a general fear that this is in fact a slippery slope in either direction. This is reflected in the Atkins Report as it describes the difficult balancing of domain and CS interests in building CI:

> If the organization is weighted too heavily toward the domain scientists, the focus overemphasizes procurement of existing technologies, and computer scientists become viewed as "merely" consultants and implementers. If the weight shifts too heavily toward computer science, the needs of end users may not be sufficiently addressed, or effort shifts too heavily toward creating new technologies with insufficient attention to stability and user support.(Atkins 2003: 50)

That computer scientists may become 'mere' technicians is an active site of concern for the IT participants. They have characterized geoscientists as "being used to having techies around and having them do their tech-work" (SDSC information technologist). This is the primary experience of scientists in working with information technologists: as information managers, software engineers and technical support staff they are seen as auxiliaries to scientific research. A scientist may delegate particular tasks – such as updating a website – or request functionalities – such as setting up a server – but these are not science or research[29]. In the boundary work of delegating tasks and roles technicians

[29] An IT PI described this concern relative to the geosciences, characterizing earth scientists as long established scientific field while CS remained nascent and relatively uninstitutionalized:

work for scientists, pure computer scientists may have no contact whatsoever with scientific applications and the SDSC's information technologists struggle to balance both and *collaborate* with scientists.

For example, an ostensive hierarchical structure emerged precisely from geo PIs requesting basic services from the SDSC team. On many occasions in the early period of GEON's development geo-scientists would request particular 'low-end' services from the IT team such as building websites. This resulted in high-priced and high-end computer science researchers specializing in, for instance, semantic mapping across databases instead creating front ends for web service databases. The results were unsatisfying to all participants: computer scientists with no interest in writing web-pages producing mediocre outputs while geo-scientists became dissatisfied with the supercomputer center's ability to provide service.

The problem was of poorly aligned expectations as earth scientists treated SDSC participants as service oriented technicians. The IT resources of GEON became minor battlegrounds for interpretation. For some of the earth science GEON members the SDSC appeared as a source for bountiful high-tech resources; requests started coming in for assistance on their various projects: web-pages, user-interfaces and dataset clean-up.

> Maybe in the beginning we were a little bit more defensive because now you're trying to get *involved* because you're afraid that, "If I get involved in that [geoscience], they'll eat me up. Geology has been around for hundreds of years". But no, computer science is now a solid thing so no, you shouldn't be afraid of being … disappearing or losing your identity or something.

Many of these requests by geologists were communicated directly to the IT team of GEON, end-running around the IT PIs.

As the larger goals and organizational vision of GEON have solidified these minor projects have come to be defined as outside the bounds of GEON's work. And yet how does one pull-out IT resources from PI's work without causing strife? This resulted was mixed feelings about the role of the IT team within GEON, perceptions of efficacy at the SDSC and the re-opening of further debates about GEON's mandate.

Once again a re-opening of discussion around GEON's mandate for simultaneous IT and geoscience research became necessary. In this case more than discussion resulted, and novel organizational structure and routines were developed to manage the distribution of IT resources and to establish a system of accountability. In this case organizational 'protective boundary' was been built around the IT resources. Resources would be tied to the science research goals in the test-beds, thus enacting the two-tier approach.

As specific science goals have solidified around the two original test beds in the second and third years of GEON these have become a focal points for resource distribution. Decisions about what does or does not count as a relevant test-bed development of IT resources has been delegated to two earth-science PIs. Thus, a routine emerged in which two geoscience PIs served as gatekeepers for access to IT resources. Requests for development efforts (whether for computing resources, ontology workshops or visualization tool development) on the part of geoscience participants would go through one of the two test-bed representatives. These PIs would in turn make executive decisions about the importance of the request to timeline, importance to project, necessary work-force and so on, and then pass on the requests to the IT team. Visions of science-driven information technology development came to be grounded in practical human organizational distributions, in this case a loose hierarchical flow. This is an enactment of what GEON actors have come to call ‘two-tier approach’.

--

The sum of these two trends – information technologists indifference to domain, and individualist tendencies of domain scientists – amounts to what has become for GEON the crucial general problem with IT/domain collaborations. Each act of characterization is also an act of problematization. The interests of computer and domain science are articulated in order to demonstrate how these diverge from a practice of infrastructuring. Problematization in the two-tier approach is neatly captured as this IT PI summarizes the typical dilemmas of both computer and geo science as they attempts collaborate.

> In other words a computer science researcher will say that's really not my thing, “go hire a programmer”, which happens all the time, and its very bad. I think it s very bad. […] when you get a bunch of domain guys just doing software […] you get some really bad products. Sometimes you get good things but sometimes you get bad stuff, because they don't know the state of the art. And they don't know what mistakes they're making. Otherwise, we build tools without knowing how usable or how useful they are, which happens all the time with computer scientists. Even now we have that problem. We have people like [Janus] who is a professor in computer science involved with GEON. His interest, he loves the kind of stuff we encounter when we talk to geo-scientists, but his interest is in creating this more perfect, rounded ball […] But the scientist, he doesn't get want to have to wait 3 years to get that rounded ball!

Three laminated problems are outlined in this excerpt. First, a computer scientist's unwillingness to work with domain scientists in the development of usable tools. Second, domain scientist's short-sightedness when software tools are developed without close collaboration with information technologists (i.e. when the domain parcels out tasks to a technician, see above). The effect here, as described by the PI, can be software products made with few considerations towards infrastructure such as interoperability, extensibility, and so on. Third, the desire of the computer scientist to create a 'rounded ball' product, by which this speaker means a theoretically tidy and technically elegant solution, but which is may be three years in the making and perhaps without practical functionalities for the user. In the excerpt the speaker switches freely back and forth across the domain / IT boundary, problematizing imbalances in either direction: software development guided too strongly by either domain or computer science actors.

Basic research and traditional career trajectories are counter posed with community infrastructure. A lead IT PI frames this in terms of the larger CI endeavour and contrasts this with traditional research ventures:

> How much are we building for community at large and how much are we building so that people can blossom into their own independent thing? We have people completely at the extremes. One view is that everybody does their own independent thing, then we haven't changed anything, right? What have you done? Have you changed anything? No. Because we just got a bunch of money and everybody went off and did their own thing. That's the extreme view. But, where is the balance? The other extreme view is everything is community. And already people are saying 'well of course *everything* cannot be centralized, *everything* cannot be dictated.' So it has to be loosely coupled, and how do you that? (IT PI)

In this excerpt the speaker is implicitly speaking not of GEON in particular, but of the larger vision of CI[30]. The Atkins report specifically characterizes the task of creating an institutional environment in a manner which mirrors the problematization in the two-tier approach. As a review of the PACI (supercomputer centers) program, the goal was to articulate the difficulties in the centers thus far and outline recommendations for future changes. Without using the terms 'two-tier approach' the Report prefigures much of the thought within GEON:

> Turning now to issues of concern, the PACI program has exhibited, from its beginning, a tension between two needs that cannot easily be reconciled: providing production systems for the current generation of high-end users, and moving to the next highest level of computing capability. Since the program's core funding has never been adequate to support more than one generation of computer system, tradeoffs have been inevitable. (Atkins 2003:63)

As with the two-tier method here the domain (high-end users) stands for short term basic research results while CS stands for a higher level of computing research and accomplishment. And as with the two-tier method we have seen that Report seeks to dovetail these two problems in the creation pf domain-centered information infrastructures that are IT based.

The Report recommends shifting from a model of primarily core-funded supercomputer centers and that instead support would be project driven by individual

---

[30] This PI later goes on to argue that "we are at a cusp" at which "if these cyberinfrastructure projects fail then we go back to what we've always been doing: making tools to answer a science question." As I have noted, almost all participants have a sense that there is something new about GEON. Here this novelty is described as part of a larger CI movement, a different vision of technological development around the notion of community infrastructure and collaboration, rather than this actor's historical characterization of a past in which technologists developed single-use tools to answer a 'science question'.

domain based grants. The centers since 1985 have been funded by 'core grants' from the NSF. Still coupled with individual project and research funding until recently the core grants have been understood to give each center a certain amount of discretion in developmental trajectory. The Atkins report suggests that this core funding would be substantially replaced purely by individual project grants. These grants, or CI projects, would be headed by practicing scientists while still relying on the centers for the application of high-end computing technologies. The "tension" of providing production systems for users and moving to the next level of computing would be resolved in the practice of science driven implementation.

GEON is just such a project as the Report describes. It is a domain centered general infrastructure project, organizationally housed at the SDSC, and seeking to produce high end computing tools for science applications. It is in the practice of constructing the emerging infrastructure that the dual problems of practical usability and novel IT application is to be resolved.

--

The two-tier approach is a *methodology* of IT / domain collaboration in the enactment of a long-term community infrastructure. It contains a theory of the actors and their interests; of the practical methods for technology development, and; a strategy for ensuring technology adoption:

> The science driver is providing the context in which to do your work. There has to be a context. The other thing is looking at the science problem you're solving. *The two-tier model is simply how you go about doing that. (emphasis in original) (IT PI)*

The methods in the two-tier approach are very similar to those of the designer in Woolgar's configuring the user. Woolgar notes that in building technologies the designer will model the future user, and then build these uses into the system. Here the designers – the GEON participants – come to model themselves, and they do so along the domain / IT boundary: what are our interests? how can we meet both? Participants model themselves and are in turn modelled. Disciplinary difference, at the level of career rewards and research interests, are first be characterized in order to be bridged.

For example, while infrastructure development is a long-term endeavour science research is an immediate concern for the GEON PIs; the two tier approach bridges an extended developmental trajectory with short-term research interests:

> But the thing with GEON or BIRN trying to work with the scientists. The scientists think, they're gonna get something that they can use right away, to produce science. And the IT guys are thinking, well they are using the scientists to get some interesting context and we'll now build this nice little tool. And "I want a nice tool, if the tool isn't cool then what's the point." And so the scientist is now[..] not getting the job done. They both get frustrated. And the IT guys are saying, "I don't want to help you put your data into a data base. I want to write something new, a mediation tool." But both are needed. In the sense if this project is to succeed you have to keep the scientists engaged. (IT PI)

Discussions regularly return to the individual motivations for engaging in infrastructure projects or community standards for contribution. These are often considered the leading cause for the failure of infrastructure projects more generally.

Solutions such as transformations in disciplinary culture – such as taking domain specific tool design into consideration of contribution – or institutionalizing new reward structures – such as enrolling journals in the publication of data – are regularly discussed. However, these are usually seen to be long term solutions.

The primary, and immediate, solution to the diverging interests is the two-tier approach:

> The two-tier I'm talking about is very simple. It's like the short-term approach to getting it done and the longer-term approach to getting it done. So if you think in terms of how do you keep the IT guys happy, I'll argue that [two SDSC members] are the hot-shot IT guys who also want to get a paper and want some cool technology out of it. If I tell those guys that 'all you're going to do is just unpack a database and put some data into it and make it so the scientists can build a nice website' that's not interesting for them. Even though they're not trying to be pure computer scientists; they're trying to be in the middle**.** You have to give them something that looks sexy from a technical point of view. But if scientists say build this database, give them a website, something they can access and draw maps, they [domain] get something out of it. And then we [IT] know that we have these more fancy problems and mediation and data integration problems. But you can't do those [~'fancy problems'] if you don't do this [~'build a database' &c.] because you won't have any scientific users. (IT PI – emphasis added)

The key to the two-tier approach is making stakeholders in the development of community infrastructure building while also meeting existing commitments. The computer scientist must be configured to 'care' about implementation success as measured from *within* the domain. The information technologies must be designed with real uses in mind and brought to the point of development such that they are usable. The individual scientist must be configured to 'care' about securing resources for her broader scientific community. Looking beyond single uses in a particular research question the technologies must be useful to a broader community user base.

Becker describes the generation of new commitments through the intertwining of emerging and established interests, or the making of 'side bets': "The committed person has acted in such a way as to involve other interests of his, originally extraneous to the action he is engaged in, directly in that action," (Becker 1960: 35). He further defines the emergence of new commitments as requiring i) that activities be tied to consequences relevant to *other* interests and lines of action; ii) that the actor must place themselves in that position by previous actions, and; iii) that the committed person must be aware of the multiply linked consequences.

Each commitment is relative to one side of the boundary domain/IT, and each side-bet crosses that border. A geoscientist in attempting to discover something new about the Rockies assists in the development of an ontology for integrating geochemical databases. An immediate geoscientific research question is coupled with the application of information technologies. Contrariwise, an information technologist draws on geoscience knowledge to provide a context for application of developments in knowledge representation. The domain becomes a 'science driver':

> In the case of GEON, say, pick the Rocky Mountains as the variable. That's the driver. […] But how do we achieve that? The driver just tells us "this is the problem we are trying to solve. We're trying to understand the tectonics of the Rocky Mountains." That tells me I need these kinds of datasets of the Rockies, I need datasets of that, I need these kinds of models. But then how do I build it? And I think the two-tier is going more towards how you build it. Okay, we now we know the driver and it will take 5 years to build it.

> During which time you will lose all the scientists. So I say no, we'll do something quick. Put together whatever datasets are available today, they might be dirty, people might not like them, but that's what we've got. In 6 months we have something while you're building the more fancy thing in two years. That's why I'm calling it the two-tier – the fast tier and the slow tier. (IT PI)

In this approach two sets of basic research results are produced, this is tier one, and the resulting ontology is infrastructure, tier two, capable of integrating datasets for other domain researchers. Each participant makes a 'side bet' that in order to meet established interests for producing basic research results both parties must invest in an emerging interest: the production of infrastructure.

Does this fulfil Becker's definition of 'being committed'? Did the actors bring themselves into this condition by their previous actions and are they aware of alternate possibilities? There is a general sense in GEON that this project is different, that it is not everyday science but rather a future possibility for changing everyday science. In 'signing on board' with GEON PIs are committing, in the conventional sense of the term, to community infrastructure building and a vision of CI. With alternative options available, such as seeking funding sources for basic research rather than infrastructure, actors are choosing a pathway. However this is not enough, for Becker commitment requires that established interests be at stake. Commitment is achieved only if a side bet comes to affect other aspects of an actor's lifeworld. A similar sentiment was expressed by an IT PI when he noted that an ethos for infrastructure was insufficient to keep participants working on GEON "seven days a week" (see above). The interdependence of interests – basic research for CS and geology – is achieved through the investment of efforts across the domain / IT boundary. For this we must trace technical developments at a finer scale of granularity than is usually expressed when actors speak of the two-tier approach, this is the topic of the next chapter.

The two-tier approach is a general organizing principle. It speaks to the problems of collaboration across disciplinary boundaries and provides a model for addressing these issues. Enacting this model involves i) the articulation of disciplinary difference, often in the form of individual motivations or community reward structures; ii) the problematization these motivations relative to a goal of infrastructure development, and; iii) providing a general framework for fostering commitment to infrastructure development by meeting extant interests.

We have seen characterizations of computer science as committed to basic research in the form of theory or proof of concept and disinterested in the development of production quality software. We have also seen the characterization of the domain, specifically earth science, as committed to basic research in the form of new knowledge about the earth, and disinterested in the production of general research resources for the community. In turn, we have seen the problematization of each of these characterizations relative to a goal of infrastructure building: computer scientists failing to produce usable research tools for the domain, and earth scientists unwilling to invest time in producing general tools beyond a single scientific problem. Finally, we have seen how the two tier approach offers a general method for meeting both interests through investment (or 'side bets') across the boundary IT / domain. For computer scientists the domain becomes a 'science driver' offering details of actual practice, knowledge and which in turn inform the development of novel software applications. For domain scientists computer science becomes the means by which to address particular research questions.

In the two-tier approach the result of dovetailing both interests is the production of a new commitment to the development of infrastructure. In order to more fully understand this general approach to technology development we must more closely follow an instance of its enactment in practice. This is the purpose of the next chapter which traces the development of a routine for ontology development in GEON.

Conclusion

In building an emergent infrastructure GEON participants have simultaneously invented new organizational forms as they design and enact future technologies. However GEON actors have not done so in isolation. The names of the meeting and structures of GEON are familiar to many participants in CI projects, particularly those funded under ITR: I have attended an 'All-Hands' meetings for BIRN, SEEK and LEAD. LEAD's 'PI meetings' are different from GEON's in that they are held over phone or video conference, but they are similarly structured in their intent to communicate updates of current research and make plans for future work. Structurally all these projects have 'test-beds'. What composes the test-bed varies significantly by project, but in CI circles the term has become a common short-hand for the empirical sites of application for novel information technologies.

GEON is at play in a larger field of emergent infrastructures. This too is part of the logic of interoperability: a drive to *problematize* the tension between the general and the particular. Usage of the term varies in practice, however published writings refer to *a* cyberinfrastructure, rather than 'cyberinfrastructure*s'*. Thus GEON is 'cyberinfrastructure for the geosciences' and other CI projects such as BIRN, SEEK, and LEAD are all branches of this singular cyberinfrastructure. In the visions which align with the Atkins Report what provides the *explicit* connective string is the movement of scientific and engineering data across the disparate resources and environments There is a single cyberinfrastructure *because* data encounters no institutional, technical or disciplinary boundaries e.g. in data's movement from the remote sensing devices of ecosciences' CLEANER, through geosciences' GEON suite of visualization tools, to its output via SEEK's web portal, travel is *unproblematic*. These are the goals of CI in the logic of interoperability. Written sources such as the Report or the GEON proposal will explicitly focus on the technical means to do so, but they also contain organizational solutions. In the practice of constructing infrastructure, the organizational solutions gain a greater attention in the everyday talk of participants.

--

Organizational difficulties are problematized in very similar ways across CI projects. In this chapter I have focused on how GEON participants have come to problematize 'knowing the future users' and 'generating commitment to infrastructure.' This problem is not native to GEON, but is problematized in a very similar manner within parallel CI projects.

First, in an emerging infrastructure with no 'functionalities' there are no users, and yet there is a mandate to serve a future constituency. In GEON this is the geoscience community. In order to come to know 'the community's needs and requirements' the geo PIs have served as surrogate users, representing and standing in for the interests, science questions, and disciplinary languages of the community. Similarly in her studies of BIRN Charlotte Lee has noted that “domain experts who will be the first users of the cyberinfrastructures are also intimately involved in development. These expert domain scientists are simultaneously managers and end users,” (Lee 2006:2). As with GEON, BIRN has its technical core (Scott 1992) located at the SDSC. Many GEON IT participants are also members of BIRN. There is an exchange of individuals, expertise, and technologies across these projects. Both have also come to frame future development in the same terms: how to know its future users in order to design for that community. This problematic is outlined most clearly in the Atkins Report: of developing functional tools (rather than CS theory) which will serve a broader constituency (rather than a ‘one-off’ solution for a science question).

Second, the reward structures, and the very mandate of GEON, demand the production of basic research for both earth and computer sciences. And yet in the face of these two short term demands for novel findings, how does one also produce long-term infrastructure for the community? We have seen how GEON has problematized a requirement to produce ‘twice basic research and infrastructure to boot’. This framing of problem is typical across ITR programs. In her studies of LEAD, funded one year after GEON by a large-ITR, Katherine Lawrence has noted :“The most apparent tension in the project was the necessity of balancing research versus development priorities,” (Lawrence 2006 (forthcoming)). As with GEON and the Atkins Report, she frames this problem as a ‘balance’ and describes some of the strategies for resolving this tensions “the most typical way of balancing the two priorities appeared to be episodic burst of development, in anticipation of impending demonstrations” (Lawrence 2006 (forthcoming)). In GEON such hectic activity in developing ‘demos’ was also common previous to, say, a Geological Society of America or Supercomputing conference. However the explicit plan for twice basic research and long-term infrastructure development has been the ‘two tier approach’: the distinction of interests around the IT/domain border, and posing their common resolution through the development of technical mediation applications which are basic research for earth and computer science, and also infrastructure for a larger community. The problematic of interoperability is framed generally across cyberinfrastructure projects, it is the local solutions which are particular.

--

In the logic of interoperability disciplinary difference is articulated and its consequence is then problematized. Disciplinary difference, while necessary for domain research, leads to difficulties in communication, collaboration and sharing data. This is not a problematization of heterogeneity across the domains, per se, but rather it is the boundaries themselves that are problematized. Thus the solution is not to transform the disciplines, but rather to mediate them. The problematic of interoperability is coupled with its resolution.

The logic of interoperability is common across the CI vision. The Atkins Report draws examples from physical, earth and biological sciences. The language of the report is framed in generalities: the natural, engineering, and social sciences are ‘the domains’. However the ‘conduct of science’ is particular. It is computer science and information technology that are to the domain and their practitioners travel freely across disciplinary boundaries. The Report is able to address a vision of development more generally, but as with GEON, each project must be ‘tailored’ to the domain.

In the ‘two-tier approach’ the diverging of interests of computer and earth scientists need not be changed. Both will continue to conduct basic research in their respective fields. Rather, it is the organization of relations between these groups that must come to structure novel commitments. In the two-tier approach both ‘something new about the Rockies’ and ‘a tool that is cool’ are sought. At the intersection of this work to produce twice basic research disciplinary difference is mediated, not transformed, by an interoperable layer of information infrastructure.

In the next chapter I trace the production of mediation technologies in GEON. These are the general technologies of the particular. I focus on ontologies, a ‘knowledge representation’ tool that emerges at the intersection of conducting basic research in computer and domain science. These tools are envisioned to enable a bridging of the heterogeneous languages and concepts of the sciences. In the logic of interoperability, the languages of the divided earth sciences remain distinct, technical, particular to their conduct, but mediated and translated by a general technical substrate enabling ‘semantic interoperability’ at the level of knowledge itself.

– Chapter V –

Knowledge Representation and The General Technologies of the Specific

–

The trouble concerns the fact that the "truths" of the modern scientific world view, though they can be demonstrated in mathematical formulas and proved technologically, will no longer lend themselves to normal expression in speech and thought.

-- Hannah Arendt

The method proposed here must have the peculiarity that it is applicable to all; it must, so to speak, specialize in generality.

-- Ross Ashby

Dropping science like Galileo dropped the orange.

--Michael Diamond

–

In the previous three empirical chapters I have focussed on actor's work in drawing the boundaries between computer science and domain, as well as within the geoscience domains. We have seen how disciplinary difference is enacted and problematized. Earth scientist render themselves as disciplinarily divided, they have varying methods, language and criteria for scientific knowledge. Collaborative and multidisciplinary research is hindered by these differences, data will come to be lost, and the field may fall into stagnation. Similarly, computer science and the geosciences are disciplinary divided by varying research interests, language and reward structures. The production of infrastructure is hindered by these differences in goals, with computer scientists producing theoretical tools useless to the domain and the domain unwilling to invest in long-term development. But in interoperating the geosciences, characterizing disciplinary difference through boundary work is accompanied by a plan to establish boundary relations. Within this model a resolution to disciplinary difference is proposed alongside its problematization. This chapter will focus on boundary relations and mediation.

Disciplinary difference is characterized at both boundaries: working across computer and earth sciences *and* within the geosciences is found to be difficult at the level of language, expertise and topical interest. However, in the vision of universal informatics the domain/ IT boundary has a special outcome: its crossing is an occasion for building persistent relations across the divided domains.

Figure 14: Relations of difference in the logic of interoperability.

Here, the heterogeneous data of the sciences can be made tractable through mediation by information technology. Conceptual, linguistic and even epistemic differences in the earth sciences can be bridged by working across IT/domain. Crossing this boundary has the particular quality of resulting in a solution which persists: infrastructure. This is the boundary work and mediation within the logic of interoperability. We will see how this vision is enacted in the technical work of building *ontologies*: GEON's principal knowledge and data integration technology effort.

In the last chapter I described the 'two-tier approach.' This is GEON's strategy for generating commitment to infrastructure building. Boundaries are bridged by making 'stakeholders' of participants. The two tier approach involves characterizing the diverging 'basic research' interests of practitioners and then meeting both needs through a *side bet* investment in the development of infrastructure. Here I trace how this approach has been enacted in efforts to develop mediating technologies.

In its first three years GEON participants developed a general routine for representing specific knowledge in software technologies known as ontologies. As with the *strategy* of the two tier approach, in *enacting* a data integration technology basic research interests for computer and domain scientists had to be met in the production of community infrastructure. Cast as the 'science drivers' geologists invest in the development of an integration tool with particular research questions in mind. Meanwhile information technologists engage in novel computer science research for ontology and knowledge representation. As with the strategy of the two tier approach it is the practice of coloration across the domain / IT divide that will result in a persistent solution to the domain / domain divides. These are the general technologies of the particular.

--

Ontologies are a general technology for representing particular knowledge. They are software tools to facilitate the manipulation of material from heterogeneous databases and to enable operation across multiple disciplines or scientific institutions. The application of ontologies to science is relatively novel, and is as much a computer science research endeavour as a site for tool development in the geosciences. They are an ambitious and complex solution to the problem of diverse languages, methods of data collection and diverging epistemic criteria. In GEON they are a key piece in the production of interoperability across geoscience disciplines.

This approach is often referred to as *semantic* interoperability or integration because ontologies represent knowledge in the form of *meaning based relations*.

Meanings, knowledge or relations which are represented in ontologies are 'captured' or 'acquired' from domain sources such as human experts, published articles or even textbooks. In a stripped down technical understanding of *knowledge acquisition* building ontologies involves taking a domain knowledge, formalizing this knowledge into a machine computable format, and encoding it into machine language. But in the work of building ontologies GEON participants engaged in a much broader range of activities than any such technical definition.

From the first days of the GEON project it became clear that domain practitioners were not readily prepared for ontology building, that work of capturing knowledge is complex, and that representing a 'community of knowers,' such as the geosciences, required a substantial investment in securing consent from that community. This chapter traces over time the development of a general *routine* within GEON for capturing specific knowledges.

An ontology itself is considered infrastructure: when functioning it will permit the integration of datasets, meaning based searches or generating tailored interfaces. But GEON has the goal of building multiple ontologies. Over the first three years of the project a general routine emerged for how to go about building an ontology. Organizational routines are "repeated patterns of interdependent actions, performed by multiple actors," (Feldman 2000; Feldman and Pentland 2003). A routine must always be enacted. It has no existence outside action. However, as we shall see, over time GEON actors came to articulate their work as a routine, transforming it into an actor's category: the ontology workshops. As an actor's notion it serves as a resource, a 'template', for how to go about a specific task. Rather than having to begin anew for each ontology development project a set of 'best practices', software tools, slideshows, and even specific arrangement of a room came to support this activity. The routine rendered the complex and uncertain activities of knowledge representation into a set of steps, reducing the work of reinvention on each new occasion of ontology building. A general routine was developed to match the general technology of ontologies. Each enactment of this general routine came to produce a specific knowledge representation.

--

An implicit theory of knowledge is wrapped into ontology development. To name only a few, ontology is premised on an understanding that knowledge can be captured and represented in formalisms, that a community is the broker for that knowledge, and that a general informational language can serve to represent a broad spectrum of knowledges. In this formulation knowledge is rendered objective, visible and manipulable. Knowledge representation experts become 'meta' relative to knowledge itself, capturing in one domain and then another, and with a relative indifference to content.

All this said, I do not wish to critique knowledge representation endeavours, per se. First, because as we will see knowledge representation experts are not epistemologically naïve. Philosophy, sociology, anthropology and even science & technology studies (STS) are endogenous actors in the fields of knowledge representation. What is knowledge, how it should be captured and what might be the consequences of such a capture are endogenous concerns for GEON's IT PIs, and over time its geo participants as well. Rather my interest is in tracing the work of actors: how is knowledge represented, how do conceptualizations of knowledge come to inform practices of knowledge capture and how do these shape their downstream products: ontologies. Ontologies will be respecified as existing at the intersection of (technical) material arrangements, the practices of their capture and the work of enrolling a broader community of knowers.

This chapter is divided into two sections. The first will serve as an introduction to ontologies and the technologies of knowledge representation. A ‘pop’ example will serve to introduce the notion of ontologies at a very general level. Next, a ‘walk through’ user-scenario of knowledge integration in GEON funding proposal will serve to illustrate GEON’s vision of the interoperated geosciences.

The second section of this chapter will outline the emergence of a routine for knowledge integration within GEON. The development of the routine is traced from the kick-off meeting to outreach to a broader geoscience community. For heuristic purposes the emergence of the routine is described in three parts: *learning ontology, enacting ontology, and enacting community*. First is the steep learning curve for the domain. To bring experts on board is to inform them of the technology of ontology, its strengths in the face of other interoperability strategies, and the particular work it will require. Enrolling practitioners is securing an investment in technological direction by a domain community. Second is the practical work of knowledge acquisition or capture. Written sources such as textbooks and technical treatises are often not precise enough for transformation into description logics: there may be competing accounts of the same phenomena, overlapping taxonomies and standards, or outright contradictions (Bowker 2000). Similarly in consulting authorities in a domain, a programmer may find that these experts are not immediately able to state domain knowledge in the terms necessary for ontology building. In short, the ontology specialist often finds that what participants in a domain consider their validated and structured knowledge is not readily compatible with ontology building. Finally in GEON the scientific and technical activity of ontology building became coupled with a backgrounded work of identifying and informing a broader community of future ontology users.

5.1 – Knowledge in The Vision of Ontology

This section serves to outline the vision of a unified geosciences in which integration is always mediated by information technologies. In this vision language, concepts, data, tools and even practice come to be linked through knowledge mediation. Knowledge mediation, and in particular ontologies, will produce links amongst the heterogeneity of the earth sciences, epistemically, methodologically and practically.

In order to understanding GEON's knowledge representation component it will be necessary to comprehend its technical trajectory. First, the following section begins with a 'pop example' of knowledge mediation. In knowledge representation circles this is a commonly used device for conveying the imagined functionalities of an effective knowledge mediation system. Secondly, this section outlines a *walk-through* of knowledge mediation as drawn from GEON's funded proposal. This example shows a 'plausible user scenario' of an ontology in action. For GEON participants it serves to produce a model of the user, a guide for the construction of ontologies, and it puts forth a vision for how these will come together in scientific practice. For us it is a maps a of process and mechanism for reconciling the general and the specific. Finally this section will describe the dovetailing of 'social science' and knowledge representation. The insights and empirical research on ‘knowledge’ within philosophy, sociology and even STS have become intertwined with ongoing research and application of the technology of ontologies. The result is a novel experiment, an intervention on ‘knowledge’ itself (Hacking 1983).

Capturing Ontology

In their economic analysis of the history of the American electrical network David and Bunn describe ‘gateway technologies’(David and Bunn 1988). These are a good heuristic for understanding how ontology and the goals of knowledge mediation have been conceptualized in GEON. A gateway links between existing standards, schemas or formats permitting these to work as though a single operating unit. Gateway technologies are “some means (a device, or a convention) for effectuating whatever technical connections between distinct production subsystems are required in order for them to be utilized in conjunction, within a larger integrated production system.,” (David and Bunn 1988:170). David and Bunn outline a simple example of an electric adaptor which permits you to plug in, say, your laptop in the U.S., England or Continental Europe[1]. Gateway technologies have become relatively common in our lives, particularly within informational applications, such as disk drives that read both PC and Mac formats or software such as Trillian which permit a single interface for Microsoft’s (MSN) and America Online’s (AOL) instant messenger.

1 David and Bun distinguish two sorts of ‘subsystem compatibility’:
i- compatible complements, that is, when subsystems A and C can be used together (e.g. plug and socket, as above), and/or
ii- compatible substitutes, that is, when subsystems A and B can each be used with third component C to form a productive system (e.g. IBM clones with DOS).

Egyedi has applied the notion to internet standards such as the Extensible Mark-up Language (XML) and the Open Systems Interconnection standard (OSI - See appendix A). She notes that gateway technologies (which she considers a type of standard) add flexibility (Hanseth, Monteiro et al. 1996), permitting local adaptation and tailoring while maintaining compatibility with the larger system. Gateway technologies increase the ability of information to move across heterogeneous platforms, environments and institutions.

Egyedi provides a tripartite division of types of gateway technologies by the 'scope' of the solution; I will outline these very briefly. A single instantiation of a gateway which provides a specific compatibility (such as with the plug adaptor above) is a 'dedicated gateway'. These technologies link between two specific standards. If a gateway technology is in the form of a standard or language, thus permitting varied content, it is "generic". For example the Extensible Mark-up Language (XML) which provides a description language for documents. IT provides a reference model for the definition of headers (or fields) the content of which is left open. Finally a 'meta-generic' standards can generate compatibilities amongst lower order gateway technologies, such as generic and dedicated e.g. OSI. (Egyedi 2001). A meta-generic gateway permits changes to the standards by which it is composed.

Egyedi makes the seemingly counterintuitive argument that standardization can lead to greater flexibility, particularly if the standards are designed in such a manner as to support linkage through gateways. Gateways link across existing technical standards; to the extent that they are generic they are able to be extended to new standards, and add flexibility across the limits of individual standards: “the inherently generic character of the standard is seen as the technology's main revolutionary feature,” (Egyedi 2001).

The notion of a ‘gateway technology’ comes very close to actor’s understandings of ontology and its enabling consequences. In the introduction of this dissertation I described the architecture of CI, in which the application layer had the dual ‘responsibility’ of linking to the conduct of science and to other applications. Thus the application layer should act both as a user interface, and as a technical gateway amongst software applications or hardware platforms.

Similarly, a formal ontology may link between existing data standards or schemas, providing compatibility across them for either the computational system or the user. Flexibility is added because there is no need create a single standard, but rather existing standards can be navigated. Similarly, ontologies permit crossing multiple databases or domain specializations by automating translations of technical, conceptual and linguistic difference. This is the vision of ontology I will capture below.

There are two subtle differences relative to ontology that must be outlined with gateway technologies as described by David, Bunn and Egyedi: i- within GEON ontologies are not usually thought of as standards and ii – while ontologies are seen to be a technical solution, they do more than provide a technical interconnection, they also provide a richer interface for humans.

A laminated phrase in GEON has been “we don’t make standards”. Participants do not consider the goal of GEON to produce data standards, or even functional standards for software such as visualization. Rather, the goal of GEON is mediation. Ontologies connect diverse concepts, languages or knowledges – they do not define new ones.

Unlike Egyedi, who takes gateway technologies to be standards[2], GEON's ontologies are understood to link between existing standards[3]. Both David and Bunn, and Egyedi, describe gateways as *technical* interconnections between subsystems. However there is a specificity to the technology of ontology which differentiates it from a narrow understanding of 'the technical': ontologies have the goal of capturing semantic content and relations. They do not produce links between 'subsystems' alone, but also between people, domains of expertise, diverse languages and concepts. Ontologies are "knowledge mediation."

While in the philosophical roots of the term ontology is a theory of being, within computer science an ontology is a software-based representation of a particular knowledge domain. By capturing key portions of knowledge the computer system has available a much richer set of terminologies, concepts and linkages with which to execute queries. Human-computer interaction becomes easier and more sophisticated because the computer operates with categories and relationships which come closer to those of the user. Knowledge can be articulated, represented and encoded; it can be translated and automated; however, within GEON's knowledge mediation efforts the human is never

---

[2] David and Bunn treat gateway technologies as particular kinds of standards, and then follow to analyze these based on their consequences: "once a gateway innovation appears on the scene, even though it is technically neutral in its ability to make use of the competing technologies, its introduction need not be neutral in its impact on the competition between contending variants. It may disturb a delicate 'balance' of market advantages based on the heterogeneity of specialized user-needs, and so assist one particular variant to emerge as the standard for the enlarged network. In the case at hand, the system which happened to hold the cost advantage" (David and Bunn 1988:198).

[3] NOTE: there is a great awareness by the GEON knowledge representation team that this is not the only avenue in the development of ontologies. IT PIs will occasionally speak of movements in the broader field of knowledge representation to create standard ontologies (sometimes described as 'top-down ontologies' or 'upper-ontologies'). This approach is rarely discussed with Geo PIs, and IT PIs explicitly divorce themselves from these approaches. In this chapter I focus on efforts within GEON, capturing the debates of the broader field of ontologies is outside the scope of this dissertation. For examples of such efforts see, for example, cyc.com .

fully removed from the equation: ontologies serve to facilitate research, the discovery of resources and multidisciplinary communication and collaboration. Ontologies are an informational substratum and a tool, the goal is to help connect the heterogeneous expertises of domain experts. These are the purposes for semantic integration.

--

Let us begin with actors' definitions of ontologies: "ontology specifies a theory (as set of models), defining and relating, concepts representing features of domain of interest," (IT participant). This general language is characteristic of knowledge representation circles: 'theory', 'relating', 'concepts', 'representing' and 'domain' can each refer as much to the geosciences as to expertise about wine, music or banking. Domain knowledges are particular, but knowledge is general

This initial definition of ontology misrepresents the existence of a consensus across knowledge representation experts. Actor's categories are at play amongst participants: the definition of ontology, or what it is for, remains highly unstable within computer science, as well as in knowledge management. Many GEON participants are also participants in the broader knowledge representation community, and so the understandings of ontology within GEON have reflected many of the meanings which are.

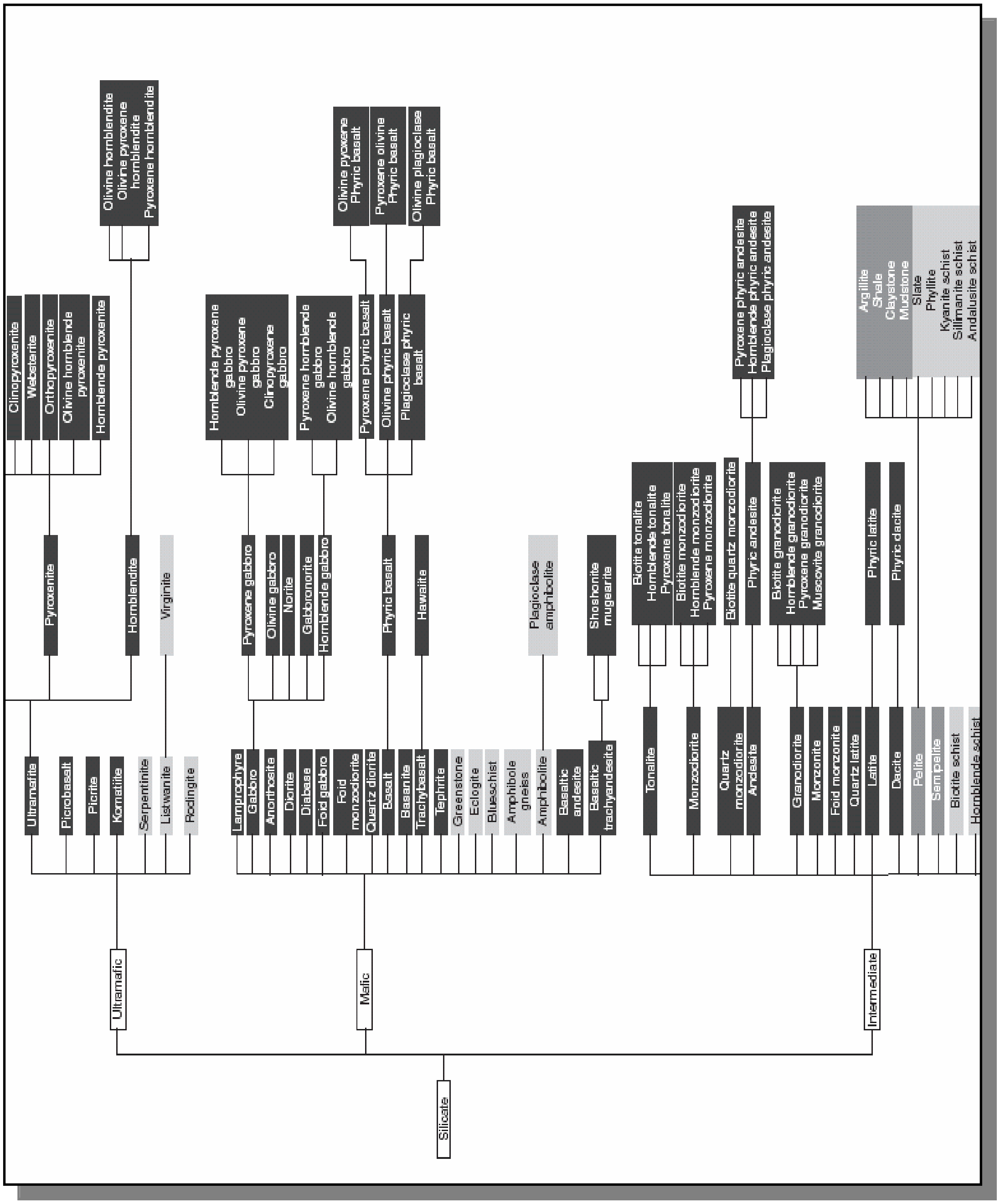

Figure 15: A selection from a rock classification ontology. This image describes the concept 'composition' in geology. This document was taken from an introductory presentation in GEON for earth science PIs (see below). The ontology itself was developed by the Geological Survey of Canada. The ontology is used regularly within GEON efforts.

currently at play within computer science. Within computer science ontologies themselves remain a going concern. Are ontologies realist, utilitarian or pragmatic (Sowa 2000)? Can automated deductive reasoning produce reliable new knowledge(Smith and Mark 2001; Smith and Welty 2001)? In this chapter I put aside these debates, and instead trace the discussions of my informants: 'what has ontology been for GEON?' It is more revealing to trace the specific enactments of ontology relative to a project than in the abstract.

Below are excepts from oral presentations and their accompanying slides as used by IT practitioners presenting for the principal investigator (PI) team of GEON in its second year. At this point, two years into the project, ontology had already become a relatively familiar term for domain scientists and the discussion is relatively sophisticated (we will see simpler understandings below):

> Other terms which loosely overlap with ontologies:
>
> controlled vocabularies, database schemas, conceptual schema, thesauri, taxonomies, inform/semi-formal representations (concept spaces, concept maps, labelled graphs, semantic networks), formal ontologies (OWL, formalization of a specification, constraints of terms).
>
> -- IT PI in a slide presentation

> What are ontologies used for?
> 1-conceptual models of a domain or application
> 2- classification of concepts, data/object instances
> 3- analysis of ontologies, automated reasoning, deductively derived knowledge
> 4- targets for semantic data registration
> 5- conceptual indexes and view for searching, browsing, querying or, integration of registered data
>
> -- IT PI in a slide presentation

In use the somewhat bombastic languages of knowledge representation are substituted by those of application. Ontologies will serve similar roles to taxonomies or thesauri only with a degree of automation and permitting access to the computational 'system'. They will formalize knowledge models in the sciences. They will link data and resources. Large claims still, but not quite so grand as ontology in the abstract (Guarino and Welty 2000). I briefly elaborate these five *uses* of ontology below, following this each understanding of usage will be clarified through the pop example and the walk through below.

The first three items from the second excerpt describe ontologies as clarifications of domain knowledge. They are explicit and visible statements about a domain's concepts, providing a manipulable model. Ontologies classify concepts, ordering them in hierarchies and taxonomies. Finally, a formal ontology, specified with logical operators, can be used for automated reasoning. New knowledge, or implicit knowledge, can be derived from such formalizations in combination with 'reasoning software'. Ontologies may not only be tools, they can serve as objects; an analysis of a domain ontology or multiple domain ontologies can serve to learn about changes or progress in the field: what subfields are most active? which are most rich?

The third and fourth items relate to data and interoperability. Semantic data registration refers to linking specific aspects of a database ('the meanings of a database') to the ontology. This will permit 'conceptual navigation' of a database or multiple databases. An ontology will serve to integrate data by permitting a user to navigate using their own conceptual categories; technical material from a *different* field becomes accessible by translation into a familiar domain terminology.

--

Imagine, for example, a personal collection of wine categorized by type: red, white, rose. This is perhaps the most common lower order categorization of wine, used in vernacular conversation, on a wine list in a restaurant or to arrange a selection in a small store. While these classifications may suffice for such collections a fine restaurant, or a specialized wine shop with a large collection, may find the category 'red' to be crude and ineffective for their purposes. Similarly a wine connoisseur, a specialist or expert, with an extensive knowledge of wine may find these categories incapable of mirroring the nuance of their experience. The category 'red' could then be broken into cabernet sauvignon, merlot, pinot noire and so on; a more detailed classification reflecting an expert domain knowledge and permitting manipulation of the wine collection at a finer granularity.

We now have two schemes: the simple red, white, rose, and the more detailed breakdown of red. An ontology links between these classification schemes. By including the broad categories red, white, rose and the fine categories cabernet sauvignon, merlot, pinot noire an ontology maintains particular *relations* between the categories (e.g. cabernet sauvignon *is* red). Thus if we have two wine collections (databases, wine lists) -- one operating solely with the category 'red,' and another with more detailed distinctions cabernet sauvignon, merlot, pinot noire -- an ontology can facilitate a search across both collections; 'find Chianti' will search in both sets of categories across the databases because they are linked in the knowledge representation.

This is in fact nothing more than a taxonomic description: a hierarchically ordered classification system. However a wine ontology could also contain information as to region of origin, producer, type of grape(s), and also links to foods with which a wine is conventionally associated. Combined with the appropriate software tools such an ontology may also enable ‘reasoning’. A ‘reasoner’ is a general term for software that interprets and operates on the data in a knowledge representation. For example, if only given the details of a Pauillac whose maker is Chateau Lafite Rothschild, together with other statements in the ontology, the reasoner will deduce additional facts: for instance, that this a Medoc wine from Bordeaux, in France, and that it is red. Conversely, given a particular meal, the reasoner could also suggest appropriate wines: pasta with a spicy red sauce will have a wine color restriction of red. The reasoner follows the coding scheme, or ‘logical operators’, of the ontology: taxonomic relations such as “Chianti ‘is a’ red”, or limits such as “spicy sauce ‘is restricted’ to red”.

Popular examples often compare ontologies to taxonomies or to thesauri: hierarchical relations or equivalencies of meaning are captured in these ontologies. However knowledge representation experts will distinguish these on the basis of their level of “formalization”: ontologies are encoded in languages (or code, or logical operators) accessible to the computational system. In other words, ontologies can serve in automated software operations such as reasoning.

Describing the ‘pop example’ serves a dual purpose here. Firstly it serves to introduce the reader to a very rudimentary model of knowledge representation and the vision of integrating information. Secondly, however, it serves to introduce the reader to the *methods* by which GEON earth scientists themselves were introduced to ontologies. During the kick-off meeting, the first appearance of ontologies was explained using the example of wine. Curious geoscientists were directed to the popular online article ‘Ontology 101’ (http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html -- accessed Sept. 4 2006) which also draws on the example of a wine ontology and the reasoning software 'Wine Agent'. This first presentation to earth scientists concluded with the statement:

> You can use Wine Agent without having to know anything about wine. If you want to know something about a bottle you have, just input the information and it will tell you the rest. *If you don't care about all that*, and you just want to know what wine goes with your meal, you can do that too. (GEON IT PI, emphasis added)

No further conclusion or direct statement about earth science knowledge was specifically drawn out in this short presentation, however, the suggestion is clear: just as the social worlds of wine are complex and can be more easily navigated with knowledge mediation tools so too are the geosciences complex in their language, knowledge, concepts and technical data, and navigating these can be facilitated through ontology development.

Very little further articulation of this argument was needed. Domain/ domain difference is a laminated boundary in GEON. Following the Atkins vision of CI, and what I have described as the logic of interoperability, the proposal was funded by earth and computer scientists who envision a unified technical substrate for the earth sciences. As we have seen the GEON project is premised upon extant disciplinary difference within the earth sciences, upon a problematization of these differences in the face of multidisciplinary collaboration, and a solution to this problem through information technology enabled mediation. We have seen the geosciences characterized as technically backwards and in danger of losing its data resources; we have seen geoscientists characterized as interested in basic research rather than general resource development, and; we have seen the various disciplines characterized as specialized in their methods, language, modes of representation and topical interests. The boundary work of disciplinary difference and its problematization is coupled to the boundary relations established at the intersection of IT and domain. Knowledge mediation – ontologies – will serve to provide the linkages between balkanized data, software and human experts.

Pop examples serve to crystallize the logic of interoperability. References to music or wine resonate with participants. They can identify a specialized knowledge, that it is unfamiliar to (most of) them, and can imagine the benefits of navigating foreign knowledge domains with such tools. An understanding of the disadvantages of such examples only comes later in the process of understanding the problematic of interoperability.

In the next section I will return to the complex interchange between computer and earth scientists at the kick-off meeting as they were introduced to problematic of interoperability, and the solution of ontology. Below I *elaborate* (see Ch. 1.2) on the 'walk through' example of knowledge mediation provided in the GEON proposal. The proposal offers a vision of technical functionality planned for the geosciences and specifies the resources and work necessary to achieve such a vision (see also user scenarios Ch. 4.1 and configuring the user Ch. 4.2).

This user scenario articulates an ideal typical vision of what ontology enabled knowledge mediation in GEON will look like at the completion of the project. It models a scientific user (Woolgar 1991) interacting with a system on the basis of conceptual queries. Rather than sorting through collections of data, the user begins with a query based on geoscientific theory: 'the Wilson Cycle'. The Wilson Cycle is Plate tectonics 'at its simplest'. Plate tectonic theory is one of plates rifting into pieces diverging apart and new ocean basins being born, followed by motion reversal, convergence back together, plate collision, and mountain building. This cycle of opening and closing ocean basins is the Wilson Cycle. However the Wilson Cycle is usually understood as historical geological theory, rather than an organizing principle for data or databases. The novelty or 'new functionality' in this walk-through are the *methods* which the user is imagined to access specific data. Rather than beginning with data and applying it to theory, a search query is initiated from a concept and the system permits the user to 'drill down' to the data.

In addition to articulating this vision of use, the walk through also defines the methods and resources by which such a knowledge mediation system will be developed. This includes:

> (a)- what is required to meet the interoperability goals: the user scenario defines the resources for mapping knowledge which include capturing domain knowledge, formalizing this into ontologies and generating links amongst heterogeneous resources such as databases, other ontologies or visualization outputs;
>
> (b)- identifying the necessary actors to accomplish the goals: this includes information technologists, domain scientists, and a broader community who must 'accept' and take-up the ontologies. Notably it also includes partner project and those institutions of geoscience which govern data.
>
> (c)- how the goals will be practically executed: in the form of knowledge capture sessions and formalization in representations. These ontologies must be linked to databases, to other knowledge representations and to the institutions of science. Less developed, but still present, is a vision of human-computer interface design.

Resources, actors and practical action are defined in tandem with a technical vision of 'revolutionized' scientific practice. In the proposal this walk-through is articulated in five steps, an example of "semantic data integration and concept navigation". A diagram, also attached below, illustrates the relationships between theory (Wilson Cycle), the GEON knowledge representations, particular databases and partner cyberinfrastructure projects (i.e.Chronos). I include direct quotes of these five steps and elaborate on each:

> [1] Example: Semantic Data Integration/Concept Navigation. Figure 2 depicts a GEON semantic data integration and concept navigation scenario applicable to data integration across the two test-bed regions.

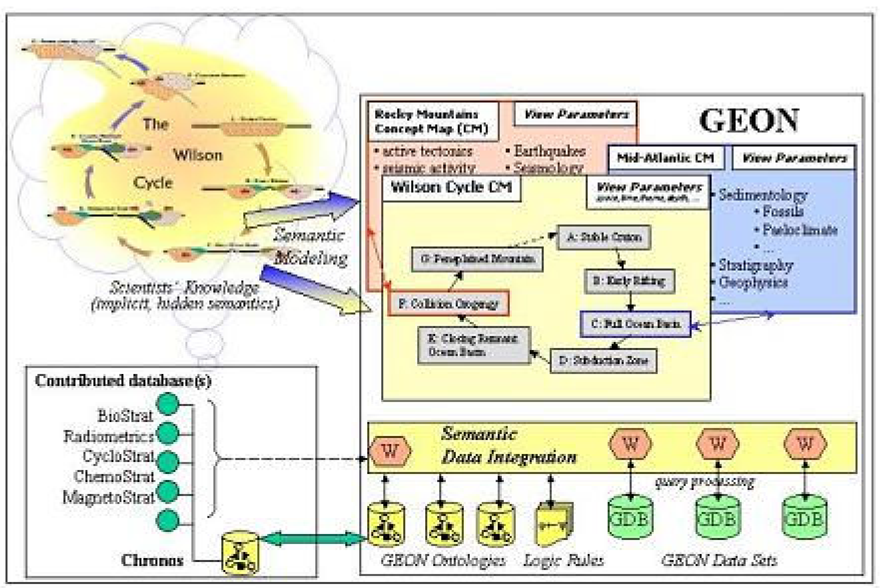

Figure 16: From the GEON proposal, a user scenario of knowledge mediation. Originally entitled 'System for Modeling Hidden Semantics and Situating Databases.'

The GEON project is divided into two physical 'test-beds' (see Ch. 4.1): these are sites of contemporary geoscience expertise and interest, located within the US and serving to "define the GEON geoscience challenges," (GEON Proposal p.1). These test-beds are the mid-Atlantic and Rocky Mountains. This example of semantic integration is of scientific work across the two test-beds -- more specifically the datasets available describing those test-beds -- which are then integrated through a concept map of the Wilson Cycle.

In the diagram above, the Wilson Cycle is first represented in the left hand corner as a domain knowledge 'think-cloud,' then to the right as a concept map, and between these is the activity of 'semantic modeling.' Semantic modeling is the process of transforming 'knowledge in the wild' into system accessible formalisms.

> [2] Our goal is to explicitly model the scientist's implicit knowledge and "hidden semantics" via formally specified ontologies and associated logic rules.

Knowledge is understood to reside in the minds of domain experts, however there it is 'informal', often remains implicitly understood amongst practitioners and may in fact be 'tacit'. Thus the task of semantic modelling cannot be easily delegated to a technical or administrative staff: the deep knowledge of database structure and a history of empirical inquiry is held in the hands of particular experts in a domain community. Ontology experts have developed a sophisticated language of inquiry around knowledge representation – often they draw on philosophical, anthropological, cognitive science or linguistic models of knowledge function and transfer (see below).

In the diagram above that the Wilson Cycle is the highest order of modelling. In this example it is connected only to other 'lower order' concept map. The ontology which represents the Wilson Cycle model is connected to more ontologies of a finer granularity. The links are to the two concept maps of the test-beds which in turn contain semantic descriptions of the databases to which they connect. These descriptions range from geoscience entities (e.g. fossils) to disciplinary associations (e.g. stratigraphy). An ontology can form links between such heterogeneous entities. No particular details of the link between fossils and stratigraphy is offered ('why do fossils matter for stratigraphers'), however this link has presumably been generated through the techniques of knowledge capture.

The Wilson Cycle concept map is linked specifically through portions of its descriptive content. It has been defined in five stages: A) Stable Craton, B) Early Lifting, C) Full Ocean Basin, D) Subduction Zone, E) Closing Remnant Ocean Basin, F) Collison Orogeny G) Peneplained Mountain. In the example the Rocky-Mountain test-bed is linked specifically to 'collision orogeny' and the mid-Atlantic to 'full ocean basin'. In turn the Rocky Mountain and Mid-Atlantic test-bed concept maps are linked to particular datasets of these regions.

> [3]Each GEON database (GDB) is then "wrapped" with software that provides both syntactic wrapping (e.g. XML representations) as well as semantic wrapping (e.g. indexing via an ontology). […]

Syntactic and semantic 'wrapping' provide different functionalities for the user in this scenario, mirroring the orders of database structure (see Appendix A for data structure, or (Sheth 1999) for an accessible introduction). In a semantic wrapper the database is described at the level of its meaning, here a relationship between theory, region and datasets. This permits an automated query to find two (or more) datasets. The system has a representation of the syntax and structure of the database available and is able to translate a single query for the differing organizations of the databases. Below is a common 'pop example' of syntactic and structural integration used within GEON (fig.17). I use the term ‘pop’ despite the technical geologic content, because the example relies on a translation of date formats, rather than conceptual knowledge. The illustration itself is drawn from the Ecological Metadata Language (EML) from the community Long-Term Ecological Research (LTER) (Millerand and Bowker forthcoming; Millerand and Ribes forthcoming):

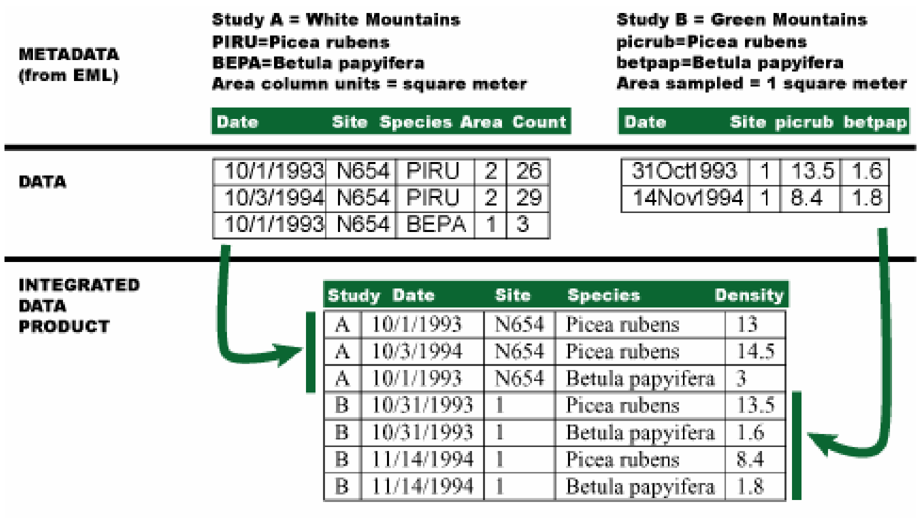

Figure 17: Example of data integration relying on the EML data standard. Drawn from an IT PI slide-set directed at the education of earth scientists in the problematic of interoperability.

In this example, two datasets have differing syntaxes for the date (left: MM/DD/YYYY and right: DD/MMM/YYYY), but because each dataset is ‘wrapped’ (using the EML metadata standard in XML) the system is able to query both datasets and produce a single synoptic representation for the user but drawn from two datasets. This is the chart below, the ‘integrated data product’. Formal maps, or schemas, of database structures permit automated integration.

Returning to the GEON example, a semantic wrapper will provide linkages at the level of meaning. Datasets are linked to a concept map based on whether they are ‘about’ the mid-Atlantic test-bed and then ‘about,’ say, fossils or paleoclimate. In turn both are defined in the ontology as 'relating to' sedimentology. The rationale for these links is not provided in these representations, it is assumed that a domain knower will have the necessary interpretive knowledge in their head to make sense of the links and chose amongst possible 'navigable' knowledge pathways:

> [4] In our example, it may not be clear to the non-expert whether and how the mid-Atlantic and Rocky Mountains test-bed regions are related to each other. However, the scientist's domain knowledge immediately allows her to relate the two, since they occupy different moments in the evolutionary path of the Wilson Cycle. The mid-Atlantic region is at stage C (Full Ocean Basin) while the Rocky Mountains region is at stage F (Collision Orogeny) in the Wilson Cycle.

Here is should be clear that the human knower is not fully removed from the equation. An ontology can capture a portion of domain knowledge. However, at some point in 'navigation' a level of familiarity with the domain will be necessary. This user scenario is described as 'concept navigation.' In short, semantic integration is about facilitating a user's search for data. However the user is not simply an abstracted 'any user' (as in the HCI examples in the previous chapter, see esp. Ch. 4.2), here it is a geoscientific user, familiar with the Wilson Cycle concept. The geoscience user has begun a query not by looking at specific datasets but through the concept of the Wilson Cycle. Beginning at a higher order conceptual level *she* is then able to 'drill down' to the next conceptual levels of the test bed, and from there through, say, sedimentology and then fossils, and then to a dataset which includes relevant data about all of the above: Full Ocean Basin=>Mid-Atlantic Region=>Sedimentology=>Fossils. She can then use this dataset, and to the extent that it is syntactically interoperable can conduct cross-database queries, visualize the data, or conduct other operations.

> [5] As mentioned above, the key to semantic data integration is to make some of this implicit and hidden domain knowledge explicit, as it provides the "glue" to connect otherwise seemingly unrelated data sets, and provides the means to "semantically browse" and query such data sets using concepts, themes, and disciplines (e.g., collision orogeny, stable craton, Wilson Cycle, tectonics, stratigraphy).

'Semantic browsing' has a very broad range of applications. In GEON demonstration software has used ontologies to render maps across state boundaries, to design tailored interfaces by domain speciality and combined with workflow tools automate data processing techniques, to name only a few.

--

The 'user scenario' example traverses three kinds of activities which in temporal flow are encountered discretely: the developmental process, automated activity of integration, and a model of use. The developmental process involves the capture of knowledge through semantic modelling, thus making explicit 'hidden' domain knowledge. Automated integration becomes possible after development is complete, providing a 'glue to otherwise seemingly disconnected' resources, such as data. A user queries by concept, and is able to receive a single integrated product. However such a query reveals only a 'clean' user interface to the geoscientist researcher. Both the particular processes of semantic mapping and the automated activities of semantic and syntactic integration will be opaque/invisible to the user in any particular query.

User-scenarios are powerful planning tools which enact a vision of system which bring together aspects of future users, automation and the methods by which it will be built, as with ‘configuring the user’. Woolgar’s concept shifts attention from the end-product – software which a user may configure freely – to the production process in which ‘a user who desires to search’ is envisioned either through the eyes of the designer or in combination with various user studies which bring forth an outline of the future user(Ch. 4.2.). More than configuring the user, the scenario outlines a plan for knowledge capture and automation: how the tool will be made, how it will function and what people will do with it are all encoded in the user scenario. The user in this example is i) a knowing domain geo-scientist; ii) familiar with a shared conceptual terminology of seismology such as the Wilson-cycle, and; iii) seeking particular data-sets across traditional disciplinary boundaries.

The automated system are the linkages between ontologies, and with datasets, the ‘wrappers’ which provide instructions in conjunction with ‘logic rules’ for semantic exploration and (syntactical, structural and systemic) data integration. This example mirrors the larger vision of cyberinfrastructure, adding greater details of ‘context’, greater detail of integration technology and a temporal development schema. Particular roles are also assigned to the various actors in the scenario. Earth-scientists will be future users of the system seeking integrated data but also the source of semantic information; information technologists (primarily invisible in this example) will assist in building these systems, and the information technology itself will serve as an automated mediator and integrator between people and data.

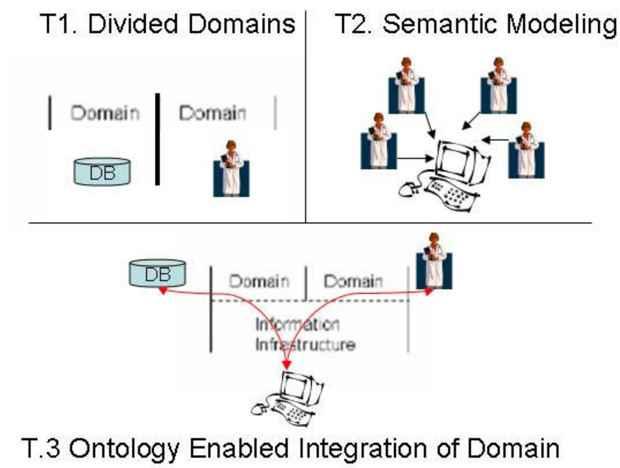

Figure 18: The logic of interoperability as actor's process: T1 Residing in their distinct domains, data and scientist are divided. T2. Semantic modeling of domain knowledge and creation of an ontology. T3. Domains remain distinct, but linked through automated integration.

The developmental practice, or 'semantic modelling', are only vaguely outlined in this example. In writing the GEON proposal the work enacting such ontologies remained to be invented, repeated and eventually routinized. In the case of ontologies these were the 'concept space' workshops, which eventually generated the routine described in the next section.

A note on knowledge 'in the wild': insights at the boundaries of social science and information technology

Following his studies of expert systems in the mid-1980's sociologist of scientific knowledge Harry Collins described the 'artificial intelligence experiment' as "not just a problem of engineering or psychology but an empirical test of deep theses in the philosophy of the social sciences," (1990, p.8). He describes how artificial intelligence abounds with implicit theories of mind, knowledge, cognition and the social order. Collins was correct in noting a confluence between social theory, psychology and engineering in artificial intelligence circles. However, in my research there has been no room for 'discovering the social' amongst knowledge representation specialists: the research, terminology and even the methods of the social sciences are explicitly referenced on a daily basis. In GEON even geoscientists have come to begin their presentations with epigraphs by Whitehead; dinner conversations following PI meetings have dismissed or embraced Wittgenstein; and, although I have never heard Polanyi's name, every GEON participant has come to understand that knowledge may be tacit.

Today's ontology work is informed by complex understandings of knowledge, of the practice of expert work and of the design of information systems. Earlier efforts in artificial intelligence, expert systems, and automated reasoning were plagued with methodological underdevelopment: the knowledge bases generated were frequently both sloppily constructed and very expensive to build up; new information gleaned by crude knowledge acquisition techniques was highly problematic and poorly substantiated, and rigid logics and design demanded untenable practices by the domain (for a critique of early knowledge engineering methods see Stefik and Conway 1982).

Recent work has attempted to ensure more robust methods for knowledge capture, acknowledging the importance of understanding the domain at a fine granularity and that the very question '*what is knowledge'*— the core of epistemology – is both a contentious and seminal question. Supplementary methods have been imported from disciplines as diverse as anthropology, sociology, psychology, cognitive science and of course philosophy. Qualitative methods, including ethnography, have become a staple of knowledge capture techniques, appearing in its various guises as participant observation (Meyer 1992), expert elicitation (Forsythe and Buchanan 1989), on-site observation

(Waterman 1986), or apprenticeship learning and teachback interviews (Boose 1989). These efforts have paralleled attempts in the larger computer science and IT community to include user studies and participation in the design process (Schuler and Namioka 1993; Star and Ruhleder. 1994; Mackay, Carne et al. 2000; Oudshoorn and Pinch 2003).

I do not wish to argue that GEON's knowledge representation circles have a conception of knowledge which matches that in the contemporary sociology of science or science and technology studies (STS)[4]. In fact, understandings and use of 'social science' research and methods within computer and information science is often received with great ambivalence (Dourish and Button 1998; Dourish 2006). But while these uses may differ, and could easily be criticized from within social science's various disciplinary standards, it is important to keep these uses in mind. That is, that for some time now social and computer science researchers have been engaged in an extended dialogue.

My purpose is first to give credit where credit is due: knowledge representation, as a field and as practitioners, should not be characterized as naïve or 'asocial' (c.f. Forsythe 1993). Secondly, and more importantly, the uses of the social have clear consequences within the practical organization of knowledge representation. For example, by explicitly understanding knowledge as the site of contestation, or controversy (Collins 1981; Scott, Richards et al. 1990), knowledge capture efforts in GEON have made conscious efforts to bring diverse participants to the table, to engage in outreach to the broader knowledge community, and have actively sought to gain support

---

[4] Of course, it would be foolish in the extreme to believe that STS has come to any form of consensus on the question of 'what is knowledge' (Callon 1991; Zammito 2004). It would be difficult even to identify a common sensibility towards knowledge, and there are many researchers who have preceded my own position and simply taken knowledge to be an actor's category rather than an independent phenomena to be explained (Lynch 1992; Lynch 1993; Latour 1999).

from the institutions of earth science. The laminated phrase in GEON 'we don't make standards' emerges specifically from an understanding that standardization is a politically charged activity (Bowker and Star 1999)[5] with downstream consequences for users and future system development (Hanseth, Monteiro et al. 1996; Monteiro and Hanseth 1997; Abbate 1999; Epstein (under review)).

Epistemography

When we think of knowledge we often think of formal encodings first: in science we think of textbooks, articles and other publications; in business we may look to statements of work and paper trails; in law to templates and precedent. Secondly, we may think of experts themselves as repositories of knowledge e.g. scientists or engineers. In researching knowledge work in action, sociologists of scientific knowledge have followed scientists and engineers across a surprising diversity of locations and kinds of activity (Star 1995). Sociologist of science Steve Shapin (Shapin 1989) opened the ground for looking at the technicians in the work of Robert Boyle and his air-pump. Shapin points to technician's significant expert work in knowledge production and simultaneously notes a moral order of the historical period which left them out of most accounts: 'invisible technicians'. Historian of science Peter Galison (Galison 1997) has extended this work in his studies of 20th century microphysics. Galison has remarked that modern physics has relied on complex collaborations between experimentalist and theoretical scientists along with another group of materially oriented technicians which

[5] That integration too may have its pitfalls is less well articulated in GEON, as it is in the social science literature.

support the work of both. Cambrosio and Keating (Cambrosio and Keating 1988) in their studies of molecular biologists have noted that much knowledge is distributed in local practices and methods of communication, accessible to negotiation and argument, but difficult to codify. Below I offer a typology of some of the most commonly used terms in STS to describe knowledge.

Knowledge:

- **has a *tacit* dimension** not immediately available to discursive expression (Polanyi 1964; Ravetz 1971)
- **is *distributed*** between people, things and systems of action, including formal and informal procedure (Douglas 1986; Latour 1988; Hutchins 1995)
- **is *immanent or local*–** it is developed in response to certain problem sets (Polanyi 1964; Lakatos 1976) and requires a skill for movement between the specific and the general (Kuhn 1977; Forrester 1996)
- **is *embodied*** – it is known as an action-in-practice not immediately accessible for transformation into a formal set of propositions (Lave 1988; Shapin 1989)
- **is *heuristic* –** knowledge claims may serve to help understand a phenomena, even when practitioners question claims of faithfully representing a reality (Sismondo 1999; Bowker 1988)
- **is *multiple* –** each sub-domain has its own set of methods, tools, heuristics and criteria for truth – styles of reasoning and roles for logic may vary (Livingstone 1986; Hacking 1990; Messer-Davidow, Shumway et al. 1993)
- **is *contentious* –** knowledge is often the focus of debate, struggle and contestation – science is both a repository for knowledge but also a site for its development and review (Collins 1985; Burchfield 1990)

Thus social studies of science have 'found' knowledge in many sites, and has come to characterize it in diverse ways, but more importantly the field has fostered a methodological tradition of openness to locations of epistemic development. By shifting the unit of analysis from formal knowledge and experts to *relationships* between these, a more complex image of knowing has emerged, the object of which is not to pin down knowledge, but rather to track its production, shifts and eddies. There is an affinity between ontology builders and sociologists of knowledge: both share an interest in ‘knowing’, and a necessity to maintain a broad field of inquiry into the possible repositories and vectors of knowledge.

It is the growing awareness and sophistication within the knowledge representation community itself that has prompted a distanced stance on epistemology in this study. In philosophy, 'ontology' is often coupled with *epistemology* -- a theory of knowing, *how* we know the world, and the characteristics of knowledge. While in this chapter I argue for the importance of 'how to know' by observing the work of domain scientists and IT specialists as they build ontologies, my argument stretches further than epistemology to the larger developmental trajectory for ontologies. This is a study of framing and its practical consequences in organizing. By necessity, then, in this dissertation I have taken as a methodological principle an agnostic, or indifferent (Lynch 1993, see also Ch. 1.2), position towards the question ‘what is knowledge?’ It is precisely this question *which is at stake* in the production of ontologies.

In knowledge representation the object of activity is to root out the location of knowledge itself, to make it available for transformation into discourse and eventually formalization in machine language. In contrast, a *sociology of knowledge representation* takes as its object an entire repertoire of action surrounding knowledge work, what sociologist of science Knorr-Cetina has called an epistemic community (Knorr-Cetina 1999). My own method is not the identification of site of knowledge for acquisition, but instead to follow my informants across the entire range of heterogeneous activities (Callon 1986) which constitute knowledge work. Historian of Science Peter Dear has called this approach ‘epistemography’ (2001).

--

In this study, ethnography is used not to understand the domain, but to understand the development of a routine around knowledge capture, formalization and ontology building. This turn towards the practice of semantic modeling, rather than towards the knowledge of the domain, has enabled observing not only the knowledge configuration of a domain, but also the process of translation into the formalizations and language of ontologies. In my research on knowledge representation work I have found as much contestation about 'what is knowledge' amongst these researchers and geoscientists as within STS.

Neither 'science' -- itself an enormously heterogeneous body -- but also not disciplines such as geology nor sub-disciplines such as geochemistry have come to solid practical or theoretical conclusions as to 'what is knowledge.' In the face of novel instrumentation, methodologies and concepts, this too is a question for scientists. In one GEON ontology workshop a geochemist remarked:

> When we talk about knowledge recovery, we talk of this: the rock record. The rock record covers four billion years of planetary history. […] If we can work on this planet and understand its ontology, we have a good chance of understanding the other ones as well […] These are knowledge reservoirs, this part that is called the lithosphere contains knowledge of different kinds. How do we integrate the knowledge reservoirs?

This comment characterizing 'the Earth' as a reservoir of knowledge sparked a minor revolt. Debates ensued over whether knowledge could be manifested materially in 'mere rocks' or whether it was a human capacity alone. Some suggested that rocks be thought of as the earth's memory, but this also led to accusations of anthropomorphism. Finally, frustrated information technologists turned the conversation *away from "metaphysics" in favour of ontology building* (the irony of the elision was noted by no one). The question of knowledge is at much *at play* in the sociology of knowledge as within knowledge representation and the domain sciences.

Putting these endogenous discussions of epistemology aside for a moment, it should be noted that the vast majority of the work in GEON is much less about such high order discussion about knowledge and instead primarily directed at 'on the ground' local applications of ontology. Knowledge representation work in GEON has focused on building local, domain specific ontologies.

Ontologies in science are seen as an aid to the research process itself. They are often built around those topics which are of contemporary interest to researchers. It is not 'stable textbook knowledge' about the earth that is typically captured in representations, but rather knowledge relevant to the study of contemporary scientific objects.

The objects of scientific research, as such, are themselves not fully specifiable. As historian of science Rheinberger has noted about scientific objects:

> Such entities, then, have a peculiar, paradoxical time structure characterized by "recurrence" [...] These research entities, for the very same reason, do not belong to the realm of deliberate construction either. The mode of scientific existence peculiar to such entities derives precisely from their resistance, resilience, and recalcitrance rather than from their malleability in the framework of our constructive and purpose ends. (Rheinberger 2000:272)

To put it another way, the 'science applications' for ontology within GEON are focused on interoperability and knowledge discovery *at precisely those points which are of* *interest to contemporary geo-scientists.* This means that there can be a great deal of 'heat' at the site of ontology building for the sciences – with the reputations and careers of particular scientists or institutions at stake, not to mention 'truth.'

As the site of explicit epistemic and ontologic development, empirical sociological investigation is facilitated because controversy and disagreement is drawn forth and made observable by informants themselves. These disagreements, sometimes controversies, come to the fore in developing ontology and in the process of outreach to 'the broader community' of geoscience knowers. However, even disagreement and controversy can be routinized in the iterative development of ontology.

5.2 – Generating a General Routine for Specific Knowledge Capture

The user scenario and walk through in the GEON proposal is an initial outline of activity. Enacting this plan is a work of situated organizing (Suchman 1987). Over the first three years of the GEON project many more forms of work were required to implement the development of ontologies than are outlined in the proposal. Participants had to face these novel situations, identifying problems in situ and developing ad hoc solutions. Eventually, as problems and solutions came to be characterized as typical of ontology building in GEON, and as 'best practices' emerged, these came to be laminated as routines. In this section I explore the emergence of a *routine for ontology building.*

Organizational routines are "repeated patterns of interdependent actions, performed by multiple actors," (Feldman 2000; Feldman and Pentland 2003)[6]. In this

---

[6] Feldman has traced the notion of an organizational routine back to Stene (Stene 1940), as well as to the 'founding fathers' of organizational theory (March 1958; Thompson 1967; Nelson and Winter 1982). The notion of routine carries significant theoretical baggage, I would like to immediately dismiss two

chapter I focus the emergence of an organizational routine in GEON through actors' talk (or accounts, Boden 1994), in the practical work of organizing (Weick 1977) and in material arrangement. The routine for ontology building includes technical action, such as knowledge encoding, but also a broader set of activities that stretch from introducing the technologies of ontology to the domain, through what is learned by a practice of building ontologies, and to the mobilization of a future domain user community.

--

The goals for GEON stretch beyond a single ontology implementation: it is not one discipline's knowledges to be mapped, a single set of databases to be integrated, or a tool to be tailored to user community. Rather, the project includes multiple ontology development efforts for diverse purposes and representing several communities. A routine emerged because of the iterative quality of the activity (Nelson and Winter 1982), and the active identification and then incorporation of 'best practices' from previous iterations (Seely Brown and Duguid 1991).

The routine for ontology development was not planned but emergent (Ashforth and Fried 1988; Cohen 1991). At no point within the first year were there attempts to formalize the process of ontology development. Over time, by the second year,

---

understandings of routine – as static and as evolutionary – in favour of a practice centered view of routine. Routines have often been understood ass a source of 'inertia',(Hannan and Freeman 1977; Weiss and Ilgen 1985) inflexibility (Weiss and Ilgen 1985) or mindlessness (Ashforth and Fried 1988). This is rooted in a structural view of routines, understood as distinct from practice or enactment. In a situated or practice centered view of organizing (see Ch. 4.1) these may certainly possibilities for routine but they are not necessarily so. The stability of routines is a product of actor's work in generating continuity (Boden and Zimmerman 1991), accountability (Garfinkel and Wieder 1992), and through material arrangement (Latour 1995). Routines are also heavily associated with an 'evolutionary' view, where they have been described as the 'genes' of an organization (Nelson and Winter 1982) and competing by mechanisms described as 'selection,' ' variation,' or 'selective retention' (Cambell 1965). Aside from an unnecessary metaphysical baggage understanding routines as operating by natural selection or as 'genes' is too rigid a metaphorical view. An interactionist account of routines must treat them not grant them an ontological existence (Taylor and Cooren 1997) but treat them as products and accounts and then observe the work of their maintenance.

participants came to perceive the outline of a routine and then began to articulate prescriptions: “we need to make sure that for the first and second round of semantic modelling people have to come to San Diego. They can’t just do it from where they are, they can’t just do it over the phone or even video,” (IT PI). Spoken, a routine can be referenced, and actively manipulated. The routine came to be embedded in material arrangements such as schedules, lines of communication and even preferred rooms at the San Diego Supercomputer Center (SDSC).

The activities here called a routine eventually came to be identified as such by participants (first as the ‘ontology workshops’ and then later as ‘community centered ontology development,’ see below), however it remained flexible and, of course, had to be locally enacted for each iteration (Feldman 2000). What I elaborate here is a trajectory (Strauss 1988; Strauss 1993) for the routine, which only became explicit in actors’ talk over time and then became loosely formalized in the organizing activities of members.

For heuristic purposes I divide the routine into three components, although in practice they should be understood as iterative:

i – *learning ontology*: this is the first step in having the domain come to understand ‘what is an ontology’ and why it is relevant to scientific work. Here participants learn the *problematic of interoperability*. This was conducted formally through presentations by the IT team to the earth scientists, and informally through discussion of the purposes ontology may serve. Collectively both IT and domain participants attempted to identify of the preconditions for ontology work. From these initial meetings the IT team also began to understand the relationship of the domain to data integration projects, or ‘data politics’;

ii- *the practice of capturing knowledge*: abstract descriptions of ontology building are not sufficient to assist domain geo-scientists in formal knowledge representation. Practical learning, or learning-by-doing, was required in order to have domain scientists begin translating their knowledge into inherited categories, logical operators, and predicates. It is at this head-to-head encounter between IT and domain that the configuration of knowledge within a domain becomes apparent to the ontology expert. Practical learning and formalizing (or ‘semantic modelling’) occur hand-in hand. Domain experts are often initially unaware of the particular configuration of consensuses, ambiguities, ambivalences, or disputes within their fields. IT experts must be prepared to offer one of multiple solutions available to assist in the formation of temporary agreements, represent multiplicities of knowledge, disagreement or uncertainty;

iii- *engaging the community*: over time GEON participants come to consider an ontology as successful only if the technical work was coupled with identifying and collaborating with a broader community of future domain users. Identifying a future community and finding means to elicit participation is an emerging skill-set in the routine. A domain community must be made to, explicitly or implicitly, consent to the ‘congealed work’ within ontologies.

As this routine emerged the definition of success for an ontology come to stretch beyond a narrow definition of technological development ('is the ontology complete?') to its uptake and usage by a domain practitioners. Success of the ontology came to be (re)defined along with the emergence of the routine for its production. In particular the third step in the routine, engaging the community, placed the success of an ontology well outside its implementation and instead in its acceptance by a community of 'knowers' and its uptake by a user base. In doing so a broader set of this community was made to understand 'what is ontology'.

The enactment of the routine is not simply iterative by virtue of building multiple ontologies, it is iterative because enrolling a community requires a return to its first steps. The efforts to enrol the community returned practitioners to the educational aspects of the routine. If success is defined by acceptance and uptake in the domain community then that community must know what is ontology, the problematic of interoperability, and the particular solution which is offered by knowledge mediation. To do so, project participants (now IT and domain alike) returned to the pedagogical tools, such as slides, demos and pop examples, used to inform the GEON's geo PIs in the 'first' iteration.

--

Just as with the previous three empirical chapters the work of interoperability is premised upon a double boundary work: demarcating roles for IT and domain practitioners, and characterizing differences of data, methods and language amongst the domains. We have seen how 'configuring the user' in the previous chapter organizes relations around requirements testing (*by* computer scientists *of* the domain – see Ch. 4.2). Similarly knowledge capture defines roles around the boundary IT / domain. The goal of knowledge capture and representation is a resource in organizing along the boundary IT / domain: geoscientists must be made to speak their knowledge in an explicit language accessible to knowledge representation; computer scientists must capture knowledge in the formal languages of ontology. However, in this chapter I focus not on boundary work so much as on its complement: the relations established across these and the methods by which actors encode these relations in the mediating informational substrate of ontologies.

In the detailed pedagogical, technical, communicative and organizational practices of this routine the boundary IT / domain is bridged. A technical outcome, a material arrangement, is sought which will enable the production, even the automation, of boundary relations across domain / domain difference. The routine stretches beyond even the broadest definition of knowledge to the work of *aligning* both *within* the domain, and *across* IT and domain boundaries. By alignment I mean simultaneously a practical know-how for building ontologies but also the substantive work of bringing together the organizations and institutions of a domain.

In the first step, learning ontology, domain practitioners and ontology experts must come to a working understanding (Star 1988) of the goals and methods of interoperability and knowledge representation. This pedagogical relationship is initiated in the initial enrolment but it is only in the learning-by-doing of building ontologies that domain practitioners come to be able to resolve their knowledge into discourse. It is through this process that domain practitioners come to see their work as a commitment (Becker 1960), and feel they must turn to their broader epistemic community.

Engaging the community involves developing techniques for communicating to a larger constituency, methods of 'community building', or 'community outreach'. What we observe is a two fold alignment: first, between *IT and domain participants* as they build ontologies, but we also see an alignment of the emerging ontology and the broader *community of the domain* participants.

Learning Ontology

I've just learned how to say "ontology" and use it in a sentence.

-- Geo NSF Officer

Although as we have seen that the proposal contained plans for ontology development and knowledge mediation it was not until the Kick-Off meeting was November, 2002 (see Ch. 4.1) that the GEON earth science participants were exposed to the details of ontology. It is difficult to capture the vast scope and ambition of this initial meeting of geoscientists, information technologists, education experts, and a lone sociologist. Held at the San Diego Supercomputer Center (SDSC) many of the IT experts were already collaborating on other projects. Meanwhile many of the geoscientists, spanning broad disciplinary differences, were unfamiliar with each other, the IT experts, and the future technologies of GEON including ontologies.

The kick-off meeting served as much as an introduction to the nascent GEON vision of a 'cyberinfrastructure for the geosciences,' as with the participants to each other, to the host organization SDSC, and to the information technologies. The following data is drawn both from the initial GEON kick-off meeting and from the all-hands meeting held six months later.

--

GEON was created with the mandate of producing data and digital resource interoperability, along with the provision of sophisticated search and query tools. The IT team had, from early on, already looked to ontologies as the means for this interoperability, but the task of enrolling the geoscience community remained. At GEON's kick-off meeting knowledge representation specialists unveiled a barrage of pedagogical presentations intended to educate geoscientists in the arcane realm of ontologies and concept-spaces. Many IT-savvy domain scientists were present, along with representatives of endeavours in digital libraries (Digital Libraries for Earth Science Education - DLESE) or national institutions which often double as standard setting bodies such as the United States Geological Survey (USGS). With such an audience the enabling possibilities of ontologies – database compatibility, inter-disciplinary collaboration, controlled vocabulary queries and the potential for softening the politics of standardization – were not lost on the audience. Thus curiosity was piqued in the domain practitioners. This said, unfamiliarity with the technical terms of ontology building, diverging presentation styles between IT and geoscience, and even the kinds of visual representations familiar within computer science but foreign to geo-trained eyes (e.g. software architecture diagrams) resulted in several miscommunications, frustrations and confusions.

The first step in engaging participants for knowledge representation work is the task of posing of the problem which ontologies seek to remedy. While stovepipes, incompatibility, and legacy databases are the everyday headache of computer science, and the bread and butter of ontology efforts, for geoscientists these problems have often been relegated to a little noticed substratum of technical infrastructure. Even the principal investigators of GEON, a highly technically informed cadre, have rarely had to encounter on such close terms its database architectures.

Knowledge representation requires a doubly-technical engagement between IT and domain: the specificities of knowledge must be brought forth and aligned with the logical requirements of ontology languages. In the first four days, with many hours of Power Point aided descriptions and followed by open discussion, it was only possible for the geoscientists to gain an impressionistic understanding of the detailed work of ontology building. A nuanced understanding would only be learned later, in the practice of knowledge acquisition; enrolment served to bring domain scientists on board, to begin a pedagogical trajectory of understanding the problematic of interoperability, and the particular solution ontologies promise.

Enrolment is not unidirectional – while domain begins an understanding of the problematic of ontology, the IT practitioners begin a familiarity with the 'data-politics' of the domain. A predecessor of GEON is BIRN – the bio-informatics network – also centered at the SDSC. Many of the IT experts within GEON are also participants in BIRN – and thus it served as an early model for the organization and technological planning of GEON. One initial vision was to begin to define a ‘unified geosciences language system’ (UGSL) in a similar fashion to the Unified Medical Language System (UMLS) already established in medicine (i.e. metathesaurus, semantic network, lexicon). However, unlike medicine, geology has not had a one-hundred year history of standardizing its specialist language across its various sub-domains (Bowker and Star 1999). While as far back as the 1830's there are tracts calling for the stabilization of a geological language – particularly fossils – it is an odd feature of geology that there has been greater disagreement in the field over time. Consequently a controlled vocabulary would be a substantial undertaking: going out in the field may involve idiosyncratic data collection methods; there are many sub-disciplines in the geosciences which do not have traditions of data sharing; data is frequently considered proprietary. This is true even in organisations such as USGS where data is mandated as public, but in practice many practical barriers stand in the way of dissemination[7]. It is only with a progressive familiarity with the domain that knowledge representation experts can begin to gain a realistic understanding of what challenges are faced, and what solutions are plausible.

Aside from data-politics IT experts began to understand the place of information technologies themselves, including diverging understandings of advancement and success. Within physics Peter Galison has described the rise of computer/model based approaches to research. Physics has a long tradition of experimentation, with established methodological criteria and means of communicating results. With the introduction of model or computer based simulation, a methodological conundrum arose in physics: was this a form of experimentation? Was this an empirical research, or some form of thought-experiment? Debates as to the status of knowledge produced by 'in-silico' means (Atkins 2003) ensued, and the future careers of hybrid graduate students was held in question. It was decades later that a place for models and simulations came be established within

[7]Years later, the UGLS was abandoned as a GEON project – partially due to funding constraints, but also in the face of realizing the enormity of the task.

physics with its corresponding criteria for communication, journals for publication and undergraduate curricula. As computing has entered the earth sciences similar sorts of debates have played out, with various specificities. For example, geo-physics, with its large quantitative datasets and ties to other branches of physics has had the benefit of learning from the experience of the broader field of physics, while paleobotany and its stronger historical and qualitative traditions is only beginning to face these questions.

In GEON's kick-off and all-hands meetings, information scientists began to understand the general orientation of the geosciences towards IT. While for information technologists, an ontology has intrinsic value – its very development is a success – in the geosciences IT projects would not be considered 'science contributions' to geology. Put another way, a built-ontology would not be considered a success, rather, only geologic knowledge derived from an ontologies' existence would constitute the success of GEON[8]. GEON's ties to its funding body, the NSF, is both through the computer science (CISE) and the geoscience (GEO/EAR) directorates, and thus it must negotiate these multiple commitments in the distribution of its development efforts; GEON must produce both IT advances and knowledge about the earth. In this first step in the routine geoscientists began to *know of* the techniques and technologies which would be employed for

[8] We have already seen that what does and does not count as a scientific contribution is often invoked by scientists themselves, sometimes within a general consensus, but often as a debate over the boundaries of the category itself: 'what counts as new knowledge in geology.' This is particularly the case with IT-science collaborations where IT developments are not considered advances – 'databases are not science'. From a historical perspective, however, it is clear that what counts as a science result is constantly shifting within any given science. For example Galison (ibid) in his studies of computer in experimental physics; Winkler has shown the rise and fall of images as knowledge within early-modern astronomical communities (Winkler and Van Helden 1992); and Epstein (Epstein 1996) has shown the reconfiguration of standards for medical clinical trials in the face of AIDS activism.

producing interoperability, while IT experts began to understand the data politics of the domain to be represented.

In enrolment there is also the creation of the particular division of labour which characterizes relations between domain and IT in knowledge representation: a relationship in which KR experts attempt to establish distance from the *content* of scientific knowledge. We will see that re-creating this division is a continuous effort, which was initially enacted primarily by IT practitioners but as ontology work continued in GEON was progressively taken-up by domain geo-scientists. Below is an excerpt from a class taught on ontology development, intended primarily for computer science students, the example is drawn directly from applications in GEON:

> B: Here we have the scientist's question [reading from the slide]:
>
> "What is the distribution of the ... I don't know ... uranium lead [hesitating] surplus of A-type plutons in Virginia?"
>
> I don't what A-type plutons are. I barely know what plutons are, Ok? You see that these guys use a language that we as non-geologists have trouble understanding, and then they use databases and they want us to help them integrate their data. What can we say? We can say put all the relevant information in the database, but still you have all these different databases. What we have to do is get them to tell us how to connect ... the A-type pluton column in this database to the uranium lead in this other database.

In this extract B is first expressing a relative ignorance regarding the details of geochemistry or geophysics, while also marking-out some familiarity with the domain: B knows, to some extent, the nature of a pluton.

Knowledge acquisition requires a comfort with the language of the domain, but specific details are left to experts. Secondly, he is instructing his students to leave aside the particular content of the scientific knowledge and to focus on the relations between concepts, and specific connections to database schemas – plutons and uranium lead become predicates connected by a query, which in turn must access particular columns in at least two databases. Here, an ontology must lay on the topology of scientific knowledge like a wet blanket, capturing a surface to support what is pragmatically necessary for interoperability or information navigation.

When domain specific IT experts create knowledge representations, e.g. a 'geoinformaticist', this division of labour is not as clearly visible. However, the ontology experts within GEON are not tied to any geoscience, and so there is an effort to capture while remaining distant as to the content of the domain. To put it more bluntly, ontology specialists often are indifferent to the configuration and controversies of domain expert knowledge per se, rather the concern is to create a functional knowledge representation which faithfully reproduces internal conceptual relations with a sufficient granularity to achieve automated computer action.

In practice a clean division between specifying content and creating representation becomes difficult to maintain but what we wish to draw attention to is the effort to inculcate in geo-scientists the sole responsibility of specifying domain knowledge. Since the act of naming objects in the world is invariably political in science (Bensaude-Vincent 1989), particular attention must be paid to the actors' own categories. In the words of a GEON ontology designer:

> One of the most important principles is to utilize terms and methods derived from the way experts communicate in their local, day to day work. In the context of the challenge problems [of building an ontology], if the experts think and refer to the first input parameter as A, then we use the term

> *A* when eliciting its estimates. Likewise, if the initial elicitation demonstrates that experts think of uncertainty as a range or interval of values, then it makes sense to elicit in those terms.

The scientist can only be enrolled in the ontology if she can see her own worldview in it.

This is not immediately apparent at this stage of the routine; in fact, IT participants may even downplay the knowledge involvement required. In the initial pedagogic stages of introduction to ontology we have often observed the use of 'pop examples' such as the as wine classifications, or 'smart' searches for commercial goods. For example, an ontology enabled search for 'beauty' may automatically include 'skin, hair and nails' but exclude 'plastic surgery.'

For IT educators these kinds of pop examples have several advantages: they are light and entertaining and maintain the attention of the audience; they require little expert knowledge and thus can travel across domains with the IT specialists; and because they are not closely tied to the domain are unlikely to foster internal debate amongst those experts present. However, it is precisely these kinds of examples that preclude an early sophisticated understanding of the future difficulties in building detailed knowledge representations. The triviality of the example for domain participants obscures the knowledge stakes involved. This is not to say that there is nothing 'at stake' in wine ontologies, rather that the use of a domain distanced knowledge (e.g. geology from wine) underplays an epistemic significance in pedagogic explanations of ontology. One can be assured that building ontologies for wine involved detailed and specific knowledge capture work, which was taken very seriously by wine aficionados. Proximity to knowledge endeavours charges epistemic issues – just as wine classification is distanced to the average geologist, we can also look within geology to see that pluton classification is epistemically-distanced for the paleobotanist.

The relatively distanced importance of ‘pop examples’ has both pragmatic uses, and unintended consequences. If in describing ontologies detailed domain examples are used, it is quite possible that domain scientists would focus on the technical knowledge, perhaps even leaving aside the ontology education at hand in favour of deconstructing the represented scientific knowledge. Thus by using pop examples such as wine ontologies, the technology becomes foregrounded rather than the knowledge representation. On the other hand, using wine ontologies can portray ontologies as 'depoliticized' hiding the kinds of detailed knowledge articulation that will be necessary to complete the learning cycle of ontology.

In summary, the initial phases of the routine bring forth the problematic of interoperability for the domain and the particular solution offered by ontologies, they also inaugurate a division of labour which keeps domain content separate from encoding practices, and they introduce the IT members to a configuration of data politics within the domain. These initial encounters are presentations *about* ontology, and can only begin to familiarize the domain with the kind of work for building; in these initial discussions I observed very little consideration of the difficulties of actually creating formal knowledge representations, and few discussions of the commitment ontologies imply for a larger domain community. This comes later in the practice of enactment. Rather, in examples such as wine ontologies or smart shopping searches, I observed a displacement of the complex issues for a later time. It is in the practice of ontology building that acquisition is learned as a skill, and the commitment of the individual domain participants to a larger community becomes clearer.

The Practice of Capturing Knowledge

The second step in the routine is a practical pedagogy, learning by doing. Domains carry with them epistemological traditions: ways of knowing, and criteria for what is considered knowledge. The methods of knowledge acquisition must be tailored to the configuration of knowledge in a domain – and in turn the domain must learn to 'speak' their knowledge in a language accessible to machine encoding. While in the received understanding of science the myth of a single scientific method abounded, more recent research in the history, philosophy and sociology of science have uncovered a plethora of domain specific methodologies, trials and vehicles for the establishment of an accepted knowledge, and great shifts over time in these methods and criteria. For example, historian of geology Martin Rudwick has traced the evolution of visual languages of evidentiary production and dissemination within geology. He argues that this language of visual coding – a means for representing topographical, distributional, structural or even causal features – emerged in tandem with an increasing knowledge base of the field itself:

> During the period in which 'geology' emerged as a self-conscious new discipline with clearly defined intellectual goals and well established institutional forms, there was thus a comparable emergence of what I shall cal a visual language for the science, which is reflected not only in a broadening range of kinds of illustration but also in a great increase in their sheer quantity (Rudwick 1976).

The visual tradition has remained strong in geology to this day, and in GEON ontologies and interoperability are closely linked to 'needs' for visualization and GIS mapping. In developing ontologies an abstracted notion of need is easy to identify – “ontologies for interoperability, for knowledge discovery or to facilitate visualization” – but it is in the practice of ontology building that domain scientists come to understand the involved complexities of specifying their knowledge with specific tasks in mind, identifying specific technology requirements, and means to achieve their goals.

In the first step of the routine, enrolment, the curiosity of scientists in GEON was piqued. Over two meetings geoscientists had established a more detailed understanding of the problem that ontologies offer to solve and a nominal understanding of a division of labor between geoscientist knowledge specification and IT encoding. Next, scientists began turning to their IT collaborators and asking to build an ontology. But at this point a significant gap was revealed between abstract understandings of ontology and a know-how for proceeding. In the following quotation by an ontology expert, many of the initial encounters with practice are summarized:

> different scientists come to me asking "we want to have a workshop to define ontologies." That’s very good, we're very happy to host that, to do that, and help them with that. But the issue is to do what? What kind of ontology do you want? What do you need it for? But sometimes it's actually useful to conserve that [ambiguity], to get people together from the domain. We've had people here from geo-chemistry, people here from seismology, so within that group, lets say seismologists, scientific representative persons from a domain, they start all of a sudden arguing heavily about the things they do, the way they view the world. But if you put them into this exercise of trying to find ontologies, of what are the things they care about, what is important for them, what are your analytic methods, how do all these things work together, and how can you create more... uh... how can you share knowledge, how can you work together in some sense. Ontologies can be that catalyst, or they can create a lot of tension, you know...

This outlines many of the principles that guide action within GEON. Ontologies, within computer science, are of course itself an active field of research with internal divergences in theoretical and methodological approach. For example, whether ontologies should be *application specific or independent* remains a general issue of contestation. In other words, is it possible to produce a general ('top-down') ontology which, if properly designed, will allow the representation of any entities? Or is knowledge specific to a domain in such a manner as that it cannot be unproblematically resolved into higher order ontologies (Gruber 1993; Guarino and Giaretta 1995; Guarino and Welty 2000; Smith and Welty 2001; Smith 2003)?

Without necessarily subscribing to one view or another, knowledge representation practitioners within GEON suggest that ontology building should be *driven* by specific scientific application (see also Ch. 4.3 and the two-tier approach). "Application drivers" are believed ensure that the ontology will be useful in specific scientific inquiries (a mandate of GEON, ibid.) but also that the scientists themselves will continue to invest their efforts into the long and often laborious task.

Application specificity ensures that that scientific interests in knowledge production will become tied to the functional completion of the ontology. Returning to the informants quote: the task of building ontologies also begins to inform scientists of their internally diverging knowledge commitments. What appeared to be a shared epistemic umbrella -- "geoscience" or even "geochemistry" -- begins to break down into a finely grained mapping of differences. It is this identification of difference within the knowledge domain, and its transformation into explicit discourse accessible to machine encoding that primarily characterizes the learning by practice of ontology building.

Lastly, from this quote we can see that ontology work in practice encourages the domain to begin ordering their priorities, and specifying their 'needs' beyond abstract notions. At a workshop directed at GEON and geo-ontology building, knowledge representation scholar Deborah McGuiness expressed the specification of needs as an emergent process over time:

> There's no quick answers to it. There's answers, but it's a dialogue. Because we have to figure out what your requirements are [...] even the most articulate person that you can encounter today, and that we'd hope to encounter tomorrow, after even an hour of conversation, or even days of conversation, we actually identify over time that there's all these other requirements. It's only then we are really able to get going [...] There's plenty of starting points for this dialogue, but there's no one best starting point. And there's no way in world you're going to pick up one shrink-wrapped [software] and have it represent your needs. The field is young. Even though there are actually all these starting places, there's tremendous variation in needs.

What tasks an ontology should facilitate, what the current requirements of a community may be, are not immediately accessible for expression, but require time and discussion. Interactive dialogue between IT and domain is necessary for elaboration on available solutions but also to curb excessive technological faith. Misconceptions of the current state of technology can lean both towards under- and over- estimating performance, and requirements are shaped in relation to shared models of technical capacities. Thus we should understand the development of *specific* needs as emergent, dialogic and concurrent to, rather than preceding, ontology practice.

--

The GEON IT team has called their particular method for knowledge acquisition 'concept-space workshops' and more recently have begun calling it 'community-based ontology development'. The distributed nature of GEON requires bringing together experts from across the continent for punctuated bursts of capture, followed by extensive revision via e-mail and video-conferencing. During the concept-space workshops geo-scientists and IT experts co-located for two to three days and sit in conference rooms, ‘hammering out’ formalized representations of geological knowledge:

> Here is the recipe that we've currently applied with some success:
>
> One, you lock up scientists for 2-plus days; add some CS or knowledge representation types to hang around there; then you create concept maps; you refine those, following the meeting, or we turn to local geoscientists [at the SDSC]; then, have other scientists to visit us [at the SDSC], so we can work on these, so we iterate. In this way we go from napkin drawings, which is a very useful start, to concept maps, to sometimes really formal approaches. We need not always go to the formal, we can sometime stop at the concept maps. But in order to go to correct knowledge representations we need this situation between scientists and domain scientists.

This method for knowledge acquisition is in the lineage of expert elicitation (Meyer and Booker 2001) and more generally is based on a social model of expertise (see Gaines 1989 for an elaboration from within the KR community) in which knowledge is considered the shared jurisdiction of a domain community. Knowledge is seen as a *distributed* phenomenon rather than localized in particular individuals or texts. This method of a collective practice of knowledge acquisition -- rather than an individual interaction between formal knowledge (e.g. textbook) and captor – is particularly apt for science, where the content of knowledge itself is often the site of debate, controversy and tension. Concept-space workshops in GEON have become a location for both debate and consensus building discussion. The IT teams hoped that through these discussions geoscientists would either come to an accord on knowledge (and thus be able to stand-in for the community) or IT would be able to represent the uncertainty in their ontologies. Concept space workshops were thus an attempt to ensure that the responsibility of knowledge determination did not force a *de facto* shift onto the shoulders of the IT team.

IT ontologies have their own epistemology: what and how the computer can 'know' is very particular, limited by the availability or development of description logics and the level of formalization in an ontology. Ontology development can range from a tool facilitating the work of a user, to system-accessible formalisms. Formal ontology requires a substantial investment of time, effort, and as we shall see, community enrolment. Furthermore, it may not always be possible to produce knowledge representations at the granularity necessary for system automation, especially within the nebulous edges of research knowledge. For this an alignment between the domain and the language of ontologies is necessary. The second step of the routine includes what I call *epistemological alignment*: a double movement of the domain specialist reflecting on her own knowledge base and beginning to reconceive it terms of a language of logical operators; in turn IT experts must establish a sufficient personal familiarity with the domain knowledge and with participants such that they may offer themselves as a resource for this translation, all this while skirting the difficult line of intervening in the domain knowledge.

In the initial phases of face-to-face ontology building at GEON, the concept-space workshops, we found that a great deal of time was spent educating the domain scientists on criteria for successful ontology building. For example, KR experts have asserted that no one should assume that scientific data are necessarily better defined than scientific theory:

> When it comes to scientific data collection – I mean when they are collecting their data of the world our there – sometimes even they don't know what they're looking at. If it's a new instrument, or if they're measuring something they've never measured before, they might only have a vague notion of what they're up to. And somehow they're supposed to be telling us how to relate this data to other data, or to general categories of knowledge? But they haven't even come up with a category for themselves!

This is not a criticism of science, rather a necessary understanding of what kinds of data to expect: research methodologies will produce *raw* data. This form of indeterminacy is a characteristic of science rather than an exception. If we conceptualize science as a *process for knowledge development* rather than a source of truth, it becomes easier to understand that contemporary developments are under continuous scrutiny, debate and revision; in the words of one ontology expert:

> It's both frustrating and exciting that we have to think about these things [ontologies] changing. After all science is about the movement of ideas, not just unchanging fact. Our tools are going to help science, and so they have to somehow match that mobility, rather than somehow holding it back by being too solid.

Given that the IT expert does not wish to intervene in scientific debates, but also requires some consensus in the domain community to do his work, this kind of controversy, or *contentiousness*, can represent a significant derailment in building ontologies. Once identified and encountered the ontology expert must avail himself to the domain and offer strategies for going on in the work of ontology development. In one example that might seem trivial to those unfamiliar with geology, an extended discussion emerged regarding the definition of the color red. Over several minutes the debate become somewhat heated. Knowing the data politics of domain can mean following such fine grained distinctions as color schemes in map-representation.

There have been various standardized legend color schemes previous to digitalization, and many since – some of these are tied to state entities such as USGS, or to similar bodies outside the US. Thus, in determining the spectral band of a particular shade of red, there is also a running backgrounded discussion of alignment with national state bodies or larger world-wide trends in map representation. Picking one standard for 'red' may exclude another. Exclusion, or mismatched categories, can lead to complications in data integration, such as leaving out an important dataset during an ontology enabled search. While in even the most technically minded geological discussions such details can be set aside when encoding ontologies detailed granularity becomes crucial and momentary agreements can become programmed commitments. Conversely, having never explicitly discussed such details, domain experts can find it challenging to produce consensus at such scales of granularity.

In encountering controversies, the inability of domain scientists to agree on a particular semantic definition or relations between definitions, I have observed several solutions employed during the workshops or suggested by knowledge representation experts:

i- *decrease granularity*: deal with the issue at a higher conceptual level where the domain has established a stronger consensus;

ii- *pragmatics*: encourage the domain experts to form a working consensus in order to continue the process;

iii- *rain check*: leave the problem aside for a later time, experts may be able to resolve the issue with a review of evidence, referring to the literature, consultation with experts or, in the long term, production of new evidence; or

iv- *represent the uncertainty*: technical solutions within ontologies work permit encoding multiple knowledges, disagreement, uncertainties, ambiguities or ambivalences [9]

The techniques for resolving difference are a skill-set. Both the identification of differences, understanding their implications, and knowing what kinds of resolutions are possible is learned by practice. Sub-communities in a discipline may hold diverging beliefs about a particular phenomenon, or commitments to particular domain institutions, and previous to exposure through the explicit formalization of ontologies these divergent beliefs or commitments may be held unproblematically. An initial conceptual understanding by the domain specialist of what is an ontology is indeed crucial, but it is never sufficient. Domain specialists have learned a specific way of knowing and have particular definitions of what constitutes valid knowledge in a field; ontologies require particular configurations of knowledge representation which are rarely readily compatible with the current structure of domain knowledges. But my real concern in this chapter is not to show that we all know differently, more practically to ask how different ways of knowing are getting represented on the ground. In epistemological alignment the IT

[9] In practice we have never seen the application of this technological solution, rather we have only heard it referred to.

specialists gather a feel for the domain and its native language, while the domain scientist begins to transform into discourse her knowledge into a machine computable format. This routine is achieved through the practice of ontology building – the communicative work between IT and domain – rather than through reading ontology guides or other formalizations which a domain specialists often find alien, impenetrable and irrelevant (see Orr 1996 for a discussion of the ecology between personal communication and formal flow charts in technical repair).

Within scientific communities it is often the concepts themselves that are at stake in debates, and I have observed patient IT specialist wading through these internal controversies while suggesting resolutions for domain scientists. It is important for the IT specialist to understand her involvement in this social dynamic, and to learn to facilitate the usual difficulties which arise -- domains experts are, by their very position, less experienced or interested in the general problems of ontology building, while for the IT specialist it is potentially a life-long career. This fine grained activity of knowledge acquisition is coupled with a larger procedural development for how to go about ontology building.

Even for IT specialists the initial framing of the ontology workshop method was supplemented over time with an evolution of the various organizational commitments which have to be secured from the domain scientists. Because transforming knowledge into accessible discourse is a learned skill, and requires extended discussion amongst the domain scientists, single ontology workshops are only the first step in building these. Initial meetings within GEON resulted in “napkin drawings” which required substantial discussion, revising, and formalization before they could be 'handed-off' to knowledge representation experts:

> D: So you guys did some recordings and some notes that you are going to pass to us?
>
> J: Ah yes, so once this meeting is done, we are going to set up some action items, and people will attempt to send the information and send it around to make sure that the intent is right, and then get it back to the GEON folks, so that we can put on the web, and then ultimately we need to build a formal ontology. But I think that we have some clear hierarchical relationships defined.
>
> D: one thing, J, works the best if you could hand-deliver that report.
>
> A: Hand deliver?
>
> D: i.e. spend a couple of days with us, and not just [give us] an ontology report, that would be the ultimate.
>
> A: oh yes, what we were just thinking, is that the next step is to start some formal ontologies. And we've already talked to participants -- that actually to formalize it we need to be there, so that we can start talking with the computer people, to make sure that what we're proposing is actually going to work.
>
> D: and that’s what I'm saying, it’s not going to work by email. When you build these things you have to bite the bullet at some point and come down [to the SDSC]

Having already educated various groups of domain scientists at the San Diego Supercomputer Center the GEON IT team was familiar and had already developed a procedure for the initial introduction of ontologies: presentations, PowerPoint slides, open discussion and demos. Similarly the framework for ontology workshops preceded GEON formal inception at the 2002 kick-off. But it was only later that the procedural organization for sketching ‘napkin drawing’s and then revising, was collaboratively determined between the geo and IT participants: can initial ontology meetings be held between geoscientists without IT experts? Can revision meetings be held over the phone or video-conference? Is it necessary to physically co-locate at the SDSC to capture more formal ontologies? Over time the method of building ontologies in GEON has grown from 'ontology workshops' to an entire repertoire of resources and procedures for organizing work over time and national distances. The particular procedures developed are specific to the kinds of resource available for conducting co-located and distanced work: email, and videoconferencing and groupware facilitates distanced work, in GEON they found that only co-located offered the 'bandwidth' to substantively create ontologies.

--

In summary, the second step of the routine – learning acquisition and the practice of ontology building – includes the practical skill of bringing knowledge into discourse, of learning techniques to establish the agreement necessary for representation, and the development of procedures for following-through on the development of ontologies. The practice of ontologies itself involves many of the modalities of knowledge that I have described in this section. I have characterized the ability to transform technical domain knowledges into a discursive form accessible to knowledge acquisition as a skill which can only be learned in doing. Even though many of the geoscience participants in GEON have dedicated their lives to education, university pedagogy has a distinctly different character than the requirements of a knowledge to be formalized for system accessibility. Similarly while geo-scientists are adept at discussing the state of the field, to disagree on findings amongst themselves, and to bring new evidence to bear on current understandings, KR requires qualitatively different expression of knowledge. This too is a learned skill. My future research will focus with finer granularity on the activity of negotiating knowledge for representation. Finally, the involvement and length of time necessary for building ontologies requires procedures for initiation, follow-up, and revision; in GEON's case organization for distributed work was also necessary.

Engaging the Community

The final step in the routine involves engaging a broader community in the use of ontology. In a project such as GEON, ontologies are constructed not only to serve the research of participating scientists but as a community resource, an infrastructure. There is always the possibility that within a community of science – where knowledge is seminal and research may be hotly contested – that an ontology may receive dissent from the group as a whole or from sub-disciplinary groups. Because expert communities are considered the arbiters of domain truths, the front stage work of formalizing knowledge, forming temporary pragmatic consensuses, or representing uncertainties must be coupled with the backstage work of securing consent, building alliances, and holding standards in place within a community.

Perhaps more importantly for the actors, over the development period they come to realize that their knowledge of ontology itself is esoteric. The problematic of interoperability, the purposes of integration, or the future benefits of multidisciplinary data sharing are understood only within the group of domain practitioners who have traversed the routine. To those outside this narrow sphere ontology remains not only esoteric but, worse, 'not geoscience.' Engaging the community returned GEON practitioners full circle to the first step in the routine, learning ontology. Here the tools and methods initially deployed by IT, such as slide-sets, articles such as 'Ontology 101', or pop examples come to be used by geoscientists in processes of outreach to their domain colleagues. By traversing the routine and in the agonistic work of formalizing knowledge, geoscientists come to see that ontologies will require community support. Ontologies themselves are not useful without a considerable investment on the part of the community. As we have seen in section one of this chapter, ontologies are intended to provide automated semantic links between heterogeneous resources. However, if the community does not contribute its resources (data, visualization tools, new concepts to update or granularize the ontology) they will have only an abstract knowledge representation without any connections to automate. The result could be a complex method for archiving and accessing knowledge but with no content.

As development continues the practice of building ontologies and securing community consent progressively begin to overlap. In forming a working consensus on debated geo-scientific knowledge GEON participants had to think not only of the scientists present, but also of a larger community of future geo-scientific users of the ontology. The properties of knowledge we listed in the first section of this chapter above are precisely what makes the movement of ontologies back to a community such a potentially explosive undertaking. The possibility of a contentious ontology should perhaps itself be sufficient motivation to take the development process extremely seriously; of course, the most likely outcome of a contentious ontology is its being thoroughly ignored by a community and thus becoming high-investment 'vapourware'. In this sense the domain community becomes seminal in defining the success of an ontology endeavour.

In an analysis of knowledge representation we must be cautious not to hypostatize *a community*, for in many senses this entity is both the object and outcome of the emerging methods for engaging community. In building information systems for GEON the term ‘community’ has had at least two general uses: i) community has referred to an already existing but amorphous body of the domain, but; ii) community has also acted as the identification of a future body of users who will be linked by interdisciplinary ties and the computing resources of GEON. The two meanings of community are used somewhat interchangeably, and reflect the mandates and purposes of the GEON project. As 'cyberinfrastructure for the geosciences' GEON is simultaneously serving a constituency of U.S. earth researchers, but is also a community building endeavour that seeks to bring these scientists together. The category 'geo-scientists' includes thousands of academic and industrial researchers distributed in universities and research institutes across the US, in disciplines as varied as seismology to archaeology. It is difficult to reconcile this use of ‘community’ with traditional understandings where community is characterized “as a network of social relations marked by mutuality and emotional bonds” in which a “sense of self and community may be hard to distinguish” and where “a community is an end in itself,” (Bender 1978 p.8). Geo-scientists only rarely know each other directly, most often do not interact at all, and in fact are usually only tied by loosely framed interest or by common organizational or funding umbrellas *(e.g. USGS or NSF).* In this first usage of the term community, GEON is referring to its mandated constituency – the broader geoscience collective as a beneficiary of its computing resources. Meanwhile there is another usage of the term which comes closer to what is usually understood as community building or outreach. We can think of that small selection of geoscientists which come together to build an ontology as a *sample* of the larger geoscience community, the work of community enrolment is partially that of making this sample representative of the larger body. In the practice of ontology building, domain practitioners come to see their work as not only an investment of their own time, but also as a knowledge commitment. In building a computable semantic map of geoscience knowledge, individual domain scientists are standing-in as representatives of a larger epistemic community, for this 'representativeness' to hold the individual geoscientists must ensure that a larger collective will stand behind their knowledge representations.

As a by-product of the 'learning by doing' of ontology work, domain scientists begin to consider methods for building community consent, inviting participation and gathering a user-base. Because ontologies remain a relatively novel technology, the act of informing a community brings GEON participants back almost full circle to the first step in the routine; the broader geoscience community must be educated about a novel technology and of the epistemic commitments made in their name:

> This is the opportunity to reach out to the geological community as a whole, whether or not they may have a view, even if they somehow, even randomly, bump into this, and say 'oh this is interesting, I didn't know this is what was going on' -- and this all comes back to accessing What's going on in as easy a way as possible, and not, say, the only way to get to these [ontologies] is go to the GEON portal, and then sign on, to do this and then press this button. And that's the challenge, just how to lay this stuff on, this is what is going on, that this is under development. (GEO PI)

This selection by a geoscientist speaks to the dual difficultly of simultaneously communicating the existence of ontology technologies and making them accessible for community participation. Ontology development is a going concern in the worlds of business, science, government (Fountain 2001) and in the public sphere (Berners-Lee, Handler et al. 2001). Over time the stakes will become clearer as the technology itself becomes more familiar, but in these early stages of science-application GEON is encountering a generally uninformed constituency and few methods for making their work accessible.

A period of ' domestication' is not uncommon following the introduction of new technology. For example, common sense would tell us that following the introduction of the printing press in the 15th Century that we would see a decrease in the number of copy errors over the previous tradition of hand copied manuscripts. After all, the printing press reproduces mechanically, while hand written texts are open to human error. However, a close inspection of early printed texts shows a distinct increase in copy-error. Historian of technology Elizabeth Eisenstein (1983) argues that over centuries of copying manuscripts by hand, a tradition and repertoire of procedure had been derived for ensuring accuracy. The introduction of the printing press, with its shift in locations of production and individuals participating, dislocated many of these correction practices, and it was only over time that new methods could emerge for the new technological order.

Today ontology visualization remains in the early stages, and conventions of representation have not yet become established. As the informant above notes, logging-in to the GEON portal, navigating to the appropriate ontology representation and taking the time to understand it, are significant barriers to participation. Enrolling of participants and engaging of community can be distinguished by the level of possible engagement – participants will experience learning by doing, the practice of ontology work, thus coming to understand first hand the kinds of difficulties of specifying knowledge as code – meanwhile engaging the community, for the moment, remains action at a distance. While a plethora of software solutions are under development for communicating ontologies, none have completed a full development arc and become accessible and familiar to non-expert users.

Over time, domain scientists and the IT team in GEON have begun to put together alternate methods of outreach, based on previous experience or by analogy:

> sometimes, I mean, I don't have that experience, some people who are in the standardization of programming languages, proposed changes for Fortran. You have some idea, you propose it in the [list serve] forum, everyone jumps on it, chews it up, you respond. Then eventually it comes up, sometimes exactly as it was proposed, but now everyone is saying ' oh you should have done that at the very beginning now it's right, now its good'. This [ontology] can be something that geologists debate, contribute, and then they feel better, and they may feel more comfortable accepting those ontologies, if it's not something that we usurpers [...] are dumping onto them. But if it's something that they are discussing themselves, it can only help.

While ontologies are often presented as an alternative to standardization, experienced practitioners quickly began to see procedural similarities. In standard building, top-down approaches can be softened by establishing multiple mechanisms of community involvement, participation, and feedback. In the selection above, changing code for Fortran is compared to building ontologies: while a joke is being made that even following a period of initial dissent by the community the final code may resemble the initial code, the informant is expressing a need to produce forums for discussion rather than having an ontology appear *deus ex machina*. This is hardly a sophisticated strategy for intervention; however it is only over time that detailed methods can emerge. In the same day as the previous quote another geoscientist proposed to widen the scope of concept-space workshop for ontology development:

> so there is this specific action item, which is, once you prepare the report, when you want to come down and do the formal representation? Or rather I should say the specific follow up. But [individual KR specialist] is going to be here tonight... and I was talking to him and one thing we want to do is, also for all the ontology work going on so far, I think we need a system for follow up. Because ontologies are not just one time things, obviously, so as a group how do we follow up? Is it subsequent meetings, *clearly these can be exposed on the website, but then we need to get this real engagement going, we might even bring in some representatives from the community. Like I said, put it up there, let the community look at it.* We need some process. (Geo PI, emphasis added).

It was following this point that the 'concept-space workshops' became the 'community-based ontology development.' This is not a purely a rhetorical move, rather I point to a processual development over time in which representing the community becomes significant to the ontology building activity. Later in the same year the first geo-ontology workshop was held as the SDSC, backed by GEON, but directed at a much broader collection of computer and geo -scientists than previous such efforts.

Apart from community outreach another method of GEON's ontology community building efforts could be characterized as an institutionalization of the procedures of ontology, and the formalization of relationships. In the larger GEON project an alignment with already existing institutions is an ongoing effort: maintaining ties with the primary earth science research institutions, leading publications, ‘sister’ cyberinfrastructure projects and other geoinformatic endeavours. Similarly within ontology work a parallel alignment must be maintained: following the leading trends in the stabilization of ontology languages, technique and applications; but also with ongoing KR projects in the earth sciences. For example GEON's ontology efforts have always drawn on already established work within the Canadian Geological Survey or the British Rock Classification, making these available as options within the configuration of its search or data-registration engines. While the geoscientists of GEON have already led efforts to publish special editions of journals and books on issues such as geo-informatics, more recently they have begun looking at a geo-ontology edition or book publication. Finally, geoscientists have considered the creation of geo-ontology 'facilities,' which would mirror the already existing data repositories and standard bodies in the field, and thus provide points of contact for the community, means for arbitrating the construction of new ontologies, the registration of data and perhaps the possibility of upper-level ontologies. This latest trajectory of action – formalizing relations through institutionalization or publication – remains novel within GEON. Few practical strategies have yet to emerge; however it is clear that these formalizations will involve building institutions beyond the initial scope of the GEON project, and will require stronger alliances with the established regulatory bodies of geoscience.

This section cannot serve as a 'how-to guide' for community enrolment, or make any claims to providing a comprehensive list of possible methods. We are in the midst of a domestication of the technologies of interoperability, ontology one amongst those. The activities described in this section will be a continuing site for sociological investigation. Today the methods and techniques for identifying and acting on a domain community are emerging, mirroring the technical evolution of ontology itself. Rather, this section points to the kind of learning which has characterized ontology development. Beyond constructing an ontology, but also beyond the first two facets of the routine -- understanding 'what is an ontology' and 'how to ontology' -- both IT and domain participants must enact (Fountain 2001) the implementation of the technology.

--

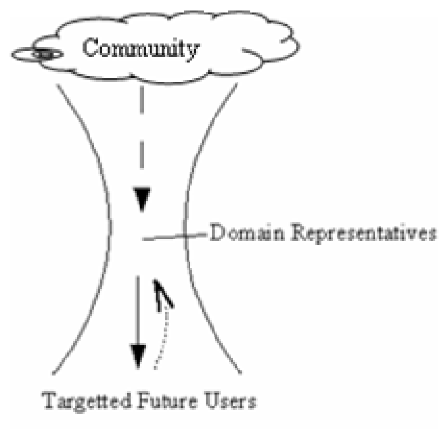

Figure 19: Knowing community: initially the domain representatives stand-in for the community and speak for the interests of future users. As these representatives 'go forth' into the community feedback is incorporated into the knowledge representation, and the community comes to be known via specific networks of communicaiton

The entire arc of ontology development can be represented as an hourglass shaped process (fig.19). Beginning as a broad amorphous community – e.g. 'geoscientists' – domain experts are selected to represent this community as a whole, or task-specific portions thereof; at the tight neck of the hourglass this small selection of individuals must come to states of agreement and gain the discursive capacity necessary for knowledge representation, and; then, in turn, the domain scientists must broaden the links built by ontologies back to the community. Enacting the specific methods for intervening in community – e.g. open workshops, forums, publications, standard bodies &c. -- leads to a specification of that community as particular groups are targeted. Engaging the community is a movement from a broadly defined domain to an identified body of future ontology users (Woolgar 1991; Mackay, Carne et al. 2000). In identification the meaning of community changes, or doubles, referring both to the amorphous entity and to particular relations with groups. These specific targeted members of the domain may act on the ontology development itself; depending on the particular mechanism of community enrolment they may become participants themselves.

In summary, the third step of the routine – engaging community – includes a growing awareness of the commitment to a larger knowledge community, of developing particular techniques for the identification of a relevant group of future users, and the development of particular techniques for engaging with that user community. In the acts of engagement, such as workshops, an amorphous domain community is transformed as specific relationships are formed which link people and organizations through ontology building or use. Technologies, and in particular novel technologies, do not simply diffuse into usage, nor does the illocutionary force of a 'best ' or 'most efficient' method lead from invention to innovation. Innovation – which I define as the progressive domestication of a technology and the work of its uptake in specific implementations – of ontology has required the simultaneous invention of technique for education, dissemination and application.

Conclusion

At the writing of this chapter GEON has been an active project for almost three years. The various phases of the routine have been slow to develop; for example, a sophisticated dialogue of investment and the methods for engaging the community have only emerged in the third year of data collection. It has taken substantial time to come to collective understanding within the project of procedures and practicalities for problem definition, knowledge acquisition, and community building (table 3).

Table 3: Summary of the elements in the routine for knowledge capture.

Learning Ontology	***The Practice of Ontology***	***Engaging Community***
• the problematic of interoperability (domain/domain difference) and the solution of ontologies (mediation) • introduction to the data politics of the domain • division of labor along IT/domain boundary– IT encoding and domain content	• how to bring forth knowledge into discourse • resolving difference (granularity, pragmatics, rain-check, represent uncertainty) • procedural development	• the commitment of ontology ('hourglass structure') • identifying community • develop methods of outreach or community building

As I have demonstrated through the development of the concept of the routine, this fitting process is integrally organizational, practical and conceptual and involves bringing in the communities and institutions of science. Formal computer-based ontologies are only as good as the knowledge that they encapsulate. In GEON, actors themselves have shaped the routine around a view of knowledge capture such that any attempt to capture knowledge recognizes the nature of this fitting. Knowledge, community and institutions must come together to authorize an ontology, through formal channels of consent (such as becoming the 'official' ontology of a particular institution), and 'informally' through up-take by users.

Many 'social' studies of knowledge representation suggest that the direct participation of sociologists in acquisition could facilitate the process (Bowker, Star et al. 1997). This is not my conclusion. Rather, it is that knowledge representation endeavors are always already a kind of *sociological work* (Latour 1996) and as we have seen are deeply engaged with theory and methods from the social sciences and humanities. Such theory and methods have become incorporated into the routines for ontology construction in very practical ways. Knowledge can be tacit and contested and so methods for elicitation and for representing debate and uncertainty are necessary. Knowledge is the jurisdiction of a community and so outreach activities are required. Knowledge is tied to institutional actors and their organization and so these must be enrolled in any ontology effort.

Information technologists have always had as part of their output the goal of producing tools – e.g. computing power, visualization, data management and storage. In this sense ontologies are no different: they are a tool for the user to execute sophisticated searches and queries. But building ontologies have been frames as requiring something that few applications have required before: the programmer must have a relative understanding of the knowledge to be represented and they require the domain experts to communicate their knowledge in a computable format. Ontology builders cannot sit on their neat and tidy side of the 'design wall' and throw over a finished product to the user (see Ch. 4.3 on characterization in the two-tier approach); an intense pedagogic relationship is required between IT and domain. The epistemic commitments of ontologies, and their purpose as community resource, means not only a close knowledge-based collaboration between domain and IT, but also negotiating a relationship with a broader domain community.

--

Interdisciplinary work is defined by actors as investments in translating between domain languages, practices and organization, as well as data, computing resources and tools. With ontologies semantic relations and equivalencies become formal and executable, and as infrastructure are intended to serve as long term investment. Anthropologist Mary Douglas notes: “Nothing else but institutions can define sameness. Similarity is an institution,” (Douglas 1986, p.55). The front stage work of local agreements must be coupled with backstage work of securing a larger consent from a community of practice and its host institutions. Sociologist Bruno Latour reminds us that there are two simultaneous meanings for translation, and that we should maintain both in mind in any such act (Latour 1987):translation, in its more linguistic sense can refer to the production of an equivalence, 'sameness' – equal meanings across languages for example. In its physical sense translation refers to a movement, to a displacement. These understandings of translation are particularly apt for building ontologies, as a translation of a form of knowledge from *wild* to formal.

It is more informative to observe the production and incorporation of an ontology into a community of practice than to ask ‘what is an ontology?’ Diana Forsythe in her studies of computer scientists constructing expert systems has noted that claims about artificial intelligence can quickly evoke high spirits (Forsythe 2001). Similarly in GEON and more generally in the field of knowledge representation, claims about ontologies often reach prophetic levels of revelation and cataclysm alike. Following Forsythe, I have elide discussions of post/humanism, Singularity, or the feasibility of the Semantic Web (Berners-Lee, Handler et al. 2001) in favour of empirically addressing the work of building ontologies.

However, as infrastructure and community resources ontologies cannot be just the words of sources but must be made the knowledge of the community. This is the basis for a boundary work of epistemic granularity between information technology and its (geo-scientific) content. Information technologists will regularly raise both hands and show empty sleeves, demonstrating no interest in intervening in the knowledge affairs of the geoscience domain they are representing. It is the form rather than the content which concerns them; the content is a matter for scientists. Returning to an excerpt from a knowledge engineer above:

> I don't what A-type plutons are. I barely know what plutons are, Ok? […]. What we have to do is get them to tell us how to connect ... the A-type pluton column in this database to the uranium lead in this other database (IT PI)

In this extract the speaker is first expressing a relative ignorance regarding the details of geochemistry or geophysics, while also marking-out some familiarity with the domain. This knowledge representation expert knows, to some extent, the nature of a pluton.

Knowledge acquisition requires a comfort with the language of the domain, but specific details are left to experts. Secondly, he is instructing his students to leave aside the particular content of the scientific knowledge and to focus on the relations between concepts, and specific connections to database schemas – plutons and uranium lead become predicates connected by a query, which in turn must access particular columns in at least two databases. Here an ontology must lay on the topography of scientific knowledge like a wet cast, setting on the surface to support what is pragmatically necessary for interoperability or information navigation. It is by no means *necessary* for ontologies to make 'reasonable' or common sense linkages, and for an ontology builder no more evidence is required than assurance from domain representatives.

The majority of information technologists within GEON are not tied to any geoscience, and so there is an effort to capture while remaining distant as to the content of the domain. To put it more bluntly, ontology specialists often are indifferent to the configuration and controversies of domain expert knowledge per se, rather the concern is to create a functional knowledge representation which faithfully reproduces internal conceptual relations with a sufficient granularity to effect automated computer action**.** The work of maintaining a clean division between specifying content and creating representation is a continuous achievement in talk and practice (Garfinkel 1967; Boden 1994), which is coupled with an effort to inculcate in geo-scientists the sole responsibility of specifying domain knowledge.

--

In what has become one of the most important statements from Science and Technology Studies (STS) sociologists Steve Shapin and historian Simon Schaffer noted that "solutions to the problem of knowledge are solutions to the problem of social order," (1985:332). With the technologies of knowledge representation this felicitous phrase takes on new meaning, most notably if we decompose Shapin's abstracted 'social' into the practical work of ontology building in GEON. This includes i) the transformation of scientific practices of data collection and mark-up; ii) the production of organizational arrangements to support technology, data curation and access; and iii) an institutionalized imaginary (Verran 2001) of interoperable infrastructure and heterogeneous but collaborating disciplinary communities.

Here is the great brilliance, the consequential transformation and the enormous difficulty in the logic of interoperability. Interoperability can only be achieved to the extent that the specificity of a science can be transformed into the generality, the universality, a single standard, or a computable translation, of system accessible machine readable code. Interoperability only ‘works’ to the extent that it meets both the criteria of being specific enough for disciplinary standards of evaluation and routines of practice, while also general enough to move across machines, software, time, disciplines, institutions, or representational technique.

– Conclusion –

–

[T]he question is why someone would engage in the strange contradictory task of severing what he wants to glue together.

– Bruno Latour

A map of the world that does not include Utopia is not worth even glancing at, for it leaves out the one country at which Humanity is always landing.

– Oscar Wilde

–

Interoperability and its Others

In their studies of the standardization of cardiopulmonary resuscitation (CPR), Berg and Timmermans have identified the simultaneous framing of ‘order and its others’ (Berg and Timmermans 2000). They pose disorder as an ‘acquired’ characteristic, and point to the work of actors in making it so. Order and disorder are each other’s ‘doublets’ (Foucault 1970). The manifestation of disorder is revealed relative to an emerging form of order.

So for example, Berg and Timmermans study the means by which the American Heart Association has implemented a standardized practice to enhance the effectiveness of resuscitation efforts. As an effort to rationalize medical work the execution of CPR is made homogenous. Local variations in execution are problematized and eradicated relative to a scientifically informed manner of proceeding. A protocol is developed, and:

> ideally, an eighteen-year old member of the Coast Guard […] who encounters a drowning, and a nurse with twenty-five years of clinical experience who notices the vanishing shallow breathing of a newborn in the neonatal intensive care unit […] will react in a similarly ordered way (Berg and Timmermans 2000:39-40).

A regime of knowledge production is established around the development of this standard protocol. Clinical trials investigate the most effective manner to proceed in CPR, replacing convention with knowledge. Decision science establishes possible contingencies and incorporates these in the protocol as well, finding sources of choice and rationalizing them. A regime of training is established around the standardization of actual – on the beach or in the hospital – practice, extending a nodal network of standardized protocol to sites of action. "[O]rders do not emerge out of (and thereby replace) a preexisting disorder. Rather, with the production of an order, a corresponding disorder comes into being" (Berg and Timmermans 2000:30). The formulation of an order is tied to the means by which a disorder is framed and known; in turn the enactment of order travels upon the characterization of its corresponding disorder. Orders 'find' and then consume disorders.

--

The *logic of interoperability* functions on a similar set of principles, with differences of note. In this dissertation I have traced the enactment of this logic across the scales of action. The three rudiments of this logic are i- a boundary work of disciplinary difference, ii- a problematization of the consequences of those boundaries and iii- posing a resolution through mediation. This formulation is related to but distinct from that of standardization. As with standards the goal is the production of a universality: all scientists will be able to access computing resources and share their data regardless of location, disciplinary affiliation or technical platform. However unlike standardization the means to achieve universality is not premised on the eradication of found local differences. Instead the logic of interoperability offers the production of a *technical mediation*.

The workings of orders and their others are specific. There is a particularity to the logic of interoperability. Similar to Berg and Timmermans, an order of interoperable computing resources and multidisciplinary collaborations is formulated as the divided sciences and its disordered data are 'found'. However unlike with CPR, in interoperability a solution is not formulated around the homogenization or standardization of practice. *The solution is mediation of difference.*

For CPR, Berg and Timmermans show increasingly granular efforts at achieving order as resources are brought to bear in the eradication of local differences. Regimes of knowledge produce certainty of protocol. Homogeneity of action is achieved through the extension of a protocol to practice. However, the logic of interoperability does not envision the eradication of found differences but rather their mediation. The sciences are necessarily specific and different but also, because of this, they are divided.

In the vision of cyberinfrastructure the disciplinary differences of the sciences are necessary to research. Metamorphic petrologists have their field research and small datasets while geophysics have their large shared instrumentation and massive databases. These differences are necessary to the 'conduct of the sciences'. Thus these are to be preserved but become interlinked – 'interoperated'. In the logic of interoperability to be disordered is to be disconnected, to be ordered is to continue unchanged in practice but linked through mediation. I have identified this logic along the enactment of the boundaries IT/domain and domain/domain. These are two distinct forms of disorder or, in the language adopted in this text, enacting these boundaries and their crossings produce distinct consequences.

In order to recapitulate this argument I return to the architecture diagram for cyberinfrastructure from which I began in the introduction:

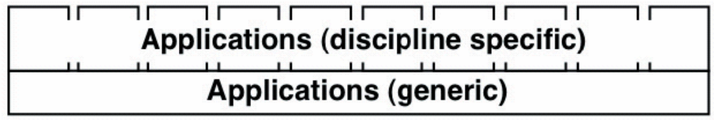

Figure 20: Selection from architecture for CI, from Atkins (2003: 49).

This (fig.20) is the application layer of the architecture; it captures the problematic of interoperability and its solution in the construction of information infrastructure. In the full diagram of the architecture (fig.1; Intro.), above this application layer is the “conduct of science”: the heterogeneous and specific methods, language and concepts of the domain sciences. Below this application layer are the “core technologies and social and technical systems” of cyberinfrastructure: storage, networking, computing resources and the institutions which house these. I have focused on the enactment of ‘an application layer’ which mediates, above, below and across. The application layer itself has the double “responsibility”(Atkins 2003:53) of serving the discipline specific needs of domain scientists *and* providing generic linkages across these applications. The application layer mediates domain differences and informational similarities: above, the differences of domain science remain as the conduct of science; below, the similarities of core technologies homogenously compute information. This is interoperability.

The Relation of Difference

In an outstanding reflexive analysis of her own discipline Marilyn Strathern has critiqued a primary tool of the anthropologist: ‘the relation’ (Strathern 1995). To relate is to establish a form of commerce across heterogeneity while preserving its difference. The anthropologist is always in the position of communicating difference. The archetypal form is *going out*, observing another people and *bringing back* to one's own people: "humanity encountered as Other" (Rabinow 1977:151).

Even as anthropology has turned its gaze back upon its own peoples, to the 'Western World', the endeavour remains the establishment of finer granularities of difference. In these ventures the anthropologist finds variation and communicates it back to us in language that *make differences at once stark and comprehensible*. Whether another culture or our own, 'across organizations' or 'within an organization' a finer granularity of difference is always to be found. Strathern has called this the holographic property of relations: "We could call it a self-similar construct, a figure whose organising power is not affected by scale"(Strathern 1995:18). At any scale of action differences can be found and these can be communicated as relations.

One configuration of this approach in anthropology has been at the level of 'culture,' with a capture of *local knowledges* in *thick descriptions* (Geertz 1973) and the communication of these in (primarily) writing to a (primarily) scholarly audience. However, there have been many others configurations, such as social anthropology with its reduction of peoples to 'primary units' of kinship. Here ownership, inheritance, and marriage rights came to be the scientific short hands for expressing a society (Clifford 1983). By capturing the establishment of kinship relations and how members come to act upon these an anthropologist could bring back a summary view of another social order. This difference, of a society, culture or order, is communicated to 'us' as relations. Kinship relations anywhere can be compared to our own relations; this mode allows for a preservation of difference while bringing the two orders together. This is the focus of Strathern's analysis of 'the relation', the methods by which relation constitutes association without necessarily calling attention to the act of linkage and its consequences.

Strathern's critique is not a raw deconstruction; she does not foreclose on relationality as a tool in the anthropologist's analytic repertoire. Rather it is closer to a reflexive inquiry, a study of the practical epistemic consequences of that tool. My own study of interoperability is similarly intentioned. Not a critique which empties possibilities of action in the field, but an epistemography of interoperability identifying it as a practical reasoning, situated action and emergent consequences.

In an analysis of the logic of interoperability Strathern's *relation* complements Berg and Timmerman's *orders and their others*. While for Berg and Timmermans a disorder comes to be replaced – 'consumed' – by an ordered practice (of CPR), in the logic of interoperability order is a product of the relation of disorders.

--

The base unit in the logic of interoperability is ternary. A boundary is characterized to be crossed. The relation "requires other elements to complete it -- relations between what? This makes its connecting functions complex, for the relation always summons entities other than itself" (Strathern 1995:18). The disorder of differences across disciplines are characterized to generate an *order through relation*. In the diagram below (fig.21) I have placed the mediation of differences in their respective associations as articulated in the architecture of cyberinfrastructure above:

Figure 21: Relations of difference in the logic of interoperability.

Two units are characterized by difference to be linked by the underlying infrastructure: IT/domain difference is permeable, enabling a mediated crossing of domain/domain difference. It is not the content of differences themselves that are problematized but their consequences in communication, collaboration or data exchange.

The goal is not a unity of science at the level of a common method, nor is it a project of reform standardizing the conduct of science: “The relation as a model of complex phenomena […] has the power to bring dissimilar orders of levels of knowledge together while conserving their difference,” (Strathern 1995:18). To become ‘more scientific’ is not a primary organizing category in cyberinfrastructure; such questions are a matter for the domain rather than for its infrastructure. The domains have developed methods, languages, concepts, and technical standards specific to their research enterprises. The conduct of science is specific, and necessarily so. This logic generates an impetus for knowing the domains, to characterize each of their specificities so that they may be interoperated. This will to know the sciences is a technical endeavour with an anthropological soul, “science is inevitably followed by its own anthropology” (Berg and Timmermans 2000:51, paraphrasing Michel Serres, 1987).

Rather than ‘of difference,’ problematization is at the boundaries of disciplinary difference. The consequences of specialization and heterogeneity are difficulties at the level of multidisciplinary communication and collaboration, in the exchange of raw data, and in the collective use of instrumentation and computing resources. This logic is holographic. Problematization applies at all scales and in all differences: cultures, institutions, disciplines, knowledge, concepts, language, conduct, data. To each of these scales cyberinfrastructure offers its mediating solutions: institutions have their 'cross-cutting' ventures; disciplines have their bio- and geo- informatics; conduct has its doubly 'responsible' applications; knowledge, language and concepts have their 'ontologies'; and data have their 'schema mapping' and 'metadata'. While some of these solutions are 'technical', and some are 'practical', both pose general approaches to specific problems – they serve to mediate the abstract issues of the sciences with the specific troubles of disciplines (see Ch.3.2, Mills 1959).

Enacting the Logic of Interoperability – A Summary of this Dissertation

The methodology of this dissertation has not been that of revealing the hidden or underlying 'truth' or 'ideology' of interoperability, or 'finding the social' within its technical activity. My goal has been the *respecification* of interoperability: its explanation as the everyday work – in talk, in practice, in design – of members engaged in its achievement.

The mediation of boundaries, particularly when automated, emphasize the relation rather than the act of association. It has been my goal in this dissertation to bring forth the work of association at three scales of action. We often see 'multidisciplinary projects', 'the needs of the earth science community' and 'interoperable data', rather than the practice of multidisciplinarity, generating a pull to cyberinfrastructure, and the routine of building ontologies. It is the time, energy and cultivation of relations – the methods of association – that I have followed:

> Their power is that […] relations can take any scale, be productive at any order of encounter, whether in a small university department or across the globe. It is a mistake to think they can be measured by size. But they do demand time, energy and cultivation, and that is what is at stake (Strathern 1995:30).

In the logic of interoperability what can be quickly forgotten are the resources and work that go into *interoperating*. In this logic the goal is to make relations proliferate at all scales of action. Data will be made to flow unproblematically across institutions, disciplines and applications. To the extent that this logic is compelling – is convincing or can be shown to be effective – then what is at stake is this distribution of resources.

Cyberinfrastructure is mediation or the "federation of the necessary multidisciplinary, multi-institutional, and geographically dispersed human expertise, archival data, and computational models," (Atkins 2003:18). The keywords of this dissertation are contained within this quote: federation (rather than standardization), across disciplines, institutions, people, knowledge and data. This programmatic statement from the Atkins Report outlines a vision of interoperated sciences. But it is only along with the practical work of participants in projects such as GEON that we can come to understand the full range of activities that are the enactment of a logic of interoperability.

Characterizing the Boundaries of Disciplinary Difference

Characterizing disciplinary difference is a boundary work which articulates extant methods, languages, data, and institutions and frames these relative to others. It is not a social construction. Rather it is a sensemaking practice with an orientation to the identification of one domain's traits via its counter-position with another's. As we saw in the quote above from the Atkins Report the type of traits, the specific nature of differences, varies broadly – institutions, locations, expertises, models – but the primary focus in cyberinfrastructure is data; the diversity of organizations, national and international distribution, disciplinary specialty and epistemic models will be traversed to make the movement of data unproblematic.

This remains the same in enacting both primary boundaries in the logic of interoperability: IT/domain and domain/domain. Both the 'internal' heterogeneity of the domains (e.g. geophysics and paleobotany) and 'external' relations of CS and domain (e.g. knowledge representation and paleobotany) are found to be plagued by linguistic, conceptual and methodological difference.

In GEON both participants and National Science Foundation (NSF) officers have contributed to the characterization of disciplinary difference. GEON is 'twice basic research and infrastructure to boot'. Over the two iterations of funding proposals to Information Technology Research (ITR) GEON was shaped as a collaboration across IT/domain boundaries, contributing to the distinct research trajectories of both computer and earth science. In its proposal GEON outlines a dual responsibility to be information technology research and to learn 'something new about the Rockies'(Ch. 2).

Over the course of various 'Geoinformatics' activities participants articulated the disciplinary differences amongst the earth sciences. These were enacted as endogenous comparisons. For example, one participant at 'Building the Geoinformatics System' drew on comparisons between seismology and paleontology to characterize differences in culture and in data: the culture of paleontology is individualistic and laboratory based while seismology's has fostered a culture of shared institutions, data and instrumentation; the data structures of seismology are simpler relative to those of general geosciences (Ch. 3.2). The differences across domains are not invented or fictional. Rather, disciplines are framed relative to each other in order to draw out differences.

It is not simply disciplines in the abstract which are characterized, but also the practitioners in them. Within GEON organizing across disciplinary difference came to require the articulation of diverging interest for computer and domain scientists. Both seeks to contribute new knowledge to their respective fields, and to reap the rewards within established 'career reward systems'. However the topics, and thus the rewards structures for research, differ across disciplinary boundaries. Writing metadata will not get a graduate student a job in geology, and producing production quality software will not get a professor tenure in computer science.

It is not simply disciplines in the abstract or experts more specifically, but also the knowledge itself of domains which are characterized. At an epistemic granularity the work of building ontologies is a formal specification of disciplinary differences in language, concept and data structure. In the routine for ontology development earth scientists learned to speak disciplinary difference in languages accessible to formalization. Domain difference is mapped in knowledge representations as they are articulated in the description logics of the ontology web language (OWL) or semantic resource description framework (RDF(S)) (Ch. 5).

--

Berg and Timmermans note that in efforts at standardization the discovery of difference is equivalent to disorder – e.g. CPR must come to be executed 'the same' everywhere or risk compromising its effectiveness: "what should be fought here is idiosyncrasy" (Berg and Timmermans 2000:46). But in the logic of interoperability the characterization of difference is not its problematization, per se. Cyberinfrastructure does not seek to erase difference, rather it operates in its interstices forming associations. Actors engaged in cyberinfrastructure make great efforts to demonstrate a preserved autonomy of the domain sciences.

To paraphrase Atkins in his broad plan to revolutionize the sciences 'the conduct of science falls beyond the scope of this report'; or in the laminated phrase in GEON 'we don't make standards,'; or in the efforts of GEON ontologists 'I don't even know what a pluton is!' demonstrating an *indifference* (Ch. 1) to the knowledge debates in the domains they seek to represent.

It is not that the conduct of science will be unchanged – after all this is 'a revolution in science and engineering'. But it is not a top-down revolution. Again, returning to Atkins, "Only domain scientists and engineers can revolutionize their own fields" (2003:50) . In a word, and in words, cyberinfrastructure is not a revolution toward centralization, but toward federation. Problematization is not *of* the diversity of scientific activities but the balkanization of science *because of* the boundaries at the intersections of those differences.

Problematization at the Two Boundaries of Disciplinary Difference

Characterization of difference is coupled to its problematization. It is not domain difference itself which comes to be the problem. Disciplinary specificity is the means by which scientists come to know the world; each domain of the sciences conducts its research respective to a phenomenon, and these methods, languages, institutions and data structures are tailored to those phenomena. Rather, it is the consequences of this specificity, the *relative* *heterogeneity* across boundaries of disciplinary difference, that are problematized. The consequences at the two primary boundaries are formulated differently: at domain/domain the progress of science is slowed; at IT/domain the development of infrastructure is inhibited.

In Ch. 2 I traced the generation of a ‘push’ to cyberinfrastructure. The ITR program was ‘cross-cutting,’ carefully structured to encourage collaborations across the directorates of the NSF, and in particular with Computer and Information Science and Engineering (CISE). I showed how this ‘structure’ was enacted through a *practice of multidisciplinarity*. Multidisciplinary collaborations are problematized as difficult organizational ventures requiring a careful mediation to initiate and maintain. A knowledge regime around multidisciplinarity has come to recommend ‘best practices’ for their organization, such as the establishment of coordinating meetings (i.e. Cummings and Kiesler 2005).

In organizing the program solicitation NSF officers drew on previous experience in arranging multidisciplinary collaborations (e.g. the Knowledge and Distributed Intelligence program – KDI), and organized to facilitate proposal evaluation and awarding. Problematization of work across boundaries has led to the development of multidisciplinarity as a skill, embodied in practice, in the reproduction of organizational routines, and in material arrangements.

In Ch. 3 I traced on the generation of a ‘pull’ to cyberinfrastructure in the earth sciences. CI is one amongst many possible visions of information technology development. GEON is one amongst many possible configurations for large-ITRs. The pull, or ‘need’, for a single informational platform in the earth sciences is an (incomplete) achievement premised on the problematization of disciplinary difference and enacted in the venues of

Geoinformatics. Problematization at the 'Building the Geoinformatics System' workshop included loss of data, the production of 'stove-pipe' solutions and cultural stagnation. Without the development of curatorial practice, and the production of an archive such as those proposed by cyberinfrastructure, data may come to be lost. Without some form of coordination across particular domains, each of these will develop its own in house IT applications that are 'one-off' or 'stove pipe' solutions, serving only a single research group, hindering collaborations and preventing multidisciplinary research. IT projects will 'reinvent the wheel' constructing solutions again and again rather than 'leveraging' innovations across disciplinary boundaries. Domain 'cultures' can be the source of stagnation in the face of novel technologies, and their 'career structures' do not recognize the work of contemporary scientists in creating metadata or ontologies. Problematization operates in the interstices of all scales of domain difference, whether institutions, culture or data structure.

In Ch. 4 we came to see how GEON earth science participants problematized their own range of disciplinary difference. The two-tier approach articulates disciplinary difference as interests, and problematizes these relative to collaboration for the production of infrastructure. Information technologists wish to produce new tools, applications of computer science theory, while geoscientists must contribute to knowledge about the earth, publishing their findings in journals and conference proceedings. The development of infrastructure is hindered because no one is 'committed' to long-term development for domain application. The two-tier approach poses these diverging interests and is also a model for mediating them by dovetailing both: GEON will produce 'twice basic research and infrastructure to boot'. Close collaborations of earth scientists with information technologists leads to, respectively, domain science focused applications with extensible and flexible ('long-term') design.

How should they design a single infrastructure to meet the needs of the range of earth sciences? Earth science PIs came to stand in, as surrogate users, for the *diversity* of the 'geological community'. Speaking in the name of their community they articulated their needs and their envisioned community needs as future GEON users. In a fine grained analysis of the design of the GEON portal I demonstrated in situ organizing as members argued the identity, architecture and computing resources of GEON. In this discussion IT/domain served to organize roles in a casual 'requirements testing' session: computer scientists would present a draft of the website, while domain scientists would respond as surrogates of the earth science community. In turn organizing around the domain/domain boundaries served to articulate an 'umbrella' infrastructure as the GEON portal was crafted to *include* metamorphic petrology, geophysics and geomorphology, as well as education, researchers and principal investigators. Domain difference along both boundaries are mediated in close collaborations.

--

Problematization at the two boundaries of disciplinary difference is relatively similar. Both within domains or across IT/domain it is differences in language, methods, and interests that hinder collaboration. It is in the consequences of their crossings that these two boundaries differ. Crossing domain/domain boundaries produces new domain knowledge: collaborations of earth scientists or sharing of data will lead to 'something new about the Rockies'. However, in the logic of interoperability to communicate across conceptual and linguistic specificity requires a mediation of difference (of location, of data, of language, of concept). This crossing of the IT/domain boundary – collaborations of computer and earth science – leads to the production of infrastructure: computing tools and resources to facilitate long term mediation of disciplinary difference.

Mediating Disciplinary Difference

Problematization is coupled to a proposal for its resolution through a *technologic mediation*. I would like to distinguish this, if only by degree, from the *practical mediation* which *characterizes the enactment* of the logic of interoperability. In building infrastructure mediation is framed as matter of practical collaboration, but the final goal of cyberinfrastructure is a technologic mediation[1].

We have seen boundary work and their crossings at the NSF, as a funding push encouraging work across IT/domain, as a practice of multidisciplinarity, and as twice satisfying the criteria of basic research (Ch. 2). We have seen boundary work and their crossings in the organizing of GEON as the configuration of future users (Ch. 4). These are a substantial investment in time and resources for participants at the NSF and in GEON. In the Atkins Report and CI circles more generally 'building infrastructure' rests at the intersection of efforts by domain, computer and social scientists. It is far more than any naïve technological reduction. But we must also remember that in CI crossings at the boundaries of disciplinary difference are ultimately to be mediated by technical means. We must not forget that for participants this practical mediation is substantially a means to an end: the explicit goal of CI and of GEON is the production of infrastructure. These tools for mediation are

[1] I do not suggest that practical mediation is devoid of technology (Strum and Latour 1987): see Ch. 4.1 for a discussion of the materiality of practical mediation. Neither is technologic mediation purified of human actors (Bowker and Star 1999). The terms point to an emphasis in the goals of participants. The goal of cyberinfrastructure ventures is usually articulated as an automated technical mediation: of distanced communication, of data integration or of knowledge mediation. In practice any such automation will require enactment, operation, maintenance, articulation work and redesign.

what I have called the general technologies of the particular. I have focused on knowledge mediation and the emergence of a routine for building ontologies.

In Ch. 5.1 I outlined the vision of an interoperated earth sciences as articulated in the GEON proposal. In the walk-through example of knowledge mediation an earth scientist is able to find, and act upon, a heterogeneous set of geological databases located in physically distributed institutions. Rather than searching for databases in multiple fields, in various institutions, in various formats, she is able to find data relevant to her research by beginning with a geoscientific concept such as the Wilson Cycle. She does not need to know where the data is located, the Grid will handle the aggregation; she does not need to know the terminology of the domain scientists that collected and organized the data, the ontology will handle the translation of semantic content; she does not need to understand the syntax and logical structure of the data, the schema is mapped to a visualization program. Disciplinary difference is traversed with the aid of an automated technical mediation. It is the "responsibility" of the application layer – the grid, ontologies, and schemas – to mediate the communication of heterogeneous users amongst themselves and with other applications. This is the vision of interoperability.

How actors work to achieve this vision of interoperability is a matter of enactment. In Ch. 5.2 I traced the development of a general routine for the capture of knowledges. The technology of ontology (semantic maps or concept maps) is a general technology for the representation of specific knowledges. The languages of ontology – OWL or RDF(S), to name only two broad categories – are understood to be contentless. The relations between concepts established by domain scientists, 'knowledge', is the content of an ontology. The 'capture' of this knowledge by computer scientists is enacted in 'concept-space workshops' (later ‘community based ontology development’): earth scientists, collected for their ability to represent the geoscience community and speak its knowledge are brought together with computer scientists specializing in knowledge capture to cross the IT/domain boundary. The boundary organizes a division of labor, in which the domain speaks knowledge while CS capture it in the description logics of ontology. If earth scientist cannot come to agreement amongst themselves, across the domain/domain boundaries, higher order description logics will be employed to capture disagreement, uncertainty or fuzziness.

Outside these concept space workshops earth scientists are returning to the community, displaying knowledge representations, securing epistemic consent and institutional authorization, and further extending the logic of interoperability to new constituencies. The ‘front stage’ work of enacting a knowledge representation is also the ‘backstage’ work of gaining institutional accreditation, fostering ‘uptake’ within a community of future users, and a ‘pedagogical action’ of teaching about ontologies and the problematic of interoperability within the earth science community.

The Relation of Difference in the General Technologies of the Particular

At the pole of generality ontology provides a method of linking semantic maps to applications, databases or customized user interfaces. At the pole of particularity ontologies connect to the technical concepts and terminologies of a science. In its search for particulars the work of ontology building and knowledge capture are an exercise in anthropological inquiry. At an ever finer granularity the technicalities of the domain must be known and captured. As with Strathern’s *relation*, ontology seeks to link orders of difference while retaining their specificities. The will to know the sciences is embedded in the logic of interoperability; as with an anthropologic enterprise it seeks a general method for rendering differences both stark and comprehensible, unlike anthropological accounts, in the enactment of interoperability actors are ultimately seeking an automation of translation.

With the kinship relations of social anthropology, the method had the advantage of mediating the general and the specific. The relations of kinship are a general tool, and each society could be described in their 'contentless' language: ownership, inheritance, and marriage rights. Kinship relations could be found in all societies, and brought back as markers of yet another social order: "the model could be enacted over and again in fieldwork," (Strathern 1995).

The methods we have seen enacted in GEON are of a similar order, for example: multidisciplinarity, the two-tier approach and the routine for ontology development. Multidisciplinarity at the NSF came to be a skill, acquired through experience and the design of tools to support work. Officers speak of a 'readiness' at the Foundation to manage larger, more heterogeneous projects, of which GEON is simply one. The practice of multidisciplinarity is a general method for mediating specific disciplinary difference. The 'two-tier approach' is not limited as a strategy for generating earth and computer science *commitment*; it speaks generally about collaborations across the IT/domain boundary. It is a method for mediating the interests of computer scientists and any domain. The technology of ontologies are coupled with a general routine for knowledge capture, equally applied to paleobotany, geochemistry or for that matter, knowledge of schizophrenia and weather prediction. It is a method which extends across the work in GEON for capturing specific earth science knowledges. But the skills, technologies and material arrangements of the routine can, and are, adapted beyond the geosciences. These are the general technologies of the particular.

They are a response to the increasing politicization of standards. We have seen how in Geoinformatics the institutions of earth science have spoken against the centralization of data collections, 'top down' efforts at IT implementation or large-scale data standards. In contrast ontologies are said to be 'community driven,' based on concepts and categories drawn from the field, and supporting existing work practices. For the followers of ontology they represent a 'gateway technology': "permitting the technical interconnection of system components that would otherwise remain isolated" (David and Bunn 1988), individual standards become linked and this is understood to provide and increased flexibility (Egyedi 2001).

Berg and Timmermans remind us that "not only does an order perform its own disorder – it also always contains it" (Berg and Timmermans 2000:36). The technologies of mediation are a response to endogenous criticisms of the centralization of computing and data archiving resources, to standardization, or 'top down' initiatives. By articulating the failings of today's standardization these debates serve to inform more subtle responses, new configurations of practice and technology. In short the logic of interoperability seeks to soften the politics of standardization. Their 'success or failure' remains in the future, and in the daily negotiations to define their criteria for evaluation (see below). My research has focused on actor's efforts to achieve interoperability, to routinize (Jordan and Lynch 1998) methods for relating difference while preserving specificity and to extend these methods across the scales.

A Logic Across the Scales

Throughout this text the Atkins Report has served as a crystallization of the vision of interoperated sciences. It is a significant articulation of the logic of interoperability and the development of a national cyberinfrastructure. A discursive analysis of the Report can help us identify the rhetoric of revolution, a general formulation of interoperability as a problem for the sciences, and a loose plan for developing CI. However, this is not the enactment of CI. To understand how the CI vision has informed a practice of multidisciplinarity, has served as a resource in organizing funds, or has been an impetus in the development of technology we must turn to an analysis at the scales of action. The exposition of this dissertation has been organized around three scales of action which I have dubbed institutional, organizational and technical. At each scale I have followed the enactment of the logic of interoperability; my methodology in doing so has been informed by ethnomethodology and actor-network theory.

I have called this methodology an *ethnography of scale* and, as a shorthand, have described my study as following the work in GEON. However, more accurately, the sites of my research have repeatedly overflowed any conventional definition of GEON's limits. It is because of the multiple scales of action, and the methodology I have deployed to render them analyzable, that I resist the definition of this research as a 'case study.' I argue that GEON cannot be analyzed as a stand-alone project circumscribed by its PIs and their research teams.

The logic of interoperability is enacted at all scales. In the emerging studies of infrastructure, cyberinfrastructure or of particular technologies, a case study is usually limited to the 'inside' of the project: 'how do they collaborate?' 'how do they organize?'

‘what are their methods, best practices or problems?’ However, in order to comprehend how the geosciences have come to need such a thing as cyberinfrastructure we must travel ‘outside GEON’ to the formation of Geoinformatics amongst the institutions of earth science. In order to comprehend how the NSF has come to funnel millions of dollars into large-scale infrastructure development we must look to the enactment of the ITR program. In order to comprehend how ‘ontologies’ are coming to be taken-up in the earth sciences we must look to work extending the logic interoperability to the domain.

In this study I have done away with the ‘insides and outsides’ of GEON. This notion can only be a hindrance in studying the activity of participants that do not acknowledge such limits, and continuously work across them. As actors work across the scales tying together institutional, organizational and technical action so too must an analysis of ‘GEON’ follow across those scales.

I have retained an ethnomethodologically informed analysis at each scale. Ethnomethodology has fostered a particular understanding of the observable (Lynch 1993) rendering analyzable activity regardless of its classification as technical, organizational and institutional. The NSF is not a ‘black box’ institution. It is practical and material organizing in a building in Arlington, Virginia. It is a spokesperson at Geoinformatics 2003 in a building on the University of California San Diego campus (Cicourel 1981). It is an email requesting updates on GEON progress.

GEON, as a distributed national organization, can be known by following the same methods as do GEON participants themselves (Boden 1994). ‘Knowing GEON’ is the everyday work of its participants, and making it act in single voice is an accomplishment in practice. It comes to be an ontological entity in the summary presentations of its PIs to each other, in the slide-show diagrams projected on screens at the SDSC (Taylor and Cooren 1997).

Technical activity is accessible to an ethnographic observer as talk, practice, and material arrangements (Suchman 1987). Competence in specific technical methods is the *unique adequacy requirement* of fields such as earth and computer science (Garfinkel and Wieder 1992). This said, technical activity is not a special case of action. For the ethnomethodological analyst technical action requires a specific treatment, but no more specific than any other form of activity (Lynch 1993). Even the most grandly labelled activities, such as ontology, are in the end conducted by practical reasoning as ordinary, mundane, in-situ activities (Garfinkel, Lynch et al. 1981). Building concept spaces is learned by drawing on materials accessible to all online and using 'pop' examples about wine. Knowledge capture is conducted in a room that often broke down into bickering or escalated into discussions of epistemology. These are all accessible to the trained ethnographer with an ear for talk, and eye for practice and a feel for material arrangements.

However, unlike the canonical ethnomethodologist I have not sought to delve forever deeper into molecular interactions. I have instead stepped from mountaintop to mountaintop as do my actors. Thus, I have retained an actor-network theory informed analysis across the scales. I have traced the methods by which actors negotiate the size of Cyberinfrastructure, Geoinformatics and GEON (Callon and Latour 1981). Members of the communities of the earth sciences do not all individually sit in a room stating their needs. Rather these come to be known through the PIs of GEON who stand-in for the community as surrogate users (Woolgar 1991). Technologies are designed based on models of the geoscience user. But these technologies do not simply 'diffuse' into the community. Rather, these technical networks are extended to particular geoscientists through the ‘outreach efforts’ and ‘community building’ of GEON participants (Latour 1988). Making GEON ‘large’ is the work of its participants, but this forever remains a practical activity, observable as enactment.

Each scale of activity has served as an organizing principle for the exposition of this text; they are not categories informing the analysis of my research. The focus on talk, practice and material arrangement is a *sensibility* fostered within ethnomethodology; the focus on the generation of scale is a *sensibility* fostered in actor-network theory. These methodologies open a terrain of research for the qualitative sociologist. In treating institutional action and technical design as practice, talk and material arrangement the scales are rendered observable to the analyst; there is nothing in advance to limit sociological studies to those questions that are deemed of the ‘correct size’ or as ‘sufficiently social’. The work of technical design, science policy, and scientific practice are, in principle, all equally observable at the scale of their enactment. In any study of a project such as GEON where members seek to act upon its size by making it *the* geosciences network the work of making it so is not only observable, more strongly, it should be observed.

Beyond the Scales: Limits of Analysis and Future Research Trajectories

A primary method in this dissertation has been to ‘follow the actors’(Latour 1987). More figurative than literal, this approach encourages the researcher to observe work where it happens irrespective of the locations and actors involved. Rather than the investigator, it is the actors that determine the sites of research; it is the actors that link across locations or times. This methodological rationale led me across the scales of action as GEON participants engage in institutional, organizational and technical work.

However, the three scales of action I have adopted are by no means the only sites where I have been led by GEON actors. In fact the list is seemingly endless. GEON has relationships to business, such as IBM, Sun and ESRI. It is tied to open source (FL/OSS) movements for grid technologies (Globus), for visual representation (GMR), and knowledge mediation (Kepler). Beyond the US, 'iGEON' has held conferences in India, and Japan, it draws on ontologies from geoscience institutions in Canada and the UK, and it has a 'sister' projects such as AEON, 'the Earth and Ocean Network' in Australia. Thus I could describe an entire set of scales such as business, FL/OSS, and international.

The choice of research sites has been pragmatic, emerging in the iterative process of data collection (Strauss 1993), informed by my own set of interests (Clarke 2005), and a set of concerns defined by the broader field of research (Edwards 2003). Below I would like to elaborate two limits of this research, sites that I have not traced in this dissertation but which are planned trajectories for the future.

'Following the Actors' Leads to Cyberinfrastructure Writ Large

We have seen how in enacting the logic of interoperability the disciplines of the earth sciences are cast as heterogeneous in method, in language and in data structure. We have seen the problematization of difference in the face of scientific advance and multidisciplinary collaboration. And we have seen a solution posed in mediation, in the development of an informational infrastructure which translates across disciplinary difference while preserving it. In this dissertation I have focused on the work within GEON – across the scales – in enacting this vision of cyberinfrastructure. However the limits of the logic of interoperability are not coextensive with those of the earth sciences. A future trajectory of this research is to follow the actors of interoperability as they travel across even the broadest of disciplinary differences – to the atmospheric sciences, to the biosciences, from science to engineering – to higher orders of interoperability.

One site to which I have been continuously led by GEON actors is to the SDSC itself. Half of the GEON PIs are themselves members of the SDSC. The SDSC is the technical core of GEON, but it also the technical core for many other CI projects. Following the actors, as a rationale, led me from GEON to another CI project, and then another and another.

For example, within weeks of beginning my research with GEON I had already gained strong interest in the technologies of knowledge mediation. An endeavour to capture and represent knowledge, to automate its actions, and translate across differences has an immediate appeal to any researcher in Science and Technology Studies (STS). A precept for any STS scholar is to understand the technical content of the research site. To do so I followed my actors. I asked the GEON IT PIs specializing in knowledge representation what I needed to know, how I should work through the technicalities of the subject. And so began a trail that led to a multiplicity of cyberinfrastructure projects, from SEEK, to BIRN, to NEES and to LEAD.

First I began with a course, taught by GEON IT PI Bertram Ludaescher. Here I learned of the history of knowledge engineering, artificial intelligence, the use of ontologies in business and, of course, of description logics, concept spaces, and knowledge bases. At the end of this course, along with graduate students in computer science, I presented my research to the class. This presentation was an initial outline of the routinization I describe in

Ch. 5. I placed a great emphasis on the ‘building of ontologies in action.’ At the end of this presentation Bertram said to me, ‘if you really want to know about ontologies in action you need to come to DAKS’. And so I followed this GEON PI to DAKS.

DAKS is Data and Knowledge Acquisition Systems, the larger team of knowledge engineers at the SDSC. They meet on a weekly basis. At this meeting I encountered a room full of computer scientists concerned with domain knowledge but with no domain scientists present. They spoke of the knowledge of the solid earth sciences, yes, but also the knowledge of the biosciences, environmental sciences, earthquake engineering and hydrology. Knowledge engineers at the SDSC are involved in CI projects which cover all these domains. To name only a few Bertram is a PI or co-PI on the Biomedical Informatics Research Network (BIRN); Real-time Observatories, Applications, and Data management Network (ROADNet); Science Environment for Ecological Knowledge (SEEK), as well as GEON. Knowledge mediation comes in contact with efforts to integrate, mediate, automate or connect by workflow a very broad range of the sciences and engineering. These general technologies of the particular are by no means limited to earth science knowledge.

In this research I most closely followed GEON’s knowledge mediation efforts, but it is apparent that a similar phenomenon could be traced out from each of GEON’s architectural divisions: systems, knowledge representation, and visualization (Ch. 4.1). The networking and Grid innovations in GEON are not limited to GEON, and neither are the visualization tools. The middleware initiatives ‘Rocks’ and ‘GeMS’ mediate a general grid with specific data and Linux computing clusters (Papadopoulos, Katz et al. 2001; Zaslavsky, Baru et al. 2005). The SDSC’s storage resource broker (SRB) is an archive for data generally, not specifically for any science (Baru, Moore et al. 1998). At the SDSC

Geographic Information Systems (GIS) technologies are by no means limited to the earth sciences, they have also been applied, for example, to 'mapping' the brains of mice in BIRN (Zaslavsky, He et al. 2004).

I have argued that in the logic of interoperability the classic 'tension' in science between general and the specific is *problematized*. It is the object of research and a source of difficulty to be challenged through the general technologies of the particular. Disciplinary boundaries are characterized in order to be traversed. In following GEON actors I have focussed on the enactment and mediation of boundaries in the geosciences: paleobotany, to geophysics to seismology. Ambitious as this effort may be, the limits of interoperability, of the unproblematic movement of information, extend far beyond. This logic is applied at all scales of disciplinary difference.

As Berg and Timmermans have noted, not only is the constitution of order also the performance of a disorder, but "a lack of uniformity emerges as that-which-needs-to-be-tamed" (Berg and Timmermans 2000:46). Similar to this but in a logic of relation rather than standardization, in the enactment of interoperability larger boundaries only mandate a larger statement of problem and a respectively larger set of territories to mediate. The CI project CLEANER plans to bring under its umbrella hydrological *scientists* and environmental *engineers* (the basic/applied boundary). The NSF TeraGrid is a higher order umbrella with 'science gateways' to multiple CI projects such as – to name only a few – LEAD for atmospheric sciences, National Virtual Observatories (NVO) for astronomy, NEES for earthquake engineering and GEON. For every order of magnitude greater in the scale of boundary crossing a correspondingly deeper translation into the homogeneity of information is needed. The boundaries of science and engineering, writ large, are the object of problematization for interoperability. This is one future trajectory of my research.

Endogenous Enactment of Evaluation in GEON

The question 'are efforts at interoperability effective?' has not been part of this study. First, because I have focused on something much more tangible or observable in the sense imparted by ethnomethodology: the reality of efforts for interoperability. This respecification of interoperability should inform its evaluation; a predecessor to evaluating effectiveness is formulating an understanding of the phenomena to be evaluated. Second, because, although it is not a feature I have traced in detail in this dissertation, evaluation is an activity endogenous to GEON. GEON has had site visits by the NSF and a regular committee of external advisors. As with its predecessor KDI, the ITR program itself is currently the object of an evaluation. As the GEON ITR lapses in 2007 it will be placed under great scrutiny as various possibilities for continuing the project are considered. The methods by which GEON will be evaluated are at play in the enactment of the project more generally. To elaborate both points I will discuss the examples of multidisciplinarity, knowledge capture and the formal evaluation of GEON.

In various forms I have taken a similar approach to the question of effectiveness relative to the specific topic of chapters in this dissertation. In Ch. 2 I argued that it is the wrong question to ask 'are we really multidisciplinary?'(Weingart 1997) or 'really entering Mode 2 science?' (Gibbons, Limoges et al. 1994). Instead I showed the work at the NSF as actors *do* organize around multidisciplinarity to make projects like GEON a reality. I also showed the means by which such multidisciplinarity is evaluated, such as in social science studies by Cummings and Kiesler.

I am not arguing that asking the question ‘whether these collaborations are effective?’ is not valuable. Instead I am adding a complement to such studies. I point to the fact that multidisciplinarity is always already an actor’s category organizing practical activities. Furthermore, I argue that understanding how multidisciplinarity is enacted today is a precursor to any gesture of evaluation tomorrow.

Thus, I have not asked ‘can we really capture knowledge in ontologies?’ Instead I have focused on the work of articulating knowledge in forms that can be captured in the formal description logics of ontology. Whether this is ‘really knowledge’ is a different question from investigating what GEON participants are doing when they say they are building ontologies. ‘Will ontologies permit the large-scale interoperation of geoscience data?’ is a very interesting question, and it is one that is discussed on an almost daily basis within GEON, and within the community of earth scientists as GEON’s vision is extended. The evaluation of ontologies as an approach to interoperability is an endogenous and ongoing activity amongst members. I hope that understanding the practical process of specific knowledge capture, and its general routinization, will contribute to discussions and enactments of its evaluation.

--

We have seen how ‘what is GEON?’ is at stake in the daily activities of its participants. Is GEON a community building endeavour for the earth sciences? If so partnerships and alliances are a marker of success. Or is the work of GEON a contribution to the earth sciences? If so, success is only to be marked as discovering ‘something new about the Rockies' and 'When Joe Blow at X University can access data sitting at his university.' In this dissertation I have focused on efforts to build GEON and in doing so how participants have come to negotiate its purposes and goals, and the institutional, organizational and technical means to achieve these. In negotiating purpose and identity, in the enactment of goals, there is the formulation of an evaluation. A future trajectory of research will track the methods and modes by which evaluation of cyberinfrastructure is conducted.

After Universal Informatics

Every instance of 'hype' such as that which has come to surround cyberinfrastructure is accompanied by its sceptics. To them CI is a 'buzzword,' a policy thrust without any clear referents, and sometimes more simply 'not real science'. For these sceptics talk of a revolution is highfalutin, to be dismissed as nonsense, rhetoric, or ideology. I have made no attempt to dissuade these critics of CI. Rather, my efforts have arisen and then departed from these criticisms in two tangents (Boltanski and Thevenot 1999). First, I argue that at the intersection of CI efforts and its critiques is a strengthening of the logic of interoperability as it incorporates its others. Second, I argue that "hype is constitutive, it mobilises the future into the present" (Brown 2003:6) that work today has consequences for an emerging order and a larger vision of the interoperated sciences.

I have shown how critics are themselves participants in constituting the trajectory of CI. The sceptics are not outside the realm of CI, they too are participants in shaping its outcomes. CI endeavours such as GEON bring together multiple charged terminologies, many of which maintain parallel meanings, are actively debated or, are the topics of contemporary research. For example, in this text I have drawn repeatedly on the actors' categories 'infrastructure', 'multidisciplinarity,' and 'knowledge.'

Each of these represent a goal of cyberinfrastructure, a contested category amongst participants in cyberinfrastructure, and a research object in multiple specialties. Around the negotiation of these categories there is considerable 'heat'. Much more is at stake than a definition of *meaning* ('what is infrastructure' or 'what is knowledge'). These debates occur at the intersections of design, planning and policy. Will GEON as an infrastructure transform the earth sciences, or will it support existing work? Is it a community building project, or the development of technical tools for conducting science? How is an umbrella infrastructure to represent the totality of geoscience? On each occasion in which I have observed the negotiation of such terms, what has been at stake are the purposes, means and future trajectory of the GEON project. Enactment of CI is as much a work of envisioning a future, of planning for a model scientific activity, and of considering the equitable distribution of resources as it is a work of on-the-ground technical development.

We have seen how GEON has come to be defined to be and then enacted as a contribution to the earth sciences. It must be a 'high risk, high reward' implementation of novel information technologies for the earth sciences, and it must produce 'something new about the Rockies.' This formulation is deeply shaped by a continuous engagement with the sceptics of cyberinfrastructure who see domain science money pouring into the coffers of computer science. The response has framed each infrastructure building project as 'also a contribution to science'.

Daily these criticisms are wrapped into its material construction. As Berg and Timmermans note "not only does an order perform its own disorder – it also always contains it […] the heterogeneous network acquires its existence partly through incorporating instability and multiplicity in its very core" (Berg and Timmermans 2000: 37-38). The critique of cyberinfrastructure is endogenous to its own execution. I have attempted to show that each activity within GEON has been negotiated, sometimes contested, and often redirected by participants themselves, or the representatives of the institutions of geoscience.

We have seen the representatives of geoscience reject a 'top-down' imposition of standards on the earth science community. In response to such criticisms the formulation of CI has becomes complex and nuanced. Institutionally GEON is not top-down, it is 'coordinated from above' and 'built by domain scientists themselves'. Organizationally it is 'twice basic research and infrastructure to boot'. And technically 'GEON doesn't make standards.' Their goal is mediation not standardization, and an entire set of emerging technologies and practical routines are the response of computer science. As each criticism, instability and multiplicity is incorporated in its very core, the logic of interoperability becomes entrenched, more subtle and nodal (Foucault 1995[1975]). Its power is not diluted but distributed as scientists, data, and knowledge are made increasingly 'collaboration ready'.

--

Negotiations and criticisms are part of the practical work of enacting CI; however it is not all the work. At times a cacophony of disagreement will be heard in GEON, while at others it will speak in one voice and act in a marshalled stride. I have shown the on-the-ground work which has come to be named cyberinfrastructure. Today the critics may still *point* to cyberinfrastructure, large-scale interoperability, or Geoinformatics as fictions but the work that has gone into their constitution is not fictitious; in other words, the critics are pointing at *something*. We need not rely on 'the success of GEON' to mark the consequences of action in GEON. I argue that this work has effects which overflow any formulation of success.

Participants in GEON have been daily engaged in the work of its construction in the institutions of geoscience, at the SDSC and in the technicalities of its knowledge. The logic of interoperability and its problematic have been extended at all three scale of action, whether institutional, organizational or technical. *They are increasingly a reality for earth science.* Today the data of the geosciences have come to be in peril, new organizational forms such as CI and Geoinformatics are recognizable institutions, and more geoscientists can 'use the word ontology in a sentence'.

Olson and Olson describe a prerequisite for the enactment of distributed multidisciplinary relationships as 'collaboration readiness.' "Using shared technology *assumes* that the coworkers *need to share* information and are *rewarded* for it. *Different fields and work settings engender a willingness to share*," (Olson and Olson 2000:164, emphasis added). In the institutions of science, in the organization of GEON and in the enactment of a technical logic each is coming to be shaped as *needing and ready* for a technically mediated collaboration.

We have seen the practical work as NSF officers encourage multidirectorate collaborations, choose to fund proposals with coordination plans deemed effective, and arrange for the rewards of basic science research to be tied to infrastructure development. We have seen the rise of Geoinformatics as an institutionalized venue for funding and bringing together informatic efforts in the earth sciences. We have seen GEON participants formulate and enact a 'two-tier approach,' a method for systematically generating commitment to a vision of interoperated infrastructure. The generation of a pull to interoperability is an effect which transcends the buzz of cyberinfrastructure.

To collaboration readiness Olson and Olson add ‘technology readiness,’ the precursor habits which are conducive to the uptake of novel technologies. This too we have seen as geoscientists learn the problematic of interoperability, the solution of ontology, and how to enact a routine for knowledge capture. This routine extends the logic of interoperability beyond GEON, it is an ‘outreach’ activity as geoscientist pioneers in ontology spread the word and indoctrinate the earth sciences in the logic of interoperability. At each scale of action cyberinfrastructure is an engine finding and consuming the differences in the sciences and generating ‘collaboration ready’ institutions, actors and technologies.

An entire generation of scientists are becoming tempered in the skills of multidisciplinary collaboration, in working across boundaries with information technologists, and in formulating their work relative to the logic of interoperability. An entire set of organizing practices are reshaping institutions to ‘incentivize’ data archiving, sharing, writing of metadata and registration. An entire panoply of technologies are emerging with the double “responsibility” of supporting specific scientific practices while linking to an increasingly generalized information infrastructure. GEON’s purposes, its methods and even its success today remain in question, but in the wake of resolving these questions a new order is being shaped in the image of interoperability.

–Appendix A –
Abbreviations of Institutions and Organizations and Glossary of Technical Terminology
–

As noted in the introduction the main body of the text will sustain a low level of technical detail, and every effort will be made to make technological nuance accessible to non-specialist audiences. In this appendix a glossary is offered, which outlines many of the technical terms of use *within GEON*. "Within GEON" is key, because these definitions do not necessarily line-up with what may be general understandings within computer and information science, they are my attempt to capture the meanings at play within GEON. Furthermore, meanings themselves may be contested, shifting or multiple e.g. discussions of ontology are never quite as grand as when information technologists are present with geo-scientists. In line with this, as much as possible I have attempted to provide references which appear within GEON usage or which are recommended by participants.

Abbreviations of Institutions, Organizations, Cyberinfrastructure Projects

Atkins Report – Revolutionizing Science and Engineering through Cyberinfrastructure, see bibliography

ACP – Advanced Cyberinfrastructure Program

The Alliance – The National Computational Science Alliance

BIRN – BioInformatics Research Network

CI – Cyberinfrastructure

CISE – Computer and Information Science and Engineering, directorate at the NSF

CHRONOS – Not an acronym (Greek for time) EAR funded CI project for History of the Earth

DLESE – Digital Libraries for Earth Science Education

DoE – Department of Energy

EAR – Earth Science Research (sub division of GEO)

Earthscope – Earth science instrumentation project. Funded as a MRE.

GEON – the GEOsciences Network

GEO– Geoscience, directorate of the NSF

Geoinformatics – intersection of geology and information technology, now a conference, a funding line in the NSF, and an umbrella organization

Hayes Report – Report of the Task Force on the Future of the Supercomputer Centers Program, see bibliography

IT – information technology

ITR – Information Technology Research - a grant program of the NSF, See Ch. 2.1

KDI – Knowledge and Distributed Intelligence – a grant program of the NSF, See Ch. 2.2

LEAD – Linked Environments for Atmospheric Discovery

LTER – Long Term Ecological Research

MRE - Major Research Equipment. Large projects funded by Congress.

NCSA – National Center for Supercomputing Applications (tied to the Alliance)

NEES -- Network for Earthquake Engineering Simulation

NPACI – National Partnership for Advanced Cyberinfrastructure Initiative (tied to SDSC)

NIH – National Institutes of Health

NSF – National Science Foundation

OCI – office of cyberinfrastructure (a division of the NSF)

PITAC – President's Information Technology Advisory Committee

RIBES – Resourceful Intelligent Bucolic Entertaining Smartass

SCEC – Southern California Earthquake Center (both a Science and Technology Center and an ITR)

SDSC – San Diego Supercomputer Center (tied to NPACI)

TeraGrid – also Terascale Grid An NSF large-scale systems development project, see below.

UCSD – University of California San Diego

USGS- United States Geological Survey

Glossary of Technical Terminology

Data/Dataset/Database/Metadata: The categories are heavily at play within GEON. For the reader I define with a risk of their reification. Note however that actor's uses of these terms differ broadly relative to their specializations (amongst scientists, information technologist, knowledge representation expert, database researcher, information manager).

- **Data** in common usage is one step removed from the empirical objects of the sciences. 'Raw data' (a contested term) comes directly from instrumentation without extensive cleanup or representation as, for example, visualization. Data usually refers to information stored on a computer, 'digital data', although in the geosciences there are large collections which have not been digitized. This is often referred to as 'legacy data' – although this term is also used to refer to a collection from a previous era of database systems. For information technologists data is the raw material itself on which infrastructure seeks to operate.
- A **dataset** is a collection of data. In geology this is usually data covering a particular geographic area for a particular time period. At times a dataset refers to data which has been organized into a database, on other occasions it refers to the informational content of a database.
- A **database** is an organized collection of data, usually in a computerized form. There are several fundamental distinctions of databases which refer to the method of organization: flat file, relational, object oriented and so on. Each of these methods enables particular sets of operations and manipulations of the data. The software used to manage and query a database is known as a database management system (DBMS). In theory databases are divided into four levels: semantic, structure, logic and system (the terms vary). Ontologies address the semantic level, structural and logical relations are described by schemas and metadata and the system is the physical storage and its address.
- **Metadata** (an emerging term) is data which describes data. It is a key term in any question of interoperability as these descriptions can be used to link across databases. Metadata can also be qualitative, including descriptions of the data, its purposes for collection, kinds of discrepancies in collection and so on. At times ontologies, or registration to an ontology, are considered a form of metadata.

See also: system, ontology, schema, data independence, and view.

Data independence: A subset of *interoperability*. The 'independence' is usually from a particular program. Physical data independence means that programs will be able to access a dataset regardless of its physical organization or location (multiple disks or Grid distribution). Logical data independence refers to the ability to change the structure of the dataset without making it inaccessible to particular programs – this is often achieved using *views* (Gray, Liu et al. 2005). See also view.

Extensibility: “In systems architecture, extensibility means that the system has been so architected that the design includes all of the hooks and mechanisms for expanding/enhancing the system with new capabilities without having to make major changes to the system infrastructure.” (http://en.wikipedia.org/wiki/Extensibility Feb 2006).

Extensible Markup Language (XML): a standard for exchanging structured documents and data on the web. Subset of **SGML**. Adopted as a standard in 1998 by the World Wide Web Consortium (W3C, http://www.w3c.org).

Geographic Information Systems (GIS) – Refers to a general set of computer mapping technologies. GIS is often described as an ‘end-to-end’ solution in that its tools have the capacity to store, assemble, manipulate and represent data. These displays are interactive i.e. unlike a physical map, or jpeg rendering, output images in GIS can be clicked on to ‘drill down’ to more detailed representations or data. GEON has a partnership with the GIS company ESRI.

Grid: Grid computing uses the resources of a many separate computers connected by a network (usually the internet) to solve large-scale computation problems. Well known examples include the SETI@home project, launched in the mid-1990s which distributed the computing of astronomical (radio) data across people’s homes in the search for extraterrestrial life. This remains a topic of research in academic and business computing circles. Its meaning often stretches beyond allocating computing to include security, access, or prioritization. Metaphors of the grid as ‘virtual organization and an economy speak to the extension of the concept (Foster, Kesselman et al. 2001; Buyya 2002; Crosby and Arrowsmith 2004; Bhatia, Chandra et al. 2005). For the Globus Toolkit, an open source platform for grid computing, see (Foster and Kesselman 1998).

Interoperability: The term is notably nebulous in computing circles. It is also a key term in cyberinfrastructure, referring primarily to the technical compatibility of differing data formats, but also movement across operating systems, software platforms or tool suites (Sheth 1999). This is a key term in this dissertation which is, in many senses, an attempt to respecify the term (Garfinkel 1991)

Logical Data Independence: See data independence.

Logical Operator: Also described as Boolean operators. Amongst GEON participants the term is used loosely, rather than as limited relative to, say, algebraic logic. In a query, it is a connector between two expressions. In an ontology they are the links between concepts or entities. By convention (and use) are three primary logical operators: AND, OR, and NOT. Efforts in knowledge representation draw on description logics, which extent the number of operators substantially.

Metadata – see data.

Ontology – also described as concept maps. "A specification of a conceptualisation of a knowledge domain." An ontology is a controlled vocabulary that describes objects and the relations between them in a formal way. They are usually understood to capture meaning in a particular domain, and in GEON they are understood to translate across technical concepts or vocabularies. By operating on a standardized vocabulary and the formal 'logically defined links' across them, an ontology can used to make queries and assertions. (Sowa 2000). Please see Ch. 5 (esp. 5.1) for an extended discussion.

OSI - Open Systems Interconnection. An internationally recognized set of standards for communication between computer systems.(Egyedi 2001).It is a standard description or 'reference model' divided into seven layers with the intention of making independent, and thus more manageable, the varying functions of communication. Furthermore each layer is designed to develop independently while still linking with the others – this is often described as a modular architecture. The standard was partially developed and is supported by the International Organization for Standardization (ISO).

Physical Data Independence: See data independence.

Platform: Usually the hardware framework, but occasionally the software, which allows software to run: this may include the architecture, the operating system, or programming language. At the SDSC the term most commonly refers to the hardware/software combination of operating systems such as PC, Mac, Linux or UNIX. Platform independence refers to the ability of software to function across platforms (such as Java) or availability of software versions for multiple platforms (such as Microsoft Word for PC and Mac).

Platform independence: A subset of *interoperability*. See platform.

Portal: A portal is a website collecting many resources, services or tools. In GEON a more specific use of the term portal refers to the login-required 'inside' of GEON's services. The term also refers to the software tools used to produce the web interfaces, in this case GridSphere (see http://www.gridsphere.org/gridsphere/gridsphere). In a GridSphere portal, the various sub-services are described as portlets. See portlets.

Portlets: Within the portal software GridSphere, and several other portal platforms, the subsections of the portal are called portlets. These can be locally developed or third-party functionalities. GridSphere provides the integrating 'framework' for the multiple portlets. See portal.

System: In GEON the physical component of the architecture, including networking, storage and computational resources. Includes also the Grid and issues of access and security (see Ch. 4.1). The term is also used to refer to the computer in general, for example, ontologies are said to make semantic information 'accessible to the system'. See also Grid.

Schema: "Database systems use a schema to implement both logical and physical data independence. The schema for a database holds all metadata including table and view definitions as well as information on what indices exist and how tables are mapped to storage volumes (and nodes in a parallel database environment)." (Gray, Liu et al. 2005) p36 see also *data independence.*

View: A view defines a 'virtual table' drawn from one or more datasets, thus excluding data (e.g. for privacy or succinctness) or re-organizing the logical data structure (see also data independence). A view for a user can synoptically and concisely represent information; a view for the system can enable interoperability. (Gray, Liu et al. 2005).

SGML -- Standard Generalized Markup Language: A standard to describe document properties using fields (heading, paragraph, lists, and figures). Initially developed at IBM, then derived at American National Standards Institute and then adopted by the Department of Defense. Became an international standard in 1986.

TeraGrid: A National Science Foundation project to build a large scale distributed grid for science and engineering research and computation. "The TeraGrid enables scientists and engineers to combine distributed, multiple data sources with computation at any of the sites or link massively parallel computer simulations to extreme-resolution visualizations at remote sites. A single shared utility lets multiple resources be easily leveraged and provides improved access to advanced computational capabilities" (Beckman 2005). The project was funded partially by ITR monies, and following the reorganization of the NSF has come under the supervision of the OCI. GEON has been a marginal participant in the testing of early phases of the TeraGrid, including being given 'compute cycles' on the grid. See grid, and OCI.

XML: see Extensible Markup Language.

–Appendix B –
The Meetings of GEON
–

This appendix provides an extended description of the meetings of GEON. These have been the primary sites of ethnographic observation over the three years of data collection in this research. Here I attempt to capture some of the laminated meanings which these gatherings have come to acquire over time. In an ambitious and distributed project the various meetings are more than simply summary, review, and planning, they have also been the occasion for renewing and transforming the purposes of GEON. Ch. 4.1 provides a sustained discussion of this rationale for an empirical focus on the meetings of GEON.

The meetings, in order of presentation are the Kickoff and All-Hands (AHM), the PI meetings (PI meets), the Internal Coordinating Committee (ICC), the workgroups, workshops and the concept space meetings. Ch. 4.1 also includes a summary of these meetings in chart form.

Kick-Off and All- Hands Meetings

Within GEON, and in this study, 'the kick-off meeting' is referred to as the official start to the project[1]. This meeting was held in San Diego, at the UCSD campus at the SDSC on November 17-20th, 2002. The kick-off was an assembly which primarily served to introduce the IT team and their planned technologies to the geo-scientists. A

[1] Note: GEON had at least one previous kick-off meeting mentioned in Ch. 2.3. However, in common parlance, it is the November 2002 meeting to which kick-off refers. The NSF award date for the GEON ITR is July 2nd 2002.

great deal of time was spent in presentations and Q&A about these novel and emerging technologies. In turn, geo-scientists presented their *science questions* and some initial descriptions of the kinds of data and integration necessary to achieve their goals. These presentations were directed as much at the IT team as at the other participating earth scientists. Many of these geo-scientists were only loosely familiar with each other as they are nationally distributed and from multiple disciplines. Many of the PIs also brought retinues of graduate students and geo or CS colleagues from their home institutions. Representatives from the United States Geological Survey (USGS), Digital Libraries for Earth Science Education (DLESE) and the Geological Survey of Canada also attended. Presentations also included appearances, for example, by the director of the SDSC Francine Berman, and partner CI projects such as BIRN. The first presentation was conducted by a lead Geo PI, with a history of GEON and Geoinformatics (see Ch. 3.1 for a later version of this history) and broad statements as to the purpose of GEON. The introductory quote for the presentation was by the eminent geophysicist Frederick Alfred Sutton: "we must look at matters on a rather larger scale if we are to see how the component parts of the mountain fit together."

Lead IT PIs identified GEON's key issues as 'semantic integration' while also encompassing grid (see below, and Appendix A), visualization and beginning to define the Unified Geosciences Language System (UGLS)[2]. Notably, an IT PI argued that all participant's goals should be directed at "*prototyping* a national information

[2] UGLS is mentioned in the proposal and the kick-off meeting, and then rarely makes an appearance in GEON again. Endogenous comparisons (in both the proposal but more so at the kick-off) were with the United Medical Language System (UMLS)(Humphreys and Lindberg 1993), which was characterized as a "metathesaurus", "semantic network", and a "lexicon". While initially this appeared to be a fascinating line for ethnographic inquiry, the project quickly lost momentum and eventually disappeared from active discussion altogether.

infrastructure for the geosciences," and went on to note with great emphasis on the word prototype that "this is all we are going to do".

--

The All-Hands Meetings (AHM) have shifted significantly in purpose over the years. The first two AHM were at the SDSC. They were intended to bring together all GEON PIs, their research retinues, and the growing collection of partner institutions. The explicit goals were for each PI to provide summary presentations of distributed work, collecting the necessary resources to govern GEON and in turn forming coordination plans for future action. An agenda, produced in advance at smaller meetings, through email exchanges and then continuously tailored, would typically include presentations from all the PIs and their research teams, as well as updates on particular IT developments, 'demos' of relatively stable IT tools such as a visualization, ontology or web service. These meetings were intensively conducted over two or three days, mostly in the form of power point presentation followed by group discussion. Smaller periods of topical break-out sessions or closed-door PI[3] meetings were also conducted. Dinners often followed meetings, and some form of geological sightseeing tour usually closed the AHM and PI meetings.

The third AHM was a significant turning point for GEON. Held at a conference center in San Diego, rather than the SDSC, it ceased to be primarily a governance meeting and became an exercise in community outreach. While still GEON centered, the third AHM included a public solicitation for papers and posters. The meeting included

[3] There was never an occasion in which I was asked not to attend. At the el Paso PI meeting I willingly excused myself (and my recorder) from a closed door PI-only meeting of approximately 1 hour.

many presentations which were not ‘GEON projects’. Finally the fourth AHM became Geoinformatics 2006, co-hosted by GEON and USGS, far from San Diego in Reston, Virginia.

During the bulk of my ethnography, AHMs maintained the sense described here for the first two meetings. A great deal of preparation and planning previous to each meeting was marked by intensified email and list-serve exchange prior to meetings, and then summation work (reports) and ‘action item’ lists were circulated following these meetings. While the term’s meaning has shifted over the years, from a governance meeting to a growing geoscience conference, participants have retained the term and it is still an occasion of extensive planning, collaboration and formalization of GEON’s progress.

PI Meetings

For the first two years PI Meetings (or “PI meets”) were held bi-annually, in addition to the AHM meetings. Locations were rotated across the geoscience PI’s home institutions. In later years these meetings were reduced to once per year, and timed to precede important geoscience conferences (such as GSA). These meetings included the GEON PIs, but also a smaller selection of their research teams and representatives of significant regular partners such as DLESE.

The PI meetings were shorter (usually two days) but otherwise structured very similarly to the AHMs. They included a loose agenda, circulated in advance and modified substantially thereafter. Three out of the PI meetings I attended were explicitly ‘science focused’ with greater time distributed to geoscience research, and outcomes than those of the IT PIs (see Ch. 2.3 for the tenuous sense of 'balance'). This said, one morning would usually be dedicated to geoscience research summaries and reports and another to IT tools, progress and planning. Breakout sessions were planned in the afternoon for topically oriented planning. The sessions were usually followed by a collective dinner (often at a PI's home), and a geological sightseeing tour in the final day's afternoon.

Internal Coordinating Committee

The Internal Coordinating Committee (ICC) was a regular distributed meeting held monthly. The group was small, usually involving only the lead PI, the project manager, the two leads for the test-beds (see below) and GEON's administrative coordinator. This was also the only regular meeting conducted over videoconferencing technology, with multiple participants co-located at the SDSC and test bed leads communicating with them via Polycom[4]. At the SDSC the technical setup for distanced communication was facilitated by an experienced staff.

The meetings were primarily administrative, including setting the agendas for AHM and PI meets, discussing new GEON partnerships, or planning for upcoming conferences, posters and demos. The small group of selected participants and usually tight agenda gave these meetings a very decisive feel, particularly in contrast to AHM and PI meets, which ranged broadly in topic, purpose and style. In the terms of enactment, this meeting offered a relatively clean vantage point for observing GEON act *as* an organization.

[4] A common tool for distanced communication; this model is a high-end conference phone, although full Polycom systems include a sophisticated and high-bandwidth video connection.

The SDSC Workgroup Meetings

The workgroup meetings are the weekly meetings for activity in GEON that is centered at the SDSC. These come closest to representing the 'everyday' site for managing GEON. These meetings bring together the top level administrative managers of GEON (lead PI, project manager, administrative coordinator), along with the central team of IT experts. GEON's larger IT team is composed of graduate students, post-docs and PhDs and non-PhDs employed at the SDSC and dedicating a portion of their time to GEON. In short the composition of this meeting was primarily (IT) technical and administrative, rather than geoscience per se.

Even with the presence of a geoscience trained project manager, and occasional geoscience visits, at this meeting geoscience was something 'out there'. In contrast with AHM or PI meets (where this was highly unlikely), in this meeting it was quite possible to bracket out geoscientists and speak exclusively of the IT aspects of GEON. SDSC participants could switch freely between their various projects (e.g. from GEON, to BIRN to SEEK), often discussing shared technical platforms.

This meeting has become the organizational nexus for GEON. Discussions and agenda items shift fluidly between 'science questions,' technical developments, organization and communications issues, NSF evaluation, recruitment to GEON, and 'politics' – it is by no means a traditional administrative or managerial body.

Over the years various efforts were made to open this meeting to the non-SDSC, primarily geoscientists. This has included establishing a phone-in line for listening and participating, and in later years a live webcast. During visits to the SDSC, GEON PIs would be invited to participate and present their work to the group.

Following a major local (SDSC) re-organization of GEON in January 2004, the single weekly meeting divided into a changing configuration of daily meetings. These daily meetings were IT topically centered. The particular meetings have shifted over the years, but the initial divisions were: portal, knowledge representation, visualization, systems, GEMS, GIS/Mapping and workflows (see below and appendix A).

The original weekly workgroup meeting was maintained to provide a point of coordination for the daily technical meetings and the "rest of GEON". This weekly meeting was subdivided into the administrative and technical portion. For example, the first hour would be dedicated to "GEON business" of administration and coordination. The second hour would be dedicated to technical matters, such as updates from the SDSC technical team and perhaps a demo. Beginning in 2004 some of these technical presentations were web-cast and archived on the GEON public website.

GEON Workshops

GEON workshops are focused and topically specific, they are events rather than regularly held meetings, and often involve many participants from outside GEON's core group. For example, GEON has had workshops on visualization and ontology, and also a larger 'Cyberinfrastructure Summer Institute for Geoscientists' for basic introduction, framing science questions, and training with emerging tools.

Concept Space Meetings

I will cover these meetings in greater detail in Chapter 5. They are the practical working meeting for developing concept spaces or ontologies. These are usually two or three day intensive small groups of specialized geoscientists (selection partially informed by on the integration problem) and knowledge representation experts. The groups are usually between 6-10 people.

Unlike the workshops, these meetings are not usually general (‘about visualization’) but primarily engaged with a specific set of tasks (‘integrating datasets’, ‘capturing knowledge’, representing knowledge’). Over time ‘concept space workshops’ were renamed ‘community based ontology development’ to reflect the growing focus on enrolling a broader geoscience community in GEON’s knowledge representation efforts.

– References –
–

Abbate, J. (1999). Inventing the Internet. Cambridge, MA, MIT Press.

Acemoglu, D. (1995). "Reward structures and the allocation of talent." European Economic Review **39**(1): 17-33.

Aki, K. (2002). "Synthesis of Earthquake Science Information and Its Public Transfer: A History of the Southern California Earthquake Center." International Handbook of Earthquake and Engineering Seismology **81**(A).

Akrich, M. (1992). The De-Scription of Technical Objects. Shaping Technology/Building Society: Studies in Sociotechnical Change. W. E. Bijker and J. Law. London, England; Cambridge Massach, The MIT Press**:** 205-224.

Akrich, M. and B. Latour (1992). A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. Shaping technology/building society: Studies in socio-technical change. W. E. Bijker and J. Law. Cambridge, MA, MIT Press.

Allen, T. and R. Katz (1995). "The project oriented engineer: A dilemma for human resources management." R&D Management **25**(2): 129-140.

Anderson, R. (1994). "Representation and Requirements: The Value of Ethnography in System Design." Human Computer Interaction **9**(2): 151-182.

Argyris, C. and D. A. Schon (1978). Organizational learning: a theory of action perspective. Reading, Mass, Addison-Wesley Pub. Co.

Argyris, C. and D. A. Schon (1996). Organizational Learning II: Theory, Method, and Practice. Reading, Mass, Addison-Wesley Pub. Co.

Ashforth, B. E. and T. Fried (1988). "The Mindlessness of Organisational Behaviors." Human Relations **41**: 305-329.

Ashworth, W. J. (1991). The Scientific Revolution: The Problem of Visual Authority. Conference on the Critical Problems and Research Frontier in History of Science and History of Technology, Madison, Wis.

Atkins, D. E. C. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, National Science Foundation.

Baker, K. S., B. Benson, et al. (2000). "Evolution of a Multi-Site Network Information System: the LTER Information Management Paradigm." BioScience **50**(11): 963-983.

Baker, K. S. and G. C. Bowker (2001). Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations. LTER Network Newsletter. **Fall**.

Baker, K. S., S. J. Jackson, et al. (2005). "Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment." Proceedings of the Hawaii International Conference for System Science, HICSS38.

Bardini, T. and A. Horvath (1995). "The Social Construction of the personal computer user." Journal of Communication **45**(3): 40-66.

Baru, C., R. Moore, et al. (1998). "The SDSC storage resource broker." Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research.

Becker, H. S. (1960). "Notes on the Concept of Commitment." The American Journal of Sociology **66**(1): 32-40.

Beckman, P. H. (2005). "Building the TeraGrid." Philos Transact A Math Phys Eng Sci **363**(1833): 1715-1728.

Bender, T. (1978). Community and Social Change in America, Rutgers University Press.

Berg, M. (1996). "Practices of reading and writing: the constitutive role of the patient record in medical work." Sociology of Health & Illness **18**(4): 499-524.

Berg, M. and S. Timmermans (2000). "Orders and Their Others: On the Constitution of Universalities in Medical Work." Configurations **8**: 31-61.

Berkenkotten, C. and D. J. Ravota (1997). "Genre as Tool in the Transmission of Practice Over Time and Across Disciplinary Boundaries." Mind, Culture, and Activity **4**(4): 256-74.

Berkley, C., S. Bowers, et al. (2005). Incorporating Semantics in Scientific Workflow Authoring. Proceedings of Scientific and Statistical Database Management, SSDBM'05, Santa Barbara, CA.

Berners-Lee, T., J. Handler, et al. (2001). "The Semantic Web." Scientific American **May**.

Beynon-Davies, P., H. MacKay, et al. (1997). User Involvement in Information Systems Development: The Problem of Finding the "Right" User. Proceedings of the 5th European Conference on Information Systems (ECIS). R. Galliers. Cork, Cork Publishing.

Bhatia, K., S. Chandra, et al. (2005). GAMA: Grid Account Management Architecture. First International Conference on e-Science and Grid Computing (e-Science'05).

Blau, P. (1973). The Organization of Academic Work. New York, Wiley.

Boden, D. (1991). "Review of "The Meeting: Gatherings in Organizations and Communities" by Helen B. Schartzman." Contemporary Sociology **20**(2): 229-230.

Boden, D. (1994). The business of talk: Organizations in action. Cambridge, UK, Polity Press.

Boden, D. and D. H. Zimmerman, Eds. (1991). Talk and social structure: Studies in ethnomethodology and conversation analysis. Berkeley, University of California Press.

Boltanski, L. and L. Thevenot (1999). "The sociology of critical capacity." European Journal of Social Theory **3**: 359-77.

Boose, J. H. (1989). "A survey of knowledge acquisition techniques and tools." Knowledge Acquisition **1**(1): 3-37.

Borgman, C. L. (2000). From Gutenberg to the global information infrastructure: access to information in the networked world. Cambridge, MA, MIT Press.

Bowker, G. C. (1988). Pictures from the subsoil, 1939. Picturing Power: Visual Depiction and Social Relations. G. Fyfe and J. Law. London, Routledge.

Bowker, G. C. (1994). Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940. Cambridge, London, The MIT Press.

Bowker, G. C. (2000). "Biodiversity Datadiversity." Social Studies of Science **30**(5): 643-683.

Bowker, G. C. and S. L. Star (1999). Sorting things out: classification and its consequences. Cambridge, Mass., MIT Press.

Bowker, G. C., S. L. Star, et al., Eds. (1997). Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. Mahwah, NJ, Lawrence Erlbaum Associates.

Bowker, G. C., S. Timmermans, et al. (1997). Infrastructure and Organizational Transformation: Classifying Nurses' Work. Information Technology and Changes in Organizational Work. W. Orlikowski, G. Walsham, M. Jones and J. DeGross. London, Chapman and Hall.

Bromley, A. (1999). "Staying Competitive." Science **285**(5429): 833.

Brown, N. (2003). "Hope against Hype: accountability in biopasts, presents and futures." Science Studies **16**(2): 3-21.

Burchfield, J. D. (1990). Lord Kelvin and the age of the Earth. Chicago, University of Chicago Press.

Bush, V. ([1945] 1960). Science -- The endless frontier. Washington, DC, National Science Foundation.

Butler, D. (1998). "Interdisciplinary research 'being stifled'." Nature **396**(6708): 202.

Button, G. (2000). "The Ethnographic Tradition and Design." Design Studies **21**: 319-332.

Buyya, R. (2002). Economic-based Distributed Resource Management and Scheduling for Grid Computing. Computer Science. Melbourne, Monash University.

Callon, M. (1980). Struggles and Negotiations to Define What is Problematic and What is Not: The Sociologic of Translation. The Social Process of Scientific Investigations. K. Knorr-Cetina. Boston, Dordrecht.

Callon, M. (1986). The sociology of an actor-network: The case of the electric vehicle. Mapping the Dynamics of Science and Technology. M. Callon and e. al., MacMillan.

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fisherman. Power, Action and Belief. J. Law and P. Kegan, Routledge.

Callon, M. (1991). Four Models for the Dynamics of Science. Handbook of Science and Technology Studies. S. Jasanoff, G. E. Markle, J. C. Petersen and T. J. Pinch. Thousand Oaks etc., Sage**:** 29-63.

Callon, M. (1994). "Is Science a Public Good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993." Science, Technology & Human Values **19**(4): 395-424.

Callon, M. (1998). An essay on framing and overflowing: economic externalities revisited by sociology. The Laws of the Markets. M. Callon. Oxford, Blackwell Publishers/Sociological Review.

Callon, M. and B. Latour (1981). Unscrewing the Big Leviathan: how actors macrostructure reality and how sociologists help them to do so. Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies. K. Knorr-Cetina and A. V. Cicourel. Boston, Mass, Routledge.

Callon, M. and F. Muniesa (2005). "Economic Markets as Calculative Collective Devices." Organization Studies **26**(8): 1229-1250.

Calvert, J. (2006). "What's Special about Basic Research?" Science, Technology & Human Values **31**(2): 199-220.

Cambell, D. T. (1965). Variation and selective retention in socio-cultural evolution. Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory. H. R. Barringer, G. I. Blanksten and R. W. Mack. Cambridge, MA, Schenkman.

Cambrosio, A. and P. Keating (1988). ""Going Monoclonal": Art, Science, and Magic in the Day-to-Day use of Hybridoma Technology." Social Problems **35**(3): 244-260.

Campbell, E. G., B. R. Clarridge, et al. (20002i). "Data withholding in academic genetics: Evidence from a national survey." JAMA **287**(4): 473-480.

Cartwright, N. (1999). The dappled world: a study of the boundaries of science. Cambridge, Cambridge University Press.

Casper, M. and A. Clarke (1998). "Making the pap smear into the right tool for the job: Cervical cancer screening in the United States, c 1940-1995." Social Studies of Science **2**(3): 255-290.

Ceci, S. (1988). "Scientists' attitudes toward data sharing." Science, Technology & Human Values **13**(1-2): 45-52.

Ceci, S. and E. Walker (1983). "Private archives and public needs." American Psychologist **38**(4): 414-423.

Chalmers, M. (2000). "A Historical View of Context." Computer Supported Cooperative Work.

Chin, G., J. Myers, et al. (2002). "Social Networks in the Virtual Science Laboratory." Communications of the ACM **45**(8): 87-92.

Chompalov, I., J. Genuth, et al. (2002). "The Organization of Scientific Collaborations." Research Policy **31**: 749--67.

Cicourel, A. V. (1981). Notes on the integration of micro- and macro-levels of analysis. Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies. K. Knorr-Cetina and A. V. Cicourel. Boston, Mass, Routledge.

Clarke, A. (2005). Situational analysis: grounded theory after the postmodern turn. Thousand Oaks, SAGE Publications.

Clifford, J. (1983). "On Ethnographic Authority." Representations **2**(Spring): 118-146.

Cohen, M. D. (1991). "Individual learning and organizational routine: Emerging connections." Organisation Science **2**: 135-129.

Collins, H. M. (1981). "Knowledge and controversy: Studies of modern natural science." Special Issue of Social Studies of Science **11**(1).

Collins, H. M. (1981). "On the Microfoundations of Macrosociology." American Journal of Sociology **86**(5): 984-1014.

Collins, H. M. (1985). Changing Order: replications and induction in scientific practice. London, Sage Publications.

Collins, H. M. (1990). Artificial Experts: Social Knowledge and Intelligent Machines. Cambridge, Massachusetts, The MIT Press.

Compte, A. (1975[1853]). Auguste Compte and Positivism: The Essential Writings. New York. G. Lenzer. Harper & Row.

Cowan, R. S. (1987). The consumption junction: A proposal for research strategies in the sociology of technology. The Social Construction of Technological Systems. W. E. e. a. Bijker, The MIT Press.

Crosby, C. J. and J. R. Arrowsmith (2004). LiDAR Data Distribution, Interpolation and Analysis on the GEON Grid - A Conceptual Framework Paper No. 60-6. Geological Society of America Annual meeting.

Crosy, C. J., J. R. Arrowsmith, et al. (2004). "Exploiting LiDAR for Regional Morphologic Correlation and Dating of Wave-cut and Fault-Controlled Landforms (Abst G13B-0810)." Eos Trans. AGU Fall Meet Suppl. **85**(47).

Cummings, J. and S. Kiesler (2003). KDI Initiative: Multidisciplinary Scientific Collaborations NSF Award No. II-S 9872996, National Science Foundation.

Cummings, J. and S. Kiesler (2005). "Coordination and success in multidisciplinary scientific collaborations." Social Studies of Science **34**): 1-20.

David, P. and A. Bunn (1988). "The Economics of Gateway Technologies and Network Evolution: Lessons from Electricity Supply History." Information Economics and Policy **3**: 165-202.

Dear, P. (2001). Science Studies as Epistemography. The One Culture? A Conversation about Science. J. Labinger and H. M. Collins. Chicago, University of Chicago Press**:** 128-141.

DiMaggio, P. and W. W. Powell (1983). "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields." American Sociological Review **48**: 147-60.

Donzelot, J. (1979[1977]). The Policing of Families. New York, Pantheon Books.

Douglas, M. (1986). How Institutions Think. Syracuse, Syracuse University Press.

Dourish, P. (2006). Implications for Design. CHI 2006, Montreal, Que. Can., ACM.

Dourish, P. and G. Button (1998). "On "Technomethodology": Foundational Relationships between Ethnomethodology and System Design." Human Computer Interaction **13**(5): 395-432.

Durkheim, E. (1915). The Elementary Forms of Religious Life: A study in religious sociology. New York, Macmillan.

Durkheim, E. (1976). The Elementary Forms of Religious Life. London, Allen and Unwin.

Edwards, P. (2003). Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems. Modernity and Technology. T. J. Misa, P. Brey and A. Feenberg. Cambridge, MA, The MIT Press.

Egyedi, T. (2001). "Infrastructure Flexibility Created by Standardized Gateways: The Cases of Xml and the Iso Container." Knowledge, Technology & Policy **14**(3): 41-54.

Eisenstein, E. L. (1983). The Printing Revolution in Early Modern Europe. Cambridge, England, Cambridge University Press.

Elzinga, A. (1997). "The science-society contract in historical transformation." Social Science Information **36**(3): 411-45.

Elzinga, A. and A. Jamison (1995). Changing policy agendas. Handbook of science and technology studies. S. Jasanoff, G. E. Markle, J. C. Petersen and T. J. Pinch. Thousand Oaks, CA, Sage**:** 572-97.

Epstein, S. (1996). Impure Science: AIDS, Activism and Politics of Knowledge, University of California Press.

Epstein, S. ((under review)). ""One Size Does Not Fit All": Standards, Categories, and the Inclusion-and-Difference Paradigm in the U.S. Biomedical Research." American Journal of Sociology.

Fairweather, J. S. (1993). "Faculty Reward Structure: Toward institutional and professional homogenization." Research in Higher Education **34**(5).

Feldman, M. S. (2000). "Organisational Routines as a Source of Continuous Change." Organisation Science **11**(6): 611-629.

Feldman, M. S. and B. T. Pentland (2003). "Reconceptualizing organizational routines as a source of flexibility and change." Administrative Science Quarterly **48**(1): 94-121.

Finholt, T. A. (2003). "Collaboratories as a New Form of Scientific Organization." Economics of Innovation and New Technologies **12**(1): 5-25.

Finholt, T. A. (2004). "Collaboratories." Annual Review of Information Science and Technology **36**: 73-107.

Fligstein, N. (2001). "Social skill and the theory of fields." Sociological Theory **19**(2): 105-125.

Forrester, J. (1996). "If p, then what? Thinking in cases." History of the Human Sciences **9**(3): 1-25.

Forsythe, D. E. (1993). "Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence." Social Studies of Science **23**: 445-77.

Forsythe, D. E. (2001). Studying Those Who Study Us. Stanford, CA, Stanford University Press.

Forsythe, D. E. and B. G. Buchanan (1989). "Knowledge Acquisition for Expert Systems: Some Pitfalls and Suggestions." IEEE Transactions on Systems, Man, and Cybernetics **19**: 435-442.

Foster, I. and C. Kesselman (1998). "Globus: a metacomputing infrastructure toolkit." Int. J. Supercomputer Appl. **11**: 115-129.

Foster, I., C. Kesselman, et al. (2001). "The Anatomy of the Grid: Enabling Scalable Virtual Organizations." International Journal of High Performance Computing **15**(3): 200-222.

Foster, N. F. and S. Gibbons (2005). "Understanding Faculty to improve content recruitment for institutional repositories." D-Lib Magazine **11**(1).

Foucault, M. (1970). The Order of Things. London, Tavistock.

Foucault, M. (1991). Governmentality. The Foucault Reader: Studies in Governmentality. G. Burchell and C. Gordon. Chicago, University of Chicago.

Foucault, M. (1995[1975]). Discipline and Punish. New York, Vintage Books/Random House.

Fountain, J. E. (2001). Building the Virtual State: Information Technology and Institutional Change. Washington, D.C., Brookings Institution Press.

Francis, D. and S. Hester (2004). An Invitation to ethnomethodology: Language, Society and Interaction. Thousand Oaks, CA, SAGE Publications Ltd.

Frederickson, H. G. and K. B. Smith (2003). The Public Administration Theory Primer. Boulder, Co, Westview Press.

Friedman, A. (1989). Computer System Development: History, Organisation and Implementation. Chichester, John Wiley.

Fujimura, J. (1987). "The construction of doable problems in cancer research." Social Studies of Science **17**.

Fujimura, J. (1988). "The Molecular Biological Bandwagon in Cancer Research: Where Social Worlds Meet." Social Problems **35**(3): 261-283.

Funtowicz, S. O. and J. R. Ravetz (1993). The Emergence of Post-Normal Science. Science, Politics, and Morality. Scientific Uncertainty and Decision Making. R. von Schomberg. Dordrecht, Kluwer.

Gaines, B. R. (1989). "Social and cognitive processes in knowledge acquisition." Knowledge Acquisition **1**(1): 39-58.

Galison, P. L. (1997). Image and logic: a material culture of microphysics. Chicago, University of Chicago Press.

Galison, P. L. and L. Daston (1992). "The Image of Objectivity." Representations **40**(Fall): 81-?

Galison, P. L. and B. Hevly, Eds. (1992). Big science: the growth of large-scale research. Stanford, CA, Stanford University Press.

Galison, P. L. and D. J. Stump (1996). The Disunity of Science: Boundaries, Contexts and Power. Stanford, Stanford University Press.

Garfinkel, H. (1967). Studies in ethnomethodology. Englewood Cliffs, NJ, Prentice Hall.

Garfinkel, H. (1991). Respecification: an announcement of studies. Ethnomethodology and the Human Sciences. G. Button. Cambridge, Cambridge University Press**:** 10-19.

Garfinkel, H., M. Lynch, et al. (1981). "The Work of Discovering Science Constructed with Materials from the Optically Discovered Pulsar." Philosophy of the Social Sciences **11**(2): 131-158.

Garfinkel, H. and H. Sacks (1970). On formal structure of practical actions. Theoretical Sociology. J. McKinney and A. Tiryakian. NY, Appleton Century Crofts.

Garfinkel, H. and L. D. Wieder (1992). Two Incommensurable, Asymmetrically Alternate Technologies of Social Analysis. Text in Context: Contributions to Ethnomethodology. G. Watson and R. M. Seiler. Newbury Park, CA, Sage.

Gaston, J. (1970). "The reward system in British science." American Sociological Review **35**: 718-32.

Geertz, C. (1973). The interpretation of cultures. New York, Basic Books.

Gershon, D. (2000). "Laying a Firm Foundation for Interdisciplinary Research Endeavors." Nature **406**(6791): 107-8.

Gibbons, M., C. Limoges, et al. (1994). The new production of knowledge. London, Sage.

Gibson, J. J. (1977). The theory of affordances. Perceiving, Acting, and Knowing. R. E. Shaw and J. Bransford. Hillsdale, NJ, Lawrence Erlbaum Associates.

Giddens, A. (1984). The Constitution of Society: Outline of the theory of structuration. Cambridge, UK, Polity Press.

Gieryn, T. F. (1983). "Boundary-Work and the Demarcation of Science From Non-Science: Strains and Interests in Professional Ideologies of Scientists." American Sociological Review **48**(6): 781-795.

Gieryn, T. F. (1999). Cultural Boundaries of Science: Credibility on the Line. Chicago, University of Chicago Press.

Glaser, B. G. (1978). Theoretical sensitivity: advances in the methodology of grounded theory. Mill Valley, CA, Sociology Press.

Glaser, B. G. and A. Strauss (1973). The discovery of grounded theory: strategies for qualitative research. Chicago, Aldine Pub. Co.

GMT. (Accessed Aug 2, 2006). "http://gmt.soest.hawaii.edu/."

Goffman, E. (1974[1986]). Frame Analysis: An Essay on the Organization of Experience. Boston, Northeastern University Press.

Granovetter, M. (1973). "The Strength of Weak Ties." American Journal of Sociology **78**: 1360.

Granovetter, M. (1983). "The Strength of Weak Ties: A Network Theory Revisited." Sociological Theory **1**: 201-233.

Gray, J., D. T. Liu, et al. (2005). "Scientific Data Management in the Coming Decade." SIGMOD Record **34**(4): 34-41.

Grinter, R. E., J. D. Herbsleb, et al. (1999). "The Geography of Coordination: Dealing with Distance in R&D Work." GROUP.

Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specification." Knowledge Acquisition **5**: 199-220.

Guarino, N. and P. Giaretta (1995). Ontologies and Knowledge Bases: towards a terminological clarification. Towards Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing. R. Dieng and O. Corby. Amsterdam, IOS Press: 25-32.

Guarino, N. and C. Welty (2000). A Formal Ontology of Properties. Knowledge Engineering and Knowledge Management: Method, Models and Tools. 12th International Conference (EKAW 2000). R. Dieng and O. Corby. Berlin/New York, Springer**:** 97-113.

Hacking, I. (1983). Representing and intervening: introductory topics in the philosophy of natural science. Cambridge Cambridgeshire; New York, Cambridge University Press.

Hacking, I. (1990). The Taming of Chance. Cambridge, UK, Cambridge University Press.

Hakala, J. and O.-H. Ylijoki (2001). "Research for Whom? Research Orientations in Three Academic Cultures." Organization **8**(2): 373-380.

Hallett, T. and M. J. Ventresca (2006). "How Institutions Form: Loose Couplings as Mechanism in Gouldner's Patterns of Industrial Bureaucracy." American Behavioral Scientist **49**(7): 908-924.

Hannan, J. and J. Freeman (1977). "The Population Ecology of Organizations." American Sociological Review **49**: 149-64.

Hanseth, O., E. Monteiro, et al. (1996). "Developing Information Infrastructure: The Tension between Standardization and Flexibility." Science, Technology & Human Values **21**(4): 407-426.

Haraway, D. (1988). "Situated Knowledges: the science question in feminism and the privilege of partial perspective." Feminist studies **14**: 575-599.

Hayles, N. K. (1999). How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics. Chicago and London, University of Chicago Press.

Hayles, N. K. (2002). Escape and Constraint: Three Fictions Dream of Moving from Energy to Information. From Energy to Information: Representation in Science and Technology, Art, and Literature. B. Clarke and L. Dalrymple. Stanford, CA, Stanford University Press.

Hicks, D. M. and S. J. Katz (1996). "Where is Science Going?" Science, Technology & Human Values **21**(4): 379-406.

Hilbert, R. A. (1990). "Ethnomethodology and the Micro-Macro Order." American Sociological Review **55**(6): 794-808.

Hilts, P. J. (1999). A law opening research data sets off debate. The New York Times**:** July 31, sec.A,p.1.,col.4.

Hirschheim, R. and M. Newman (1988). "Information Systems and User Resistance: Theory and Practice." The Computer Journal **31**: 398-408.

Howard, K. (2000). "The bioinformatics gold rush." Scientific American **283**(1): 58-63.

Hughes, T. P. (1983). Networks of Power: electrification in Western society, 1880-1930. Baltimore, John Hopkins University Press.

Hughes, T. P. (1989). The Evolution of Large Technological Systems. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. W. E. Bijker, T. P. Hughes and T. J. Pinch. Cambridge, Massachusetts and London, England, The MIT Press**:** 51-82.

Hughes, T. P. (1999). Funding a revolution: government support for computing research. Washington, D.C., National Academy Press.

Humphreys, B. and D. Lindberg (1993). "The UMLS project: making the conceptual connection between users and the information they need." Bull Med Libr Assoc. **81**(2): 170-7.

Hutchins, E. (1995). Cognition in the Wild. Massachusetts, MIT Press.

Jacko, J. A. and G. Salvendy (1996). "Hierarchical Menu Design: Breadth, Depth and Task Complexity." Perceptual Journal of Motor Skills **92**: 1187-1201.

Jaeger, E., A. Memon, et al. (2005). Integrating Seismic Events Focal Mechanisms with Image Services in Kepler. ESRI. San Diego.

Jasanoff, S. (2003). In a Constitutional Moment: Science and Social Order at the Millennium. Social Studies of Science and Technology: Looking Back, Ahead, Yearbook of the Sociology of the Sciences. B. Joerges and H. Nowotny. Dordecht, Kluwer**:** 155-180.

Jasanoff, S. (2005). Designs on Nature: Science and Democracy in Europe in the United States. Princeton, N.J., Princeton University Press.

Jeffrey, P. (2003). "Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration." Social Studies of Science **33**(4): 539-562.

Jirotka, M. and J. Goguen (1994). Requirements Engineering: Social and Technical Issues. New York, Academic Press.

Jordan and M. Lynch (1998). "The Dissemination, Standardization and Routinization of a Molecular Biological Technique." Social Studies of Science(28): 773-800.

Kammen, v. (2003). Who Represents the Users? Critical Encounters between Women's Health Advocates and Scientists in Contraceptive R&D. How Users Matter. N. Oudshoorn and T. J. Pinch. New Baskerville, The MIT Press.

Keller, R. G. (2003). "GEON (GEOScience Network) -- A first step in creating cyberinfrastructure for the geosciences." Electronic Seismologist **July/August**.

Kiger, J. L. (1984). "The Depth/Breadth Trade off in the Design of Menu-Driven User Interface." International Journal of Man Machine Studies **20**: 201-213.

Kim, J. (2006). "Motivating and Impeding Factors Affecting Faculty Contribution to Institutional Repositories." Proceedings of the Joint Conference on Digital Libraries 2006, Chapel Hill June 11-15.

Klein, J. T. (1990). Interdisciplinarity: History, Theory & Practice. Detroit, Wayne State University Press.

Klein, J. T. (1996). Crossing Boundaries:Knowledge, Disciplinarities and Interdisciplinarities. Chatlottesville, University Press Virginia.

Kline, R. (1995). "Constructing 'Technology' as 'Applied Science': Public Rhetoric of Scientists and Engineers in the United States, 1880-1945." Isis **86**: 194-221.

Kling, R. (1991). "Computerization and Social Transformations." Science, Technology & Human Values **16**(3): 342-367.

Kling, R. (1999). "What is Social Informatics and Why Does it Matter?" D-Lib Magazine **5**(1): 1-23.

Kling, R. (2000). "Learning about Information Technologies and Social Change: The Contribution of Social Informatics." The Information Society **16**: 217-232.

Kling, R. and S. Iacono (1988). "The mobilization of support for computerization: The role of computerization movements." Social Problems **35**(2): 226-243.

Kling, R. and S. Iacono (1995). Computerization Movements and the Mobilization of Support for Computerization. Ecologies of Knowledge: Work and Politics in Science and Technology. S. L. Star. Albany, NY, State University of New York Press**:** 155-199.

Kling, R. and W. Scacchi (1982). The Web of Computing: Computer Technology as Social Organization. Advances in Computers. M. C. Yovits. New York, Academic Press. **21:** 3-90.

Kling, R. and L. B. Spector (2003). Rewards for scholarly communication. Digital scholarship in the tenure, promotion, and review process. D. L. Anderson. Armonk, NY, M.E. Sharpe Inc.

Knorr-Cetina, K. (1981). The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science. Oxford, Pergamon Press.

Knorr-Cetina, K. (1999). Epistemic Cultures: How the sciences make knowledge. Cambridge, MA, Harvard University Press.

Knorr-Cetina, K. (2002). "Global Microstructures: The Virtual Societies of Financial Markets." American Journal of Sociology **107**(4).

Kolakowski, L. (1972). Positivist Science. Harmondsworth, Penguin Books.

Kratz, M., M. Ackerman, et al. (2001). "NGI and Internet2: Accelerating the Creation of Tomorrow's Internet." Studies in Health Technology and Informatics **84**(28-32).

Kuhn, T. S. (1977). Second Thoughts on Paradigms. The Essential Tension. Selected Studies in Scientific Tradition and Change. T. S. Kuhn. Chicago, Il/London, Chicago University Press.

Kwa, C. (2005). "Local Ecologies and Global Science: Discourses and Strategies of the International Geosphere-Biosphere Programme." Social Studies of Science **56**(6): 923-950.

Lakatos, I. (1976). Proofs and refutations: the logic of mathematical discovery. Cambridge, MA, Cambridge University Press.

Latour, B. (1983). Give Me a Laboratory and I Will Raise the World. Science Observed. K. Knorr-Cetina and M. Mulkay. Beverly Hills, Sage.

Latour, B. (1986). "Visualization and Cognition: Thinking with Eyes and Hands." Knowledge and Society: Studies in the Sociology of Culture Past and Present **6**: 1-40.

Latour, B. (1986). "Will the Last Person to Leave the Social Studies of Science Please Turn On the Tape-Recorder." Social Studies of Science **16**(3): 541-548.

Latour, B. (1987). Science in action: how to follow scientists and engineers through society. Cambridge, Mass., Harvard University Press.

Latour, B. (1988). The pasteurization of France. Cambridge, Mass., Harvard University Press.

Latour, B. (1991). Technology is society made durable. A Sociology of Monsters: Essays on Power, Technology and Domination. J. Law. London, UK, Routledge.

Latour, B. (1995). Foreword:The Flat-Earthers of social theory. Accounting and Science: National Inquiry and Commercial Reason. M. Power.

Latour, B. (1995). The Sociology of the Door-Closer. Ecologies of Knowledge. S. L. Star. New York, State of New York Press.

Latour, B. (1996). Aramis, or, The love of technology. Cambridge, Mass., Harvard University Press.

Latour, B. (1996). "On Interobjectivity." Mind, Culture, and Activity **3**(4): 228-245.

Latour, B. (1999). Pandora's hope: essays on the reality of science studies. Cambridge, Mass., Harvard University Press.

Latour, B. and S. Woolgar (1979). Laboratory life: the construction of scientific facts. Beverly Hills, Sage Publications.

Lave, J. (1988). Cognition in Practice: Mind, mathematics and culture in everyday life. Cambridge, Cambridge University Press.

Lave, J. and E. Wegner (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge, Cambridge University Press.

Law, J. (1994). Organizing Modernity. Oxford, Blackwell.

Law, J. and M. Callon (1988). "Engineering and Sociology in a Military Aircraft Project: A Network Analysis of Technological Change." Social Problems **35**(3): 284-297.

Lawrence, K. A. (2006 (forthcoming)). "Walking the Tightrope: The Balancing Acts of a Large e-Research Project." Computer Supported Cooperative Work.

Lee, C. P. (2006). "The First Usability Challenge for Cyberinfrastructure." CHI 2006, April 22-27, Montreal, Can. - Position Paper.

Lee, E. and J. MacGregor (1985). "Minimizing User Search Time in menu Retrieval Systems." Human Factor **27**: 157-162.

Levinson, S. (1983). Pragmatics. Cambridge, Cambridge University Press.

Livingstone, E. (1986). The ethnomethodological foundations of mathematics. London, Routledge & Kegan Paul.

Livingstone, E. (1988). Making Sense of Ethnomethodology. London, Routledge, Kegan and Paul.

Lynch, M. (1985). Art and artifact in laboratory science. Boston, MA, Routledge and Kegan Paul.

Lynch, M. (1992). Extending Wittgenstein: The Pivotal Move from Epistemology to the Sociology of Science. Science as Practice and Culture. A. Pickering. Chicago, Chicago University Press.

Lynch, M. (1993). Scientific practice and ordinary action: Ethnomethodology and social studies of science. New York, NY, Cambridge University Press.

Lynch, M. (2000). "Against reflexivity as an academic virtue and source of privileged knowledge." Theory, Culture & Society **17**(3): 26-54.

Mackay, H., C. Carne, et al. (2000). "Reconfiguring the User: Using Rapid Application Development." Social Studies of Science **30**(5): 737-57.

MacKenzie, D. (2001). Mechanizing Proof: Computing, Risk and Trust. The MIT Press, Cambridge, Massachusetts.

March, J. G. a. H. A. S. (1958). Organizations, Carnegie Institute of Technology: Graduate School of Industrial Administration.

Martin, M. (1991). "Hello Central?" Gender, Technology and the Culture in the Formation of Telephone Systems. Montreal, Que, McGill-Queens University Press.

McGinley, L. (1999). Scientists challenge provision opening access to data. The Wall Street Journal**:** March 1, Sec. A, p. 24, col.1.

Merton, R. (1942). The Normative Structure of Science. The Sociology of Science (1973). N. W. Storer. Chicago, University of Chicago Press.

Merton, R. (1968). Social Theory and Social Structure. New York, Free Press.

Metzger, N. and N. Zare (1999). "Interdisciplinary Research: From Belief to Reality." Science **1999**(5402): 642-43.

Meyer, A. M. (1992). "How to Apply the Anthropological Technique of Participant Observation to Knowledge Acquisition for Expert Systems." IEEE Transactions on Systems, Man, and Cybernetics **22**(5): 983-990.

Meyer, J. and B. Rowan (1977). "Institutionalized Organizations: formal structure as myth and ceremony." American Journal of Sociology **83**(340-63).

Meyer, M. A. and J. M. Booker (2001). Eliciting and Analyzing Expert Judgment: A Practical Guide. Philadelphia, PA, Society of Industrial and Applied Mathematics.

Miles, I. and G. Thomas (1995). User resistance to new interactive media: Participants, processes and paradigms. Resistance to New Technology. M. Bauer, Cambridge University Press.

Mill, J. S. (1843/1874). A System of Logic. Ratiocinative and Inductive.Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. New York, Harper and Brothers Publishers.

Millerand, F. and G. C. Bowker (forthcoming). "Metadata" et cognition sociale. Sciences sociales et cognition. C. Rosental. Paris, Editions de l'Ecole des Hautes Etudes en Sciences Sociales (coll. Enquete).

Millerand, F. and D. Ribes (forthcoming). "Making Metadata a Public Issue." The Information Society.

Mills, C. W. (1959). The Sociological Imagination. New York, Oxford University Press.

Mintzberg, H. (1992). Structure in Fives: Designing Effective Organizations. Englewood Cliffs, N.J, Prentice-Hall.

Mintzberg, H. and A. McHugh (1985). "Strategy formation in an Adhocracy." Administrative Science Quarterly **30**(160-97).

Moerman, M. (1988). Talking culture: Ethnography and conversation analysis. Philadelphia, University of Pennsylvania Press.

Monteiro, E. and O. Hanseth (1997). "Inscribing behaviour in information infrastructure standards." Science, Technology & Human Values **21**(4): 407-426.

Montgomery, J. D. (1991). "Social Networks an Labor Market Outcomes: Towards an Economic Analysis." American Economic Review **81**: 1408-1418.

Mukerji, C. (1989). A Fragile Power: Scientists and the State. Princeton, Princeton University.

Nelson, R. R., Ed. (1992). National innovation systems. Oxford, Oxford University Press.

Nelson, R. R. and S. G. Winter (1982). An evolutionary theory of economic change. Cambridge, MA, Belknap Press of Harvard University.

Norman, D. A. (1988). The psychology of everyday things. New York, Basic Books.

NSF (2001). Information Technology Research (ITR) Program Solicitation. NSF 01-149.

Oldow, J. S., X. Xu, et al. (2003). "Ground-based LIDAR imaging and surface analysis of Late Quaternary wave-cut terraces and high-angle faults: Constraints on active deformation in the Alvord Desert, southeastern Oregon." Geological Society of America Abstracts with Program **34**: 445.

Olson, G. M. and J. S. Olson (2000). "Distance Matters." Human-Computer Interaction **15**(2-3): 139-179.

Oost, E. v. (2003). Materialized Gender: How Shavers Configure the Users' Femininity and Masculinity. How Users Matter. N. Oudshoorn and T. J. Pinch. New Baskerville, The MIT Press.

Orlikowski, W. (1996). "Improvising organizational transformation over time: A situated change perspective." Information Systems Research **7**: 63-92.

Orr, J. E. (1996). Talking About Machines: An Ethnography of a Modern Job. Ithaca, Cornell University Press.

Orton, D. J. and K. E. Weick (1990). "Loosely coupled systems: A reconceptualization." The Academy of Management Review **15**(2): 203-223.

Oudshoorn, N. and T. J. Pinch (2003). How Users Matter: The Co-Construction of Users and Technology. New Baskerville, The MIT Press.

Oudshoorn, N. and T. J. Pinch (2003). Introduction: How Users and Non-Users Matter. How Users Matter: The Co-Construction of Users and Technology. N. Oudshoorn and T. J. Pinch. New Baskerville, The MIT Press.

Papadopoulos, P. M., M. J. Katz, et al. (2001). "NPACI Rocks: tools and techniques for easily deploying manageable Linux Clusters." Concurrency and Computation: Practice and Experience (Special Issue 'Cluster 2001') **15**(7-8): 707-725.

Pinch, T. J. and W. E. Bijker (1984). "The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other." Social Studies of Science **14**(1): 399-441.

Polanyi, M. (1964). Personal Knowledge: Towards a Post-Critical Philosophy. New York, Harper Press.

Popper, K. R. (1965). The Logic of Scientific Discovery. New York, Harper and Row.

Porter, T. M. (1995). Trust in numbers: the pursuit of objectivity in science and public life. Princeton, N.J., Princeton University Press.

Prentice, R. (2005). "The Anatomy of a Surgical Simulation: The Mutual Articulation of Bodies in and through the Machine." Social Studies of Science **35**(6): 837-866.

Price, D. J. (1963). Little Science, Big Science. New York, Columbia University Press.

Rabinow, P. (1977). Reflections on Fieldwork in Morocco. Berkeley, CA, University of California Press.

Reskin, B. F. (1977). "Scientific Productivity and the Reward Structure of Science." American Sociological Review **42**(3): 491-504.

Rheinberger, H.-J. (2000). Cystoplasmic Particles: The Trajectory of a Scientific Object. Biographies of Scientific Objects. L. Daston. Chicago, University of Chicago Press.

Rhoten, D. (2003). A Multi-Method Analysis of the Social and Technical Conditions for Interdisciplinary Collaboration, National Science Foundation BCS-0129573.

Rhoten, D. (2004). "Interdisciplinary Research: Trend or Transition." Items and Issues **5**(1-2): 6-11.

Ribes, D. (forthcoming). GEON: Organizing for Interdisciplinary Collaboration. The Collaboratories Handbook. G. M. Olson, J. S. Olson and A. Zimmerman. Cambridge, MIT Press.

Roberts, T. L., M. Wermus, et al. (2003). "The Career of the Technical Professional: A Look at the Career Path and Reward System in Technical Polish Companies." Proceedings of the Hawaii International Conference for System Science, HICSS.

Roland, A. and P. Shiman (2002). Strategic Computing: DARPA and he quest for machine intelligence, 1983-1993. Cambridge, MA, MIT Press.

Rosenberg, N. (1994). Exploring the black box: technology, economics, and history. Cambridge, UK, Cambridge University Press.

Rossini, F. A. and A. L. Porter (1984). Interdisciplinary Research: Performance and Policy Issues. Problems in Interdisciplinary Studies. R. Jurkovich and J. H. P. Paelinch. Aldershot, Gover.

Rudwick, M. (1976). "The Emergence of a Visual Language for Geological Science, 1760-1840." History of Science **14**: 149-95.

Schegloff, E. A. (1987). "Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis." Social Psychology Quarterly **50**(2): 101-114.

Schmidt, K. (2000). The Critical Role of Workplace Studies in CSCW. Workplace Studies: Recovering Work Practice and Informing Design. P. Luff, J. Hindmarsh and C. Heath. Cambridge, UK, Cambridge University Press**:** 141-149.

Schuler, D. and A. Namioka (1993). Participatory Design: Principles and Practices. New Jersey, Lawrence Erlbaum Associates.

Scott, P., E. Richards, et al. (1990). "Captives of controversy: The myth of the neutral social researcher in contemporary scientific controversies." Science, Technology & Human Values **15**: 474-94.

Scott, W. R. (1992). Organizations: Rational, Natural and Open. Englewood Cliffs, NJ, Prentice-Hall.

Seely Brown, J. and P. Duguid (1991). "Organizational learning and communities of practice: Toward a unified view of working, learning and innovation." Organization Science **2**(1): 40-57.

Seppala, P. and G. Salvendy (1995). "Impact of Depth of Menu Hierarchy on Performance Effectiveness in a Supervisory Task: Computer Flexible Manufacturing System." Human Factor **27**: 713-722.

Serres, M. (1987). Statues. Paris, Francois Bourin.

Service, R. F. and D. Malakoff (1999). "Gore Presents Plan to Spend $366 Million." Science **283**(5402): 613-614.

Shalin, D. N. (1986). "Pragmatism and Social Interactionism." American Sociological Review **51**: 9-29.

Shapin, S. (1989). "The Invisible Technician." American Scientist **77**(Nov/Dec): 554-563.

Shapin, S. and S. Schaffer (1985). Leviathan and the air-pump: Hobbes, Boyle and the Experimental Life. Princeton, N.J., Princeton University Press.

Sheth, A. P. (1999). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. Interoperating Geographic Information Systems. E. Goodchild, Fegeas, Kottman. Boston, Kluwer Academic Publishers.

Sieber, J. E. and B. E. Trumbo (1995). "(Not) giving credit where credit is due: Citation of data sets." Science and Engineering Ethics **1**(1): 11-20.

Slaughter, M. M. (1982). Universal Languages and Scientific Taxonomy in the Seventeenth Century. Cambridge, UK, Cambridge University Press.

Slaughter, S. (1993). "Beyond basic science: Research university president's narratives of science policy." Science, Technology & Human Values(18): 278-302.

Slaughter, S. and G. Rhoades (1996). "The emergence of a competitiveness research and development policy coalition and the commercialization of academic science and technology." Science, Technology & Human Values **21**(3): 303-39.

Smith, B. (2003). Ontology. Blackwell Guide to the Philosophy of Computing and Information. L. Floridi. Oxford, Blackwell**:** 155-166.

Smith, B. and C. Welty (2001). "Ontology: Towards a New Synthesis." Published in the Proceedings of FOIS '01.

Sowa, J. (2000). Knowledge representation: logical, philosophical, and computational foundations. Boston, MA, MIT Press.

Star, S. L. (1988). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. Readings in Artificial Intelligence. M. Huhns and L. Gasser. Menlo Park, California, Morgan Kaufman.

Star, S. L. (1989). Regions of the mind: brain research and the quest for scientific certainty. Stanford, Calif., Stanford University Press.

Star, S. L. (1991). Invisible Work and Silenced Dialogues in Knowledge Representation. Women, Work and Computerization. I. V. Eriksson, B. A. Kitchenham and K. G. Tijdens, Elsevier Science Publishers B.V. (North-Holland).

Star, S. L. (1991). Power, Technology, and the phenomenology of conventions: On being allergic to onions. A Sociology of Monsters. J. Law, Routledge.

Star, S. L. (1992). "The Trojan Door: Organizations, Work, and the "Open Black Box"." Systems Practice **5**(4): 395-410.

Star, S. L. (1995). Ecologies of Knowledge: Work and Politics in Science and Technology. Albany, New York, SUNY Press.

Star, S. L. (1999). "The Ethnography of Infrastructure." American Behavioral Scientist **43**: 377-391.

Star, S. L. and J. R. Griesemer (1989). "Institutional Ecology, "Translations," and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." Social Studies of Science **19**: 387-420.

Star, S. L. and K. Ruhleder. (1994). Steps Towards and Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. NC, USA, Chapel Hill.

Stefik, M. and L. Conway (1982). "Towards a Principled Engineering of Knowledge." AI Magazine **3**(3): 4-16.

Stene, E. (1940). "An approach to the science of administration." American Political Science Review **34**: 1124-1137.

Sterling, T. D. (1988). "Analysis and re-analysis of shared scientific data." Annals of the American Academy of Political Social Science **495**: 49-60.

Stokes, D. E. (1997). Pasteur's quadrant: Basic science and technological innovation. Washington, D.C., Brookings Institution Press.

Strathern, M. (1995). The Relation: Issues in Complexity and Scale. Cambridge, UK, Prickly Pear Press.

Strathern, M. (2004). Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge. Wantage, UK, Sean Kingston Publishing.

Strauss, A. (1988). "The Articulation of Project Work: An Organizational Process." The Sociological Quarterly **29**: 163-78.

Strauss, A. (1993). Continual Permutations of Action. New York, Aldine de Gruyter.

Strum, S. S. and B. Latour (1987). "Redefining the social link: from baboons to humans." Social Science Information **26**(4): 783-802.

Suchman, L. (1987). Plans and Situated Actions. Cambridge, Cambridge University Press.

Suchman, L. (1994). "Working Relations of Technology Production and Use." Computer Supported Cooperative Work **2**: 21-39.

Taylor, F. (1995). "Shifting from a heteronomous to an autonomous worldview of organizational communication: Communication theory on the cusp." Communication Theory **15**(3): 1-35.

Taylor, J. R. and F. Cooren (1997). "What makes communication "organizational"? How the many voices of the organization become the one voice of an organization." Journal of Pragmatics **27**(4): 409-438.

Taylor, J. R. and E. J. Van Every (2000). The Emergent Organization: Communication as Its Site and Surface. Mahwah, NJ, Lawrence Erlbaum Associates.

Teasley, S. and S. Wolensky (2001). "Scientific collaborations at a distance." Science **292**: 2254-2255.

Thompson, J. (1967). Organizations in Action. New York, McGraw-Hill.

Traweek, S. (1992). Beamtimes and lifetimes: the world of high energy physicists. Cambridge, Mass., Harvard University Press.

Trigg, R. (1994). From Implementation to Design: Tailoring and the Emergence of Systematization in CSCW. Proceedings of ACM 1994 Conference on Computer-Supported Cooperative Work. New York, ACM Press.

Van House, N. A., M. H. Butler, et al. (1998). Cooperative work and practices of trust: Sharing environmental planning data sets. CSCW '98: ACM conference on computer supported cooperative work, Nov14-18, Seattle, WA, New York: ACM.

Vann, K. and G. C. Bowker (2001). "Instrumentalizing the Truth of Practice." Social Epistemology **15**(3): 247-262.

Vaughan, D. (1999). "The Role of the Organization in the Production of Techno-Scientific Knowledge." Social Studies of Science **29**(6): 913-943.

Verran, H. (2001). Science and An African Logic. Chicago and London, The University of Chicago Press.

Waterman, D. A. (1986). A Guide to Expert Systems. Reading, MA, Addison-Wesley.

Weedman, J. (1998). "The Structure of Incentive: Design and Client Roles in Application-Oriented Research." Science, Technology & Human Values **23**(3): 315-345.

Weick, K. (1969). The Social Psychology of Organizing. Reading, Mass, Addison-Wesley.

Weick, K. E. (1976). "Educational organizations as loosely coupled systems." Administrative Science Quarterly **21**: 1-19.

Weick, K. E. (1977). Enactment processes in organizations. New directions in organizational behavior. B. M. Staw and G. Salancik. Chicago, St. Clair**:** 267-300.

Weick, K. E. (1997). Enactment processes in organizations. New directions in organizaitonal behavior. B. M. Staw and G. Salancik. Chicago, St. Clair**:** 267-300.

Weick, K. E., K. M. Sutcliffe, et al. (2005). "Organizing and the Process of Sensemaking." Organization Science **16**(4): 409-421.

Weingart, P. (1997). "From 'Finalization' to 'Mode 2': Old Wine in New Bottles?" Social Science Information **36**: 591-613.

Weiss, H. M. and D. R. Ilgen (1985). "Routinized behavior in organizations." Journal of Behavioral Economics **14**: 57-67.

Weller, A. C. (1996). The human genome project. From print to electronic: The transformation of scholarly communication. S. Y. Crawford, J. M. Hurd and A. C. Weller. Medford, NJ, Information Today**:** 35-64.

Welsh, E., M. Jirotka, et al. (2006). "Post-Genomic science: cross-disciplinary and large-scale collaborative research and its organizational and technological challenges for the scientific research process." Philos Transact A Math Phys Eng Sci **364**(1843): 1533-49.

Winkler, M. G. and A. Van Helden (1992). "Representing the Heavens: Galileo and Visual Astronomy." Isis **83**: 195-217.

Winner, L. (1984). "Myth information in the high-tech era." IEEE Spectrum **21**: 90-6.

Woolgar, S. (1991). Configuring the user: The case of usability trials. A Sociology of Monsters. J. Law, Routledge.

Yates, J. (1989). Control through Communication: The Rise of System in American Management. Baltimore, The John Hopkins University Press.

Zammito, J. H. (2004). A Nice Derangement of Epistemes: Post-Positivism in the Study of Science from Quine to Latour. Chicago, University of Chicago Press.

Zare, R. N. (1997). "Editorial: Knowledge and Distributed Intelligence." Science **275**(5303): 1047.

Zaslavsky, I., C. Baru, et al. (2005). Grid-enabled mediation services for geospatial information. Next Generation Geospatial Information from Digital Image Analysis to Spatiotemporal Databases. P. Agouris and A. Croitoru, Taylor & Francis.

Zaslavsky, I., H. He, et al. (2004). "Integrating Brain Data Spatially: Spatial Data Infrastructure and Atlas Environment for Online Federation and Analysis of Brain Images." DEXA Workshops 2004:389-393.

Ziman, J. (1981). "What are the Options? Social Determinants of Personal Research Plans." Minerva **19**(1): 1-42.