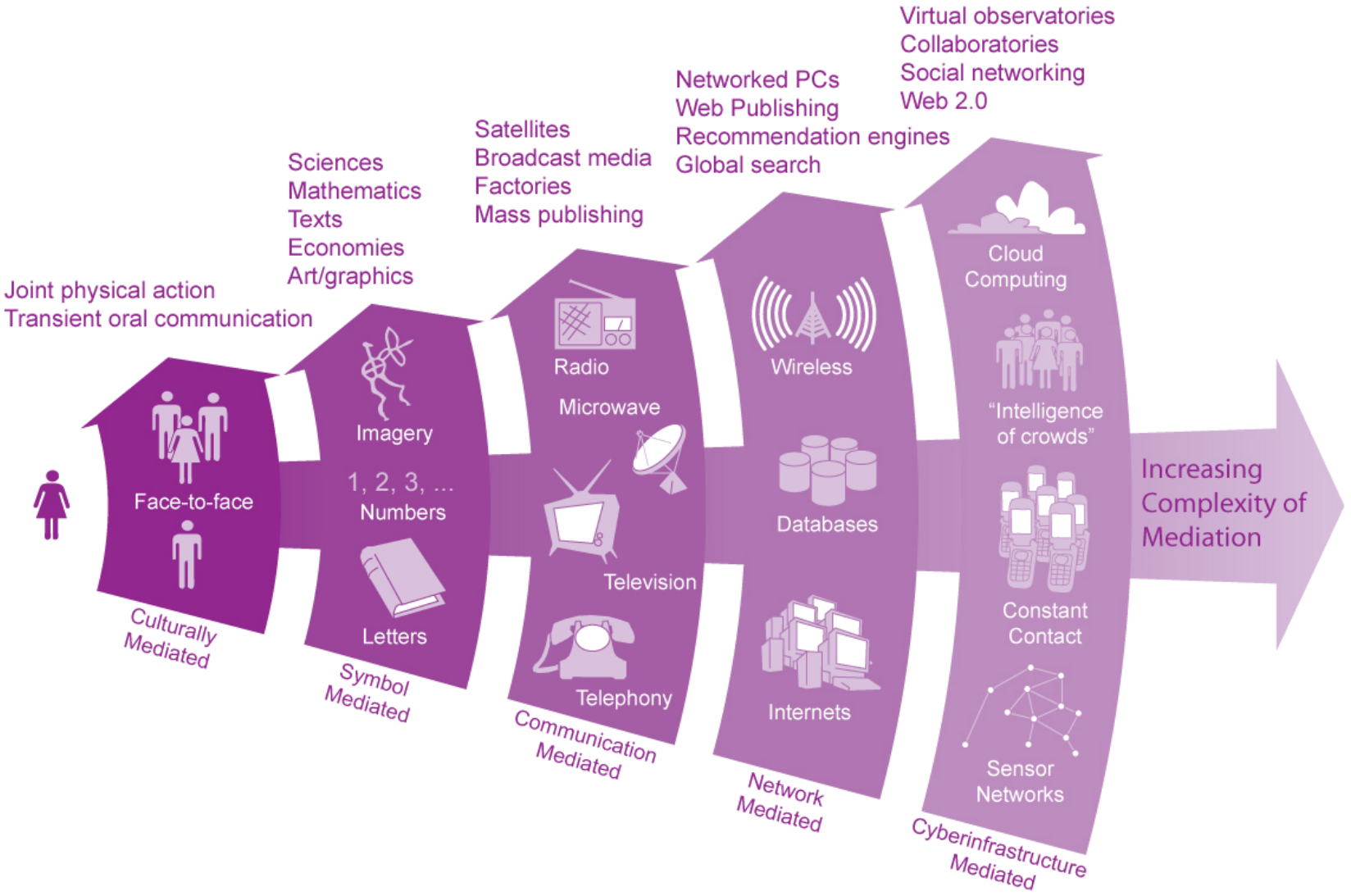
Big Data, Little Data, No Data: Scholarship in the Networked World

Christine L. Borgman
Professor and Presidential Chair in Information Studies
University of California, Los Angeles

3–6 February 2014
Melbourne Convention & Exhibition Centre
streaming with possibilities

Technological advances in mediated communication



Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., … Szalay, A. (2008). *Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge.* National Science Foundation. http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf08204

Image: Jillian Wallis and Roy Pea

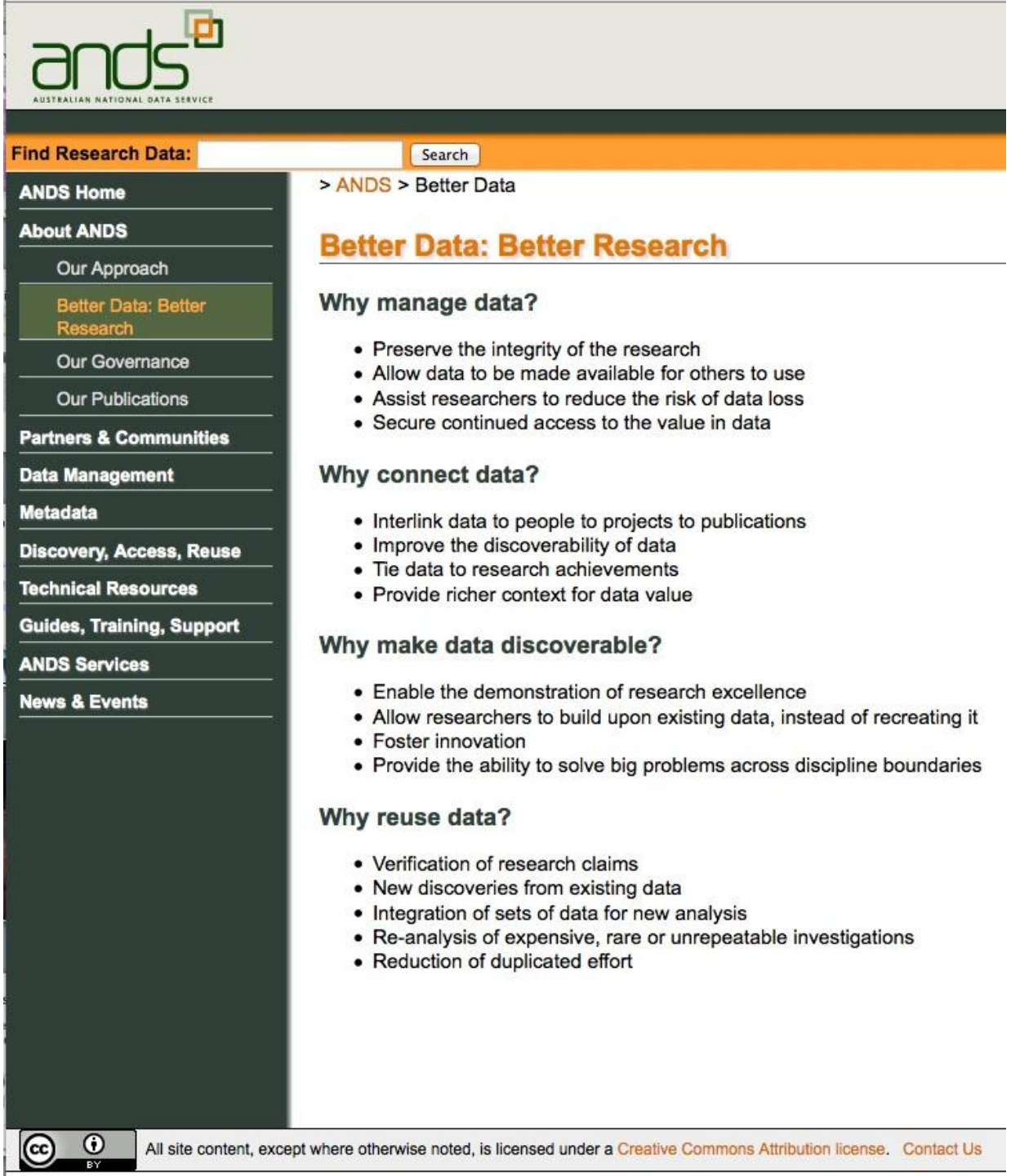

Science
12 September 2008 | $10
AAA


Science
11 February 2011 | $10
scientific
data
information
climate
research
science
new
analysis
visualization
researchers
access
many
human
knowledge
example
AAAS

Neelie Kroes, VP European Commission:

To collect, curate, preserve and make available ever-increasing amounts of scientific data, new types of infrastructures will be needed. The potential benefits are enormous but the same is true for the costs. We therefore need to lay the right foundations and the sooner we start the better.

Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D. R., … Hudson, R. L. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf


ands
AUSTRALIAN NATIONAL DATA SERVICE
Find Research Data:
Search
ANDS Home
About ANDS
Our Approach
Better Data: Better Research
Our Governance
Our Publications
Partners & Communities
Data Management
Metadata
Discovery, Access, Reuse
Technical Resources
Guides, Training, Support
ANDS Services
News & Events
> ANDS > Better Data
Better Data: Better Research
Why manage data?
• Preserve the integrity of the research
• Allow data to be made available for others to use
• Assist researchers to reduce the risk of data loss
• Secure continued access to the value in data
Why connect data?
• Interlink data to people to projects to publications
• Improve the discoverability of data
• Tie data to research achievements
• Provide richer context for data value
Why make data discoverable?
• Enable the demonstration of research excellence
• Allow researchers to build upon existing data, instead of recreating it
• Foster innovation
• Provide the ability to solve big problems across discipline boundaries
Why reuse data?
• Verification of research claims
• New discoveries from existing data
• Integration of sets of data for new analysis
• Re-analysis of expensive, rare or unrepeatable investigations
• Reduction of duplicated effort
All site content, except where otherwise noted, is licensed under a Creative Commons Attribution license. Contact Us

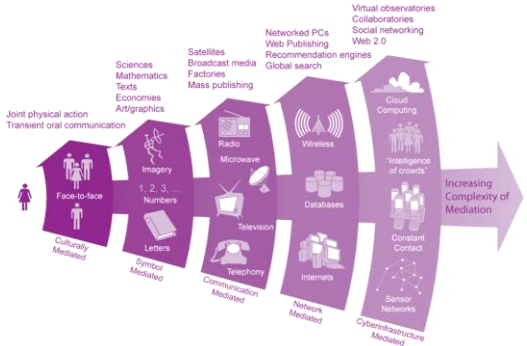
Open access policies

- Australian Research Council
  - Code for the Responsible Conduct of Research
  - Data management plans
- National Science Foundation
  - Data sharing requirements
  - Data management plans
- U.S. Federal policy-2013
  - Open access to publications
  - Open access to data
- European Union
  - European Open Data Challenge
  - OpenAIRE
- Research Councils of the UK
  - Open access publishing
  - Provisions for access to data

Supported by wellcome trust

Big Data, Little Data, No Data: Scholarship in the Networked World*

- Section I: Data and Scholarship
  - Ch 1: Provocations
  - Ch 2: What are Data?
  - Ch 3: Data Scholarship
  - Ch 4: Data Diversity
- Section II: Case Studies in Data Scholarship
  - Ch 5: Data Scholarship in the Sciences
  - Ch 6: Data Scholarship in the Social Sciences
  - Ch 7: Data Scholarship in the Humanities
- Section III Data Policy and Practice
  - Ch 8: Releasing, Sharing, and Reusing Data
  - Ch 9: Credit, Attribution, and Discovery
  - Ch 10: What to Keep and Why



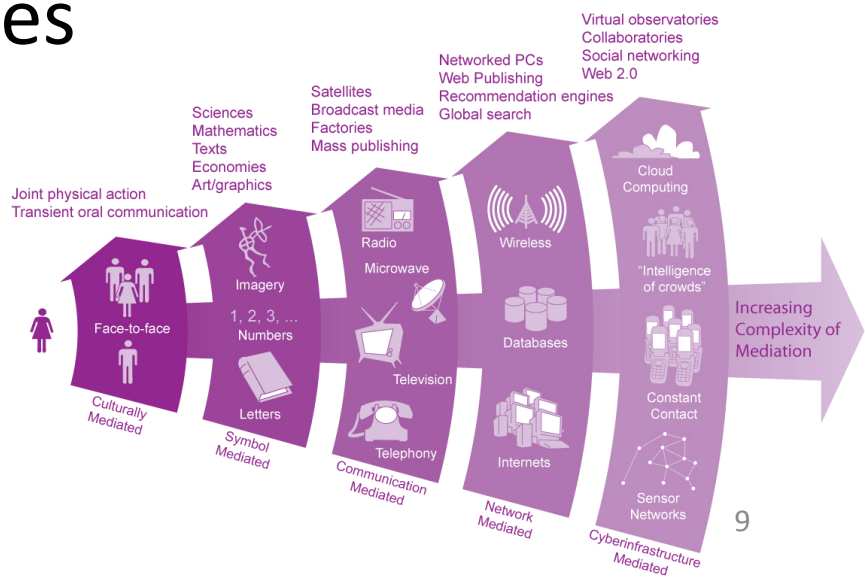
*Borgman, C.L. (2014) MIT Press

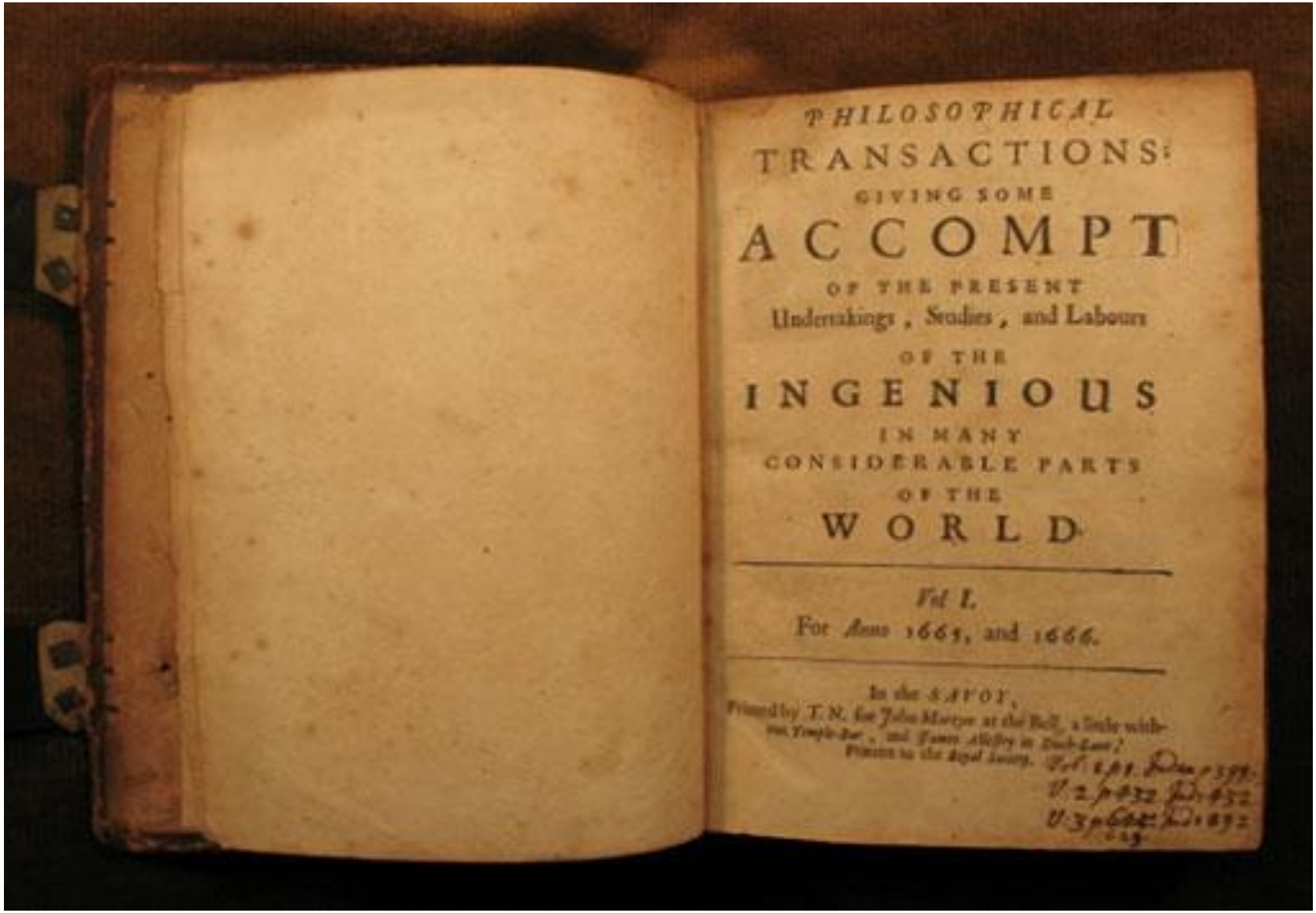
Provocations

1. How do rights, responsibilities, and risks associated with research data vary by discipline and by stakeholder?
2. How can data be exchanged across domains, contexts, and over time?
3. How do publications and data differ?
4. What are scholars' motivations to share their data?
5. What kinds of expertise are required to manage, curate, and sustain access to research data?
6. How can knowledge infrastructures adapt to the needs of generations of scholars and the demands of competing stakeholders?

Libraries, librarians, and data

1. Data are not publications
2. Data are not natural objects
3. Data are representations
4. Data sharing and reuse depend on knowledge infrastructures

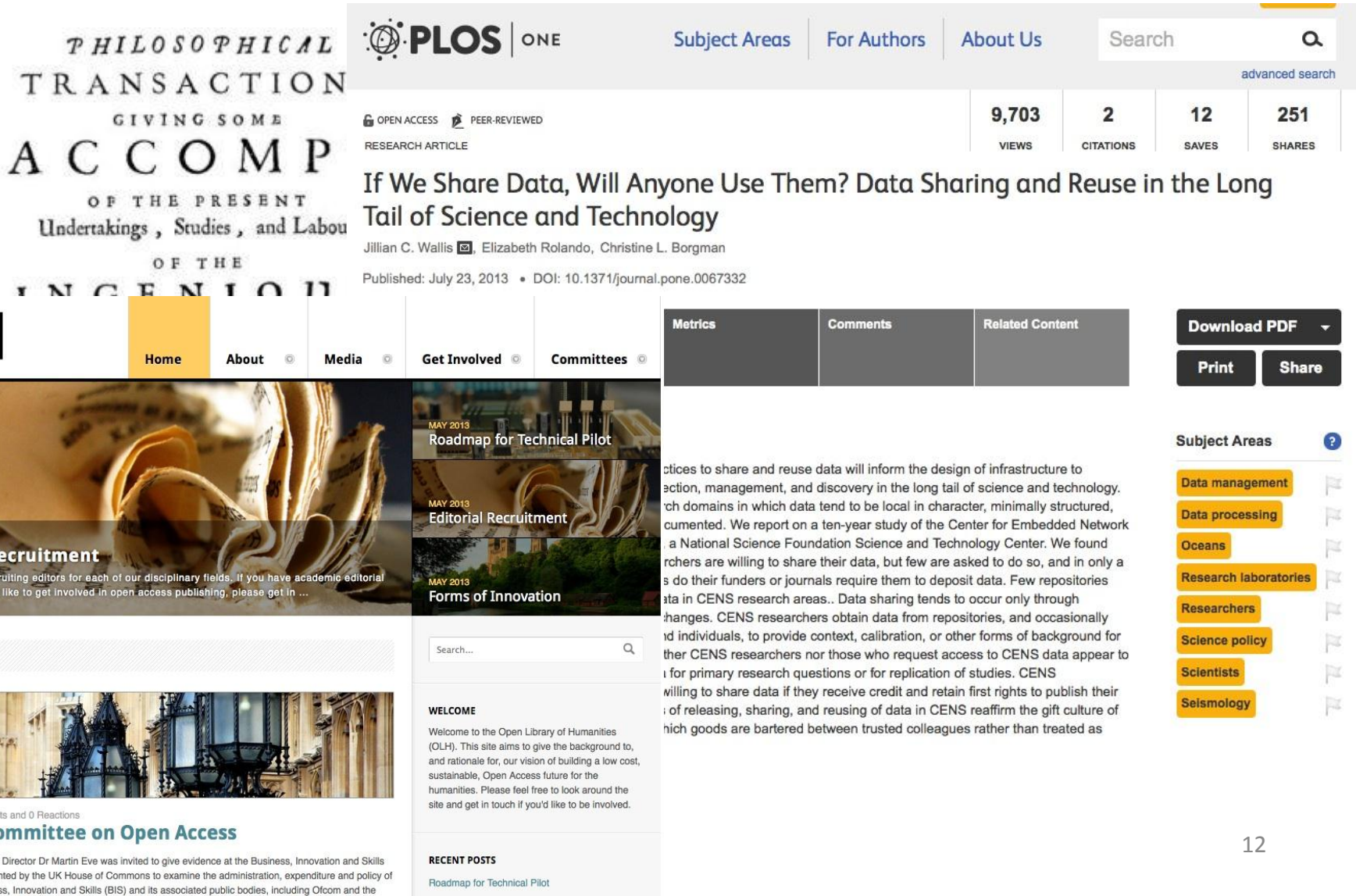



PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD
Vol I.
For Anno 1665, and 1666.
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little within Temple-Bar, and Henry Allestry in Fleet-street: Printers to the Royal Society.

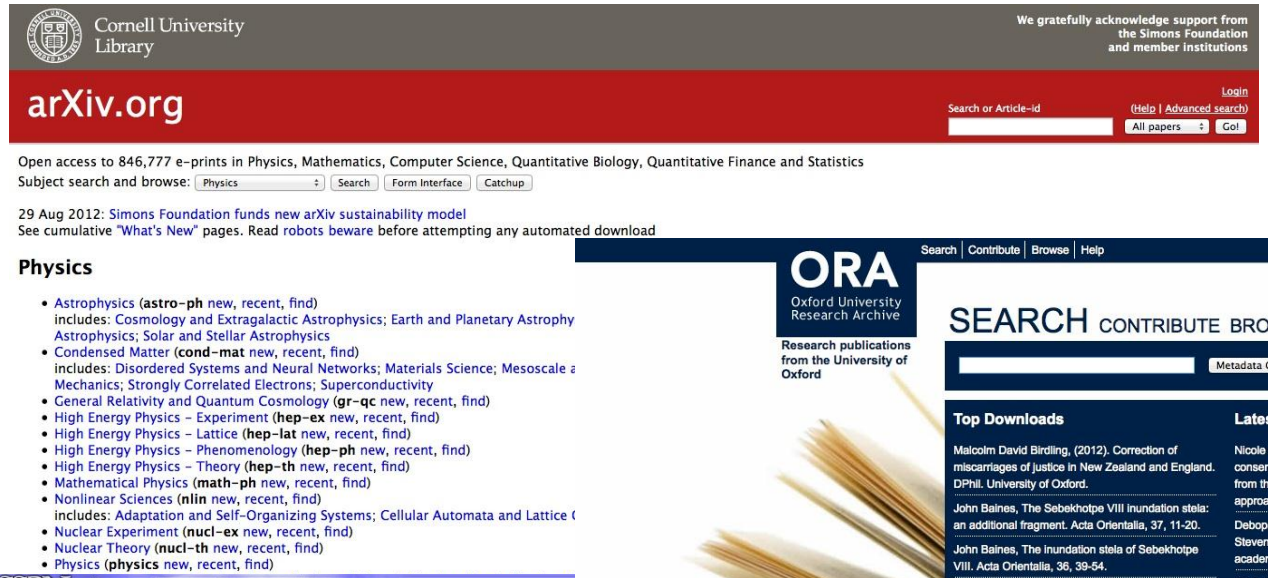
http://www.ansp.org/~/media/Images/ans/explore/online-exhibits/stories/philosophical-transactions-v1_510.ashx

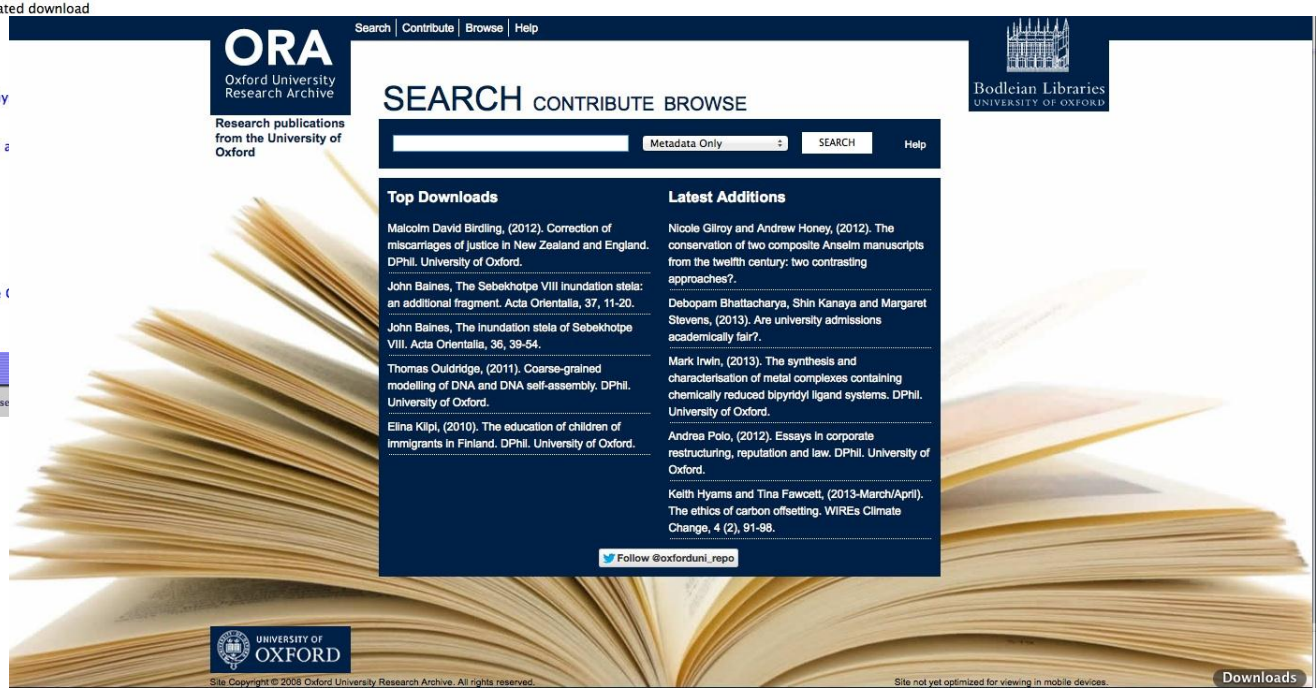
Brick inscribed with the Sutra on Dependent Origination *Gorakhpur district, late 5th century - early 6th century AD. Ashmolean Museum*

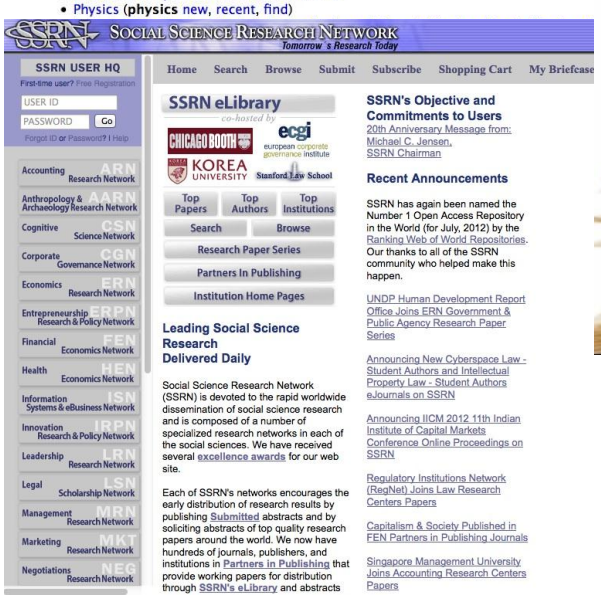
Open access publishing



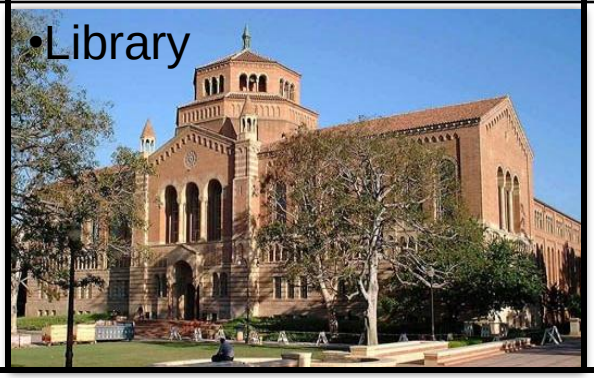
Open dissemination



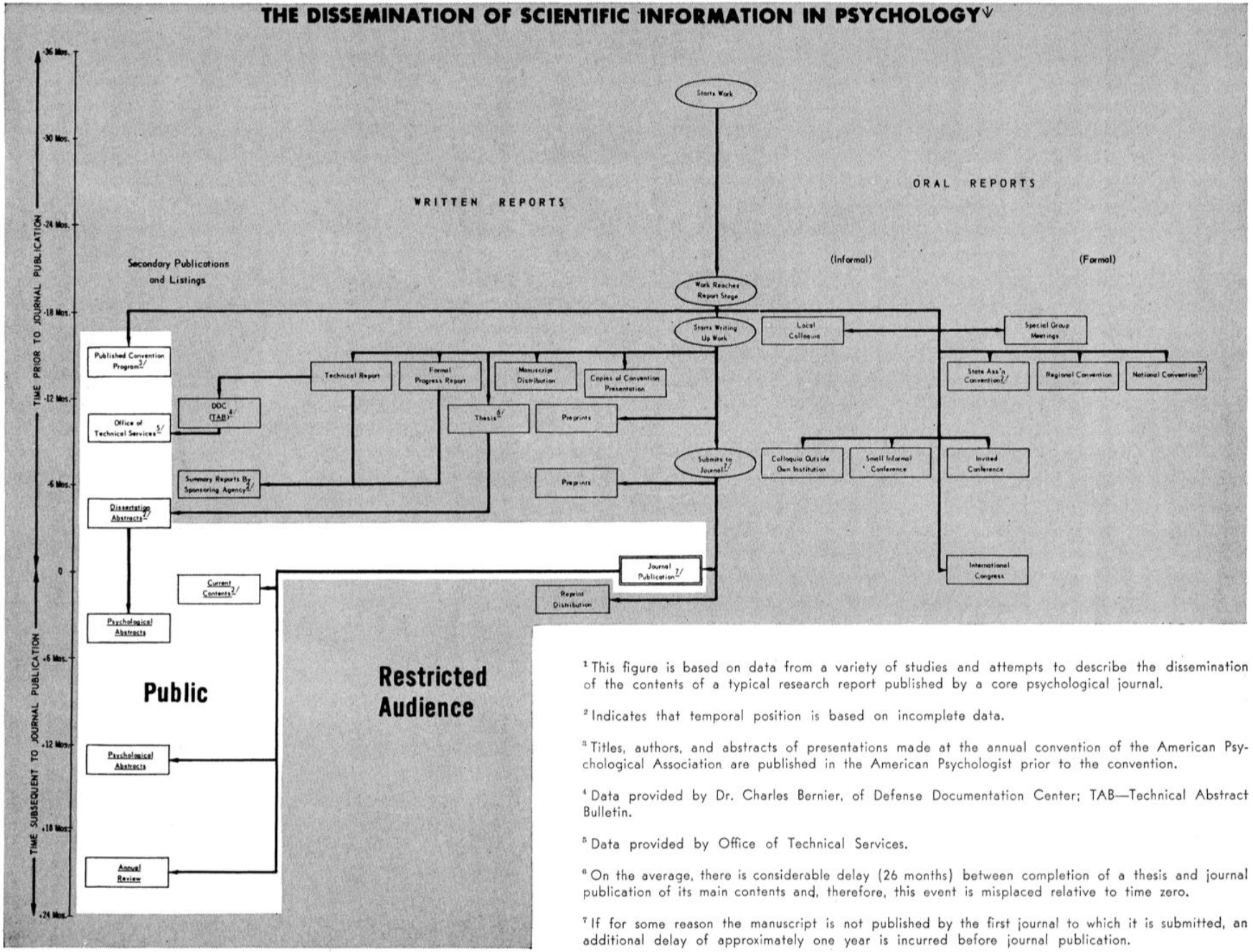




Publications in Scholarly Communication

Function	Print	Digital
Legitimization Authority, quality, priority, trustworthiness	•Peer review	•Peer review
Dissemination Awareness, diffusion, publicity	•Publisher •Pre-print distribution •Copy •Mail	•Publisher •Pre-print distribution •Post on Web •Deposit
Access, preservation, curation Availability, discovery, retrieval, persistence	•Library 	•Library •Publisher •Repository •Homepage

Borgman, C.L. (2007), MIT Press, Scholarship in the Digital Age



[1] This figure is based on data from a variety of studies and attempts to describe the dissemination of the contents of a typical research report published by a core psychological journal.

[2] Indicates that temporal position is based on incomplete data.

[3] Titles, authors, and abstracts of presentations made at the annual convention of the American Psychological Association are published in the American Psychologist prior to the convention.

[4] Data provided by Dr. Charles Bernier, of Defense Documentation Center; TAB—Technical Abstract Bulletin.

[5] Data provided by Office of Technical Services.

[6] On the average, there is considerable delay (26 months) between completion of a thesis and journal publication of its main contents and, therefore, this event is misplaced relative to time zero.

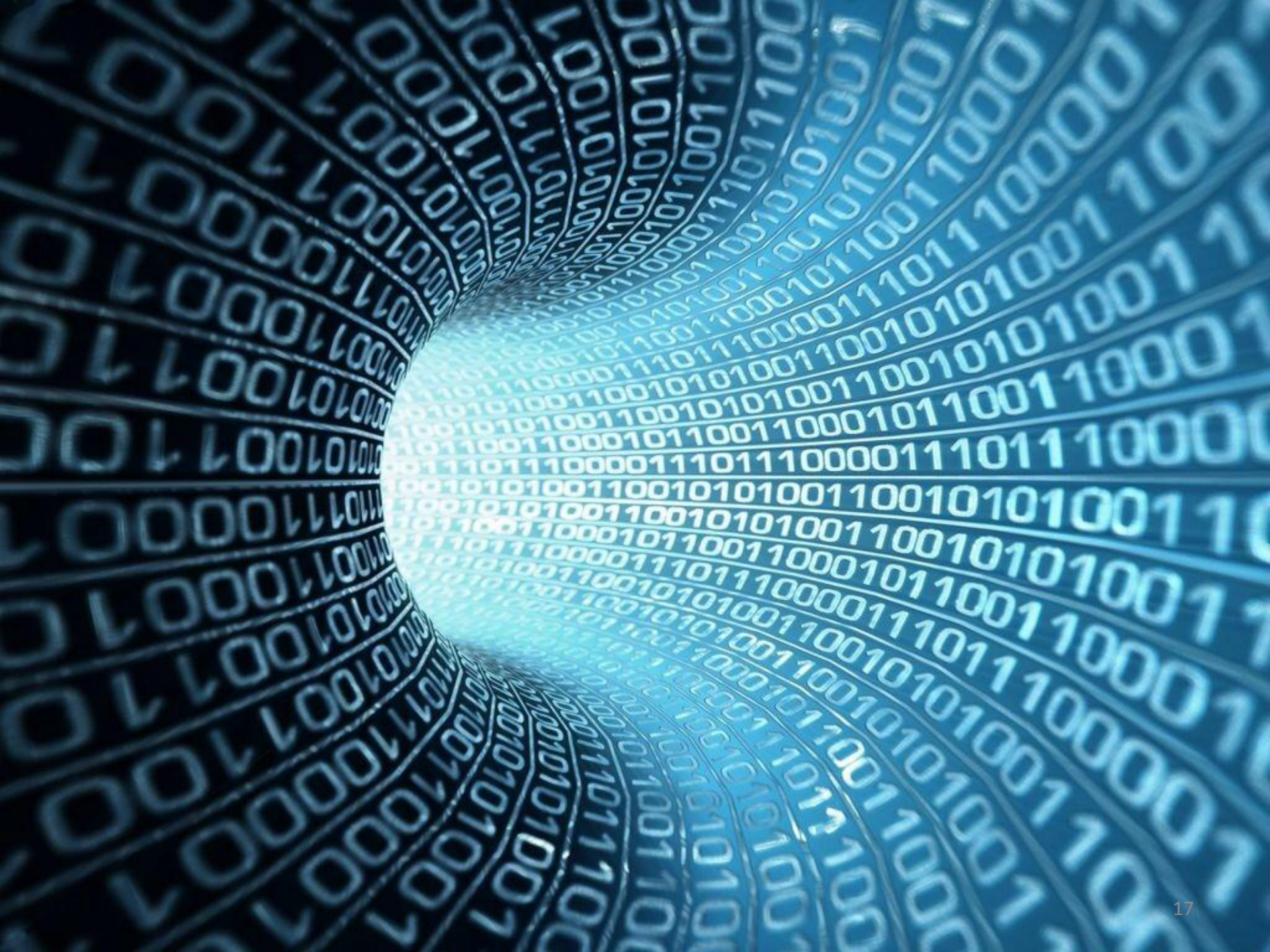
[7] If for some reason the manuscript is not published by the first journal to which it is submitted, an additional delay of approximately one year is incurred before journal publication.

Garvey, W. D., & Griffith, B. C. (1964). Scientific information exchange in psychology. *Science*, *146*, 1655–1659. Figure redrawn by J.C. Wallis for Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.

Open access to scholarly literature

- Open access literature is digital, online, free of charge, and free of most copyright and licensing restrictions.*
- Principles
  - Scholarly authors hold the copyright on their work
  - Scholarly authors write for impact, not revenue

*Suber, P. (2012). *Open Access*. Cambridge MA: MIT Press. http://mitpress.mit.edu/books/open-access



Australian National Data Service

Our Vision: More Australian researchers reusing research data more often

ANDS is enabling the transformation of:

Data that are:	to	Structured Collections that are:
Unmanaged	→	Managed
Disconnected	→	Connected
Invisible	→	Findable
Single-use	→	Reusable

More>>

Australian Research Data Commons

ANDS is building the Australian Research Data Commons: a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs.

Research Data Australia

Research Data Australia, ANDS' flagship service, provides a comprehensive window into the Australian Research Data Commons. It is an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines.

Search for research data on researchdata.ands.org.au

Open Data

- A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike (Open Data Commons, 2013).
- Data that meets the criteria of intelligent openness. Data must be accessible, useable, assessable and intelligible. (Royal Society, 2012)
- Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, and sustainability. *OECD Principles and Guidelines for Access to Research Data from Public Funding* (2007)

OA publications vs. data

- Open access to literature principles (Suber, 2012)
  - Scholarly authors hold the copyright on their work
  - Scholarly authors write for impact, not revenue

- Open access to data
  - Rights in data are unclear and often contested
  - Papers are arguments for which data are evidence

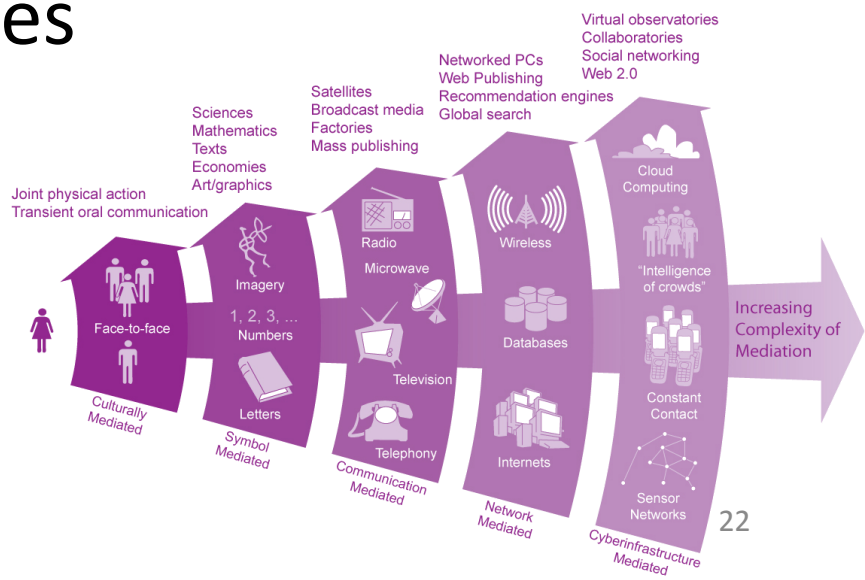
Data in Scholarly Communication

Function	Reported in a publication	Contributed to a data repository
Legitimization Authority, quality, priority, trustworthiness	•Peer review in context •Quality of method •Evidence for conclusions •Verify, reanalyze? •Author reputation	•Peer review •Quality of metadata, documentation •"test drive" the data
Dissemination Awareness, diffusion, publicity	•Description in a publication	•Repository publisher
Access, preservation, curation Availability, discovery, retrieval, persistence	•Request to author •Author maintains own data •Author directs requestor to data source	•Repository •Immediate access •Embargo period •Curation responsibility

Borgman, C.L. (2007), MIT Press, Scholarship in the Digital Age

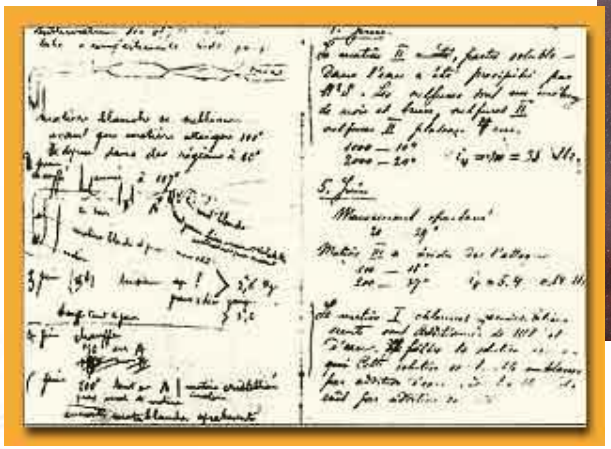
Libraries, librarians, and data

1. Data are not publications
2. Data are not natural objects
3. Data are representations
4. Data sharing and reuse depend on knowledge infrastructures



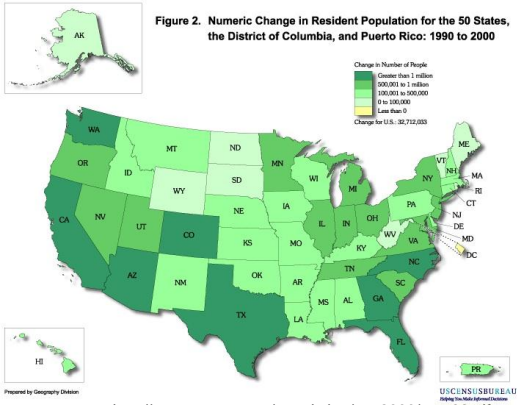
What are data?

hudsonalpha.org

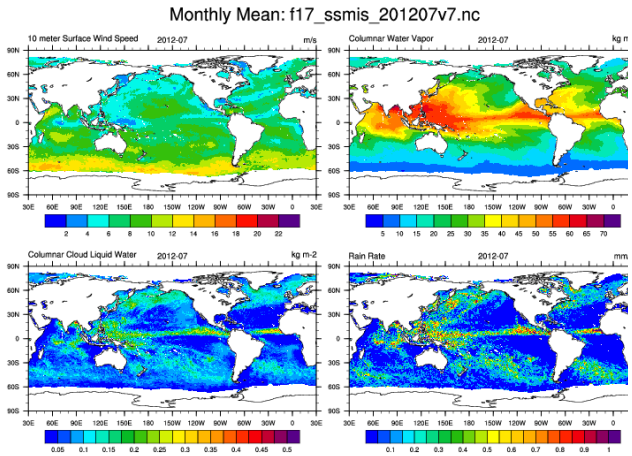


Marie Curie's notebook aip.org

NASA Astronomy Picture of the Day



http://www.census.gov/population/cen2000/map02.gif



ncl.ucar.edu

Date:1/2.07.75 Place:Sakaltutan
Zafor
He will grow old in his present house; new house is for sons - 5 sons. Not sure they want to live in village. He will only build another if they want him to. eS came from Germany and did the plastering. He arranged the carpentry in Kayseri. Çok para gitti. {much money went} Has a tractor.

Date:July1980 Place:Sakaltutan
Zafor:
Household now Zafor and wife; Nazif Unal and wife and youngest son, still a boy. They run two dolmuß; one with a driver from Süleymanli. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin de©il. { not sharp - i.e.? not profitable} I said he did very well on 8,000 TL with only two journeys a day. Nazif Unal has "bought" a Durak {dolmuß stop} from Belediye and works all day in Kayseri.

http://onlineqda.hud.ac.uk/Intro_QDA/Examples_of_Qualitative_Data.php

LETTERS

A role for self-gravity at multiple length scales in the process of star formation

Alyssa A. Goodman[1,2], Erik W. Rosolowsky[2,3], Michelle A. Borkin[1]†, Jonathan B. Foster[2], Michael Halle[1,4], Jens Kauffmann[1,2] & Jaime E. Pineda[2]

Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size ~0.1 parsecs) inside molecular clouds collapse to form star-plus-disk systems[1]. But self-gravity's role at earlier times (and on larger length scales, such as ~1 parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function[2]. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by ^{13}CO observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission[3] are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the locations of already-forming stars, or of those probably about to form, a self-gravitating cocoon seems a critical condition for their existence. Turbulent fragmentation simulations without self-gravity—even of unmagnetized isothermal material—can yield mass and velocity power spectra very similar to what is observed in clouds like L1448. But a dendrogram of such a simulation[4] shows that nearly all the gas in it (much more than in the observations) appears to be self-gravitating. A potentially significant role for gravity in 'non-self-gravitating' simulations suggests inconsistency in simulation assumptions and output, and that it is necessary to include self-gravity in any realistic simulation of the star-formation process on subparsec scales.

Spectral-line mapping shows whole molecular clouds (typically tens to hundreds of parsecs across, and surrounded by atomic gas) to be marginally self-gravitating[5]. When attempts are made to further break down clouds into pieces using 'segmentation' routines, some self-gravitating structures are always found on whatever scale is sampled[6,7]. But no observational study to date has successfully used one spectral-line data cube to study how the role of self-gravity varies as a function of scale and conditions, within an individual region.

Most past structure identification in molecular clouds has been explicitly non-hierarchical, which makes difficult the quantification of physical conditions on multiple scales using a single data set. Consider, for example, the often-used algorithm CLUMPFIND[7]. In three-dimensional (3D) spectral-line data cubes, CLUMPFIND operates as a watershed segmentation algorithm, identifying local maxima in the position–position–velocity (p–p–v) cube and assigning nearby emission to each local maximum. Figure 1 gives a two-dimensional (2D) view of L1448, our sample star-forming region, and Fig. 2 includes a CLUMPFIND decomposition of it based on ^{13}CO observations. As with any algorithm that does not offer hierchically nested or overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line

Figure 1 | Near-infrared image of the L1448 star-forming region with contours of molecular emission overlaid. The channels of the colour image correspond to the near-infrared bands J (blue), H (green) and K (red), and the contours of integrated intensity are from ^{13}CO(1–0) emission[8]. Integrated intensity is monotonically, but not quite linearly (see Supplementary Information), related to column density[14], and it gives a view of 'all' of the molecular gas along lines of sight, regardless of distance or velocity. The region within the yellow box immediately surrounding the protostars has been imaged more deeply in the near-infrared (using Calar Alto) than the remainder of the box (2MASS data only), revealing protostars as well as the scattered starlight known as 'Cloudshine'[11] and outflows (which appear orange in this colour scheme). The four billiard-ball labels indicate regions containing self-gravitating dense gas, as identified by the dendrogram analysis, and the leaves they identify are best shown in Fig. 2a. Asterisks show the locations of the four most prominent embedded young stars or compact stellar systems in the region (see Supplementary Table 1), and yellow circles show the millimetre-dust emission peaks identified as star-forming or 'pre-stellar' cores[3].

[1]Initiative in Innovative Computing at Harvard, Cambridge, Massachusetts 02138, USA. [2]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [3]Department of Physics, University of British Columbia, Okanagan, Kelowna, British Columbia V1V 1V7, Canada. [4]Surgical Planning Laboratory and Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA. †Present address: School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

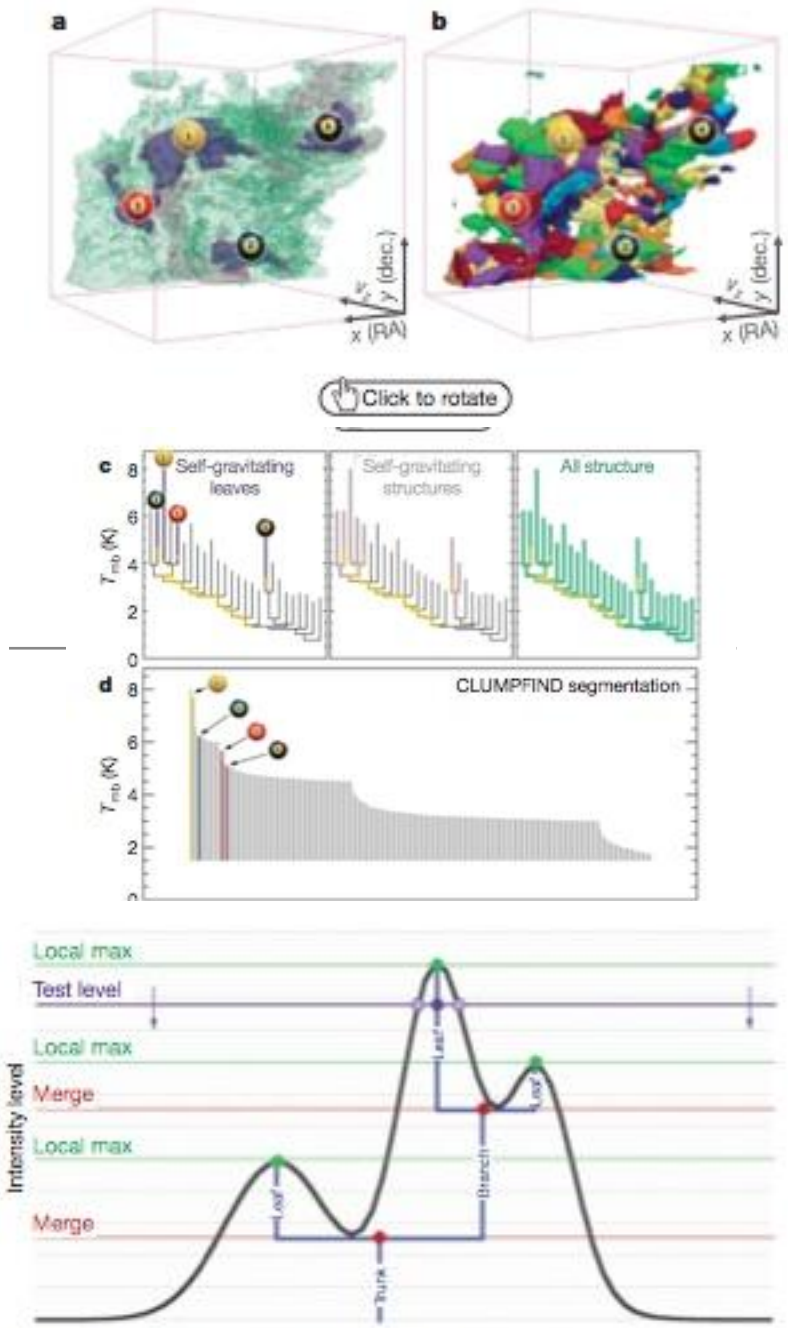
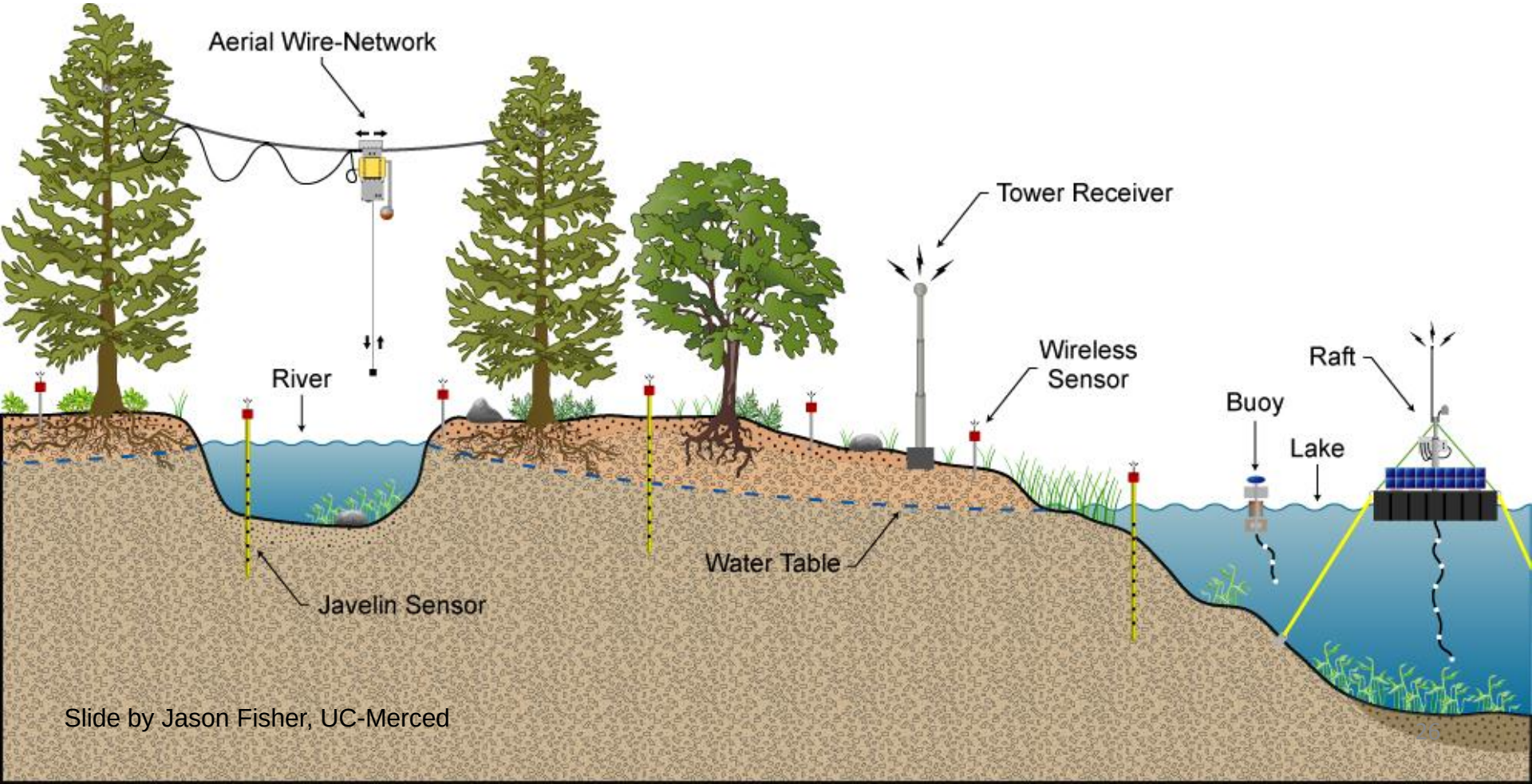


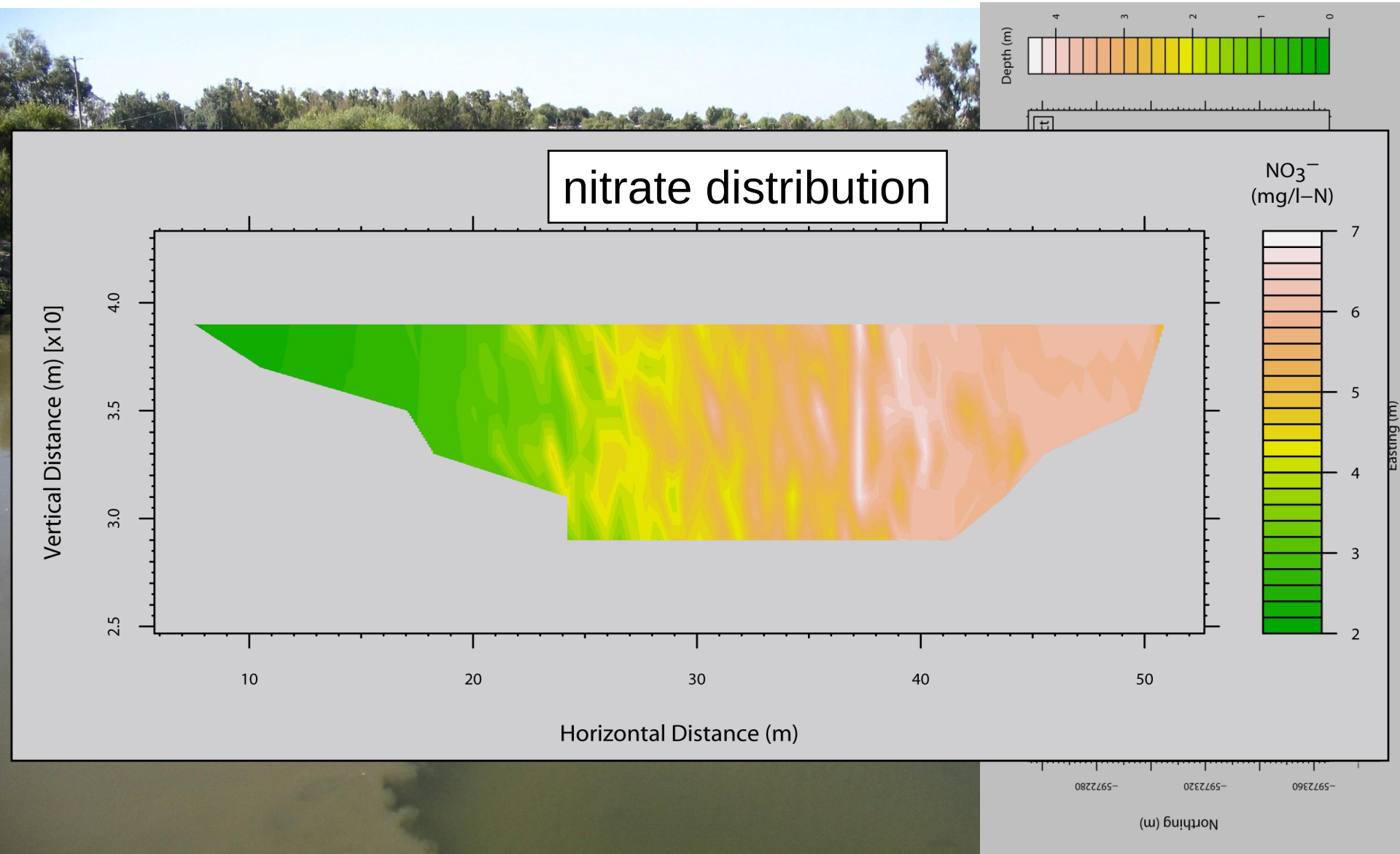
Figure 3 | Schematic illustration of the dendrogram process. Shown is the

Sensor networked science



Slide by Jason Fisher, UC-Merced

Sensor network data



Social science data

56. Generally speaking, do you usua ndependent, or what?

RESPONSE	PUN					COL: 240
			1994	1996	1998	ALL
Strong Democrat		7	423	400	370	6,046
Not very strong Democrat		1	644	577	597	8,756
Independent, close to Democrat		0	341	356	349	4,581
Independent (Neither, No response)		5	369	457	477	4,882
Independent, close to Republican		8	282	258	244	3,379
Not very strong Republican		9	519	500	484	6,265
Strong Republican		0	321	307	239	3,479
Other party, refused to say		7	44	43	63	530
Don't know		0	0	0	0	10
No answer		9	49	6	9	188

See Appendix D: Recodes, for origi Appendix N for changes
across surveys. If planning to perfor GSS Methodological Report
No. 56.

http://dss.princeton.edu/images/gss.gif

karmāntājīvā
virahitākuśalakāyavāṅmanas-
karmāntājīvāś[109] ca bhavanti sma

§ 1.66 (147c 28-148a 1)

一切衆生得平等心，展轉相瞻如父、如母、如兄、如弟、如姊、如妹，各各同心，等無偏邪，皆行慈心。

PG 4r 4-5 (Ś 18, 22-19, 1; PD 10, 1-2; PSL *kā* a 4-5): sarvasatvāś ca sarvasatveṣu samacittā abhūvan* yad uta[110] mātāpitṛbhrātṛbhaginisamacittāḥ mitrajñātisahāyasamacittāḥ[111]

§ 1.67 (148a 1-2)

一切群萌悉修十善，清淨梵行，無有塵埃。

PD 10, 2-3 (PG 4r 5-6; Ś 19, 2-3; PSL *kā* a 5): daśakuśalakarmapathasevinaś ca bhavanti sma[112] / brahmacāriṇaḥ śucayo nirāmayagandhāḥ[113]

§ 1.68 (148a 2-4)

一切黎庶悉獲安隱，所得安隱猶如比丘得第三禪。于時衆生而致智慧，而悉具足善快調定，離於卑劣，逮得和雅。

PG 4r 6-8 (PD 10, 3-8; PSL *kā* a 5-6; Ś 19, 3-8): sarvasatvās tasmin samaye sarvasukhasamarpitā abhūvan* evaṃrūpeṇa sukhena samanvāgatās[114] tadyathā {s} tṛtiyadhyānasamāpannasya bhikṣoḥ sukhaṃ sarvasatvāś ca tasmin samaye evaṃrūpayā prajñayā samanvāgatā abhūvan* yad evaṃ jānaṃti sma[115] • sādhu dānaṃ sādhu damaḥ sādhu saṃyyamaḥ[116] sādhu satyaṃ • sādhv apramādaḥ sādhu maitrī sādhu karuṇā sādhv avihiṃsā prāṇibhūteṣu[117] •

---

[110] sarvasattvāś ... yad uta: not in PD & PSL.

[111] PG wrongly repeats verbatim this latter compound. PD 10, 2 and PSL have at this point a longer reading: mitrāmātyajñātisālohitasamacittā. Note that Ś has all the words construed as one compound.

[112] PG 4r 5-6 & Ś 19, 2: daśakuśalakarmapa(tha)samanvāgatā [Ś without daśa-] abhūvan.

[113] PG 4r 6, Ś 19, 3 and PSL *kā* a 5: nirāmagandhāḥ, which seems to be the correct reading; after this word, PG & Ś + sarvākuśalavitarkavigatāḥ.

[114] PD 10, 4 & PSL *kā* a 5: idṛśaṃ sukhaṃ pratilabhante sma.

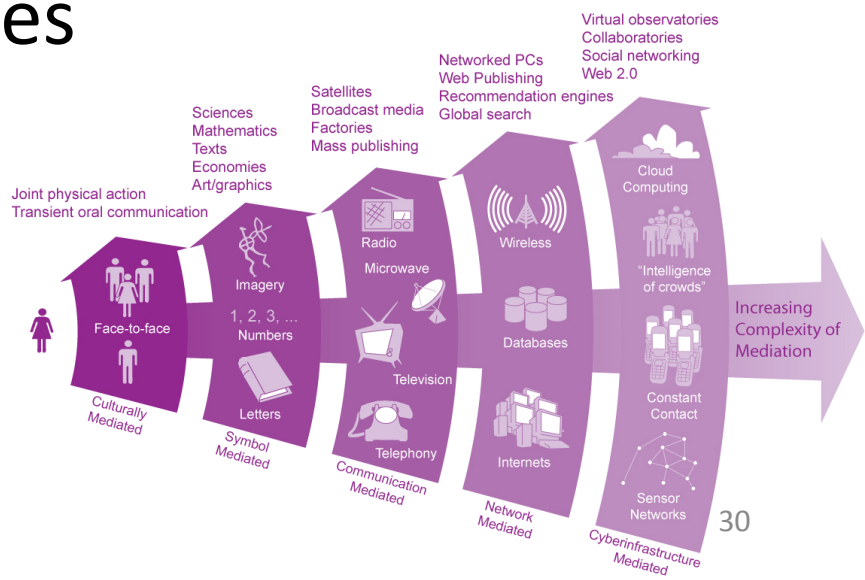
[115] yad ... sma: PD 10, 6 & PSL *kā* a 6: yad anyabuddhakṣetrasthā buddhā bhagavanta evam [PSL + udānam] udānayanti sma.

[116] Ś 19, 7: saṃyamaḥ.

[117] sādhu dānaṃ ... prāṇibhūteṣu: PD 10, 7-8 & PSL *kā* a 6: sādhu damaḥ [PSL + sādhu śamaḥ] sādhu saṃyamaḥ sādhu cīrṇo brahmacaryyāvāsaḥ sādhu prāṇibhūteṣv avihiṃseti.

Libraries, librarians, and data

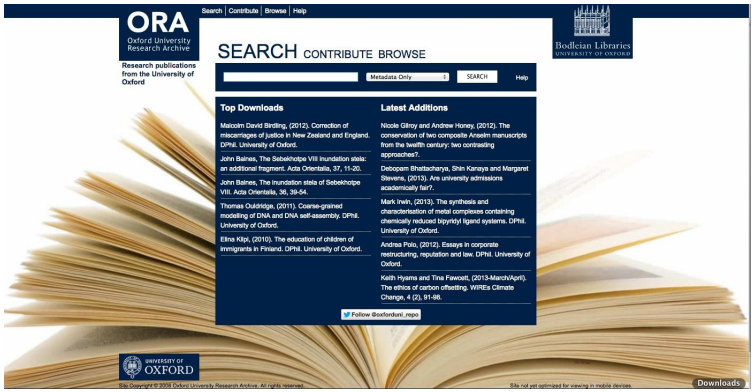
1. Data are not publications
2. Data are not natural objects
3. Data are representations
4. Data sharing and reuse depend on knowledge infrastructures

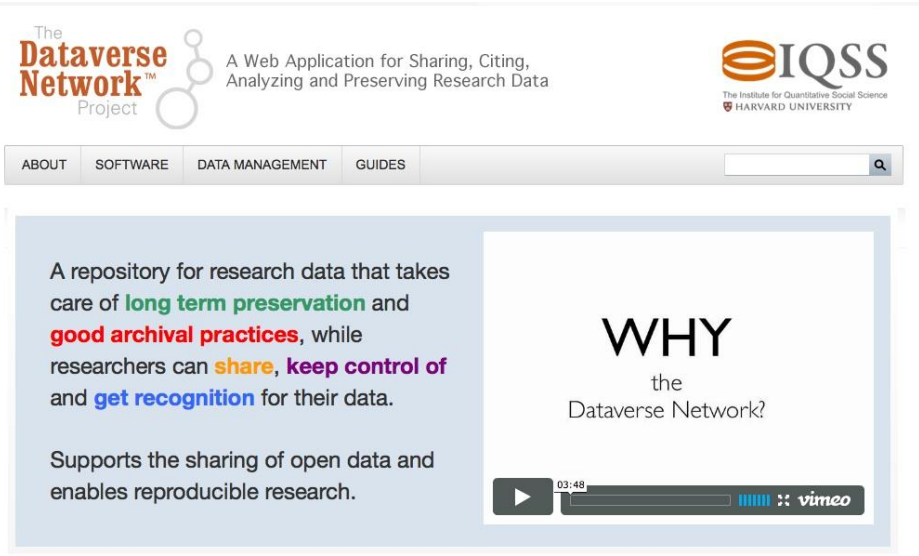


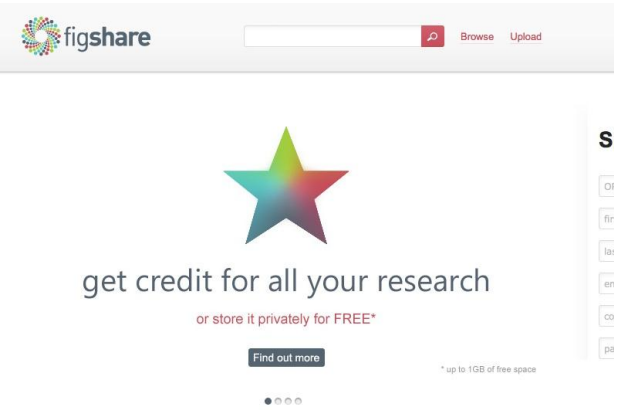
Data

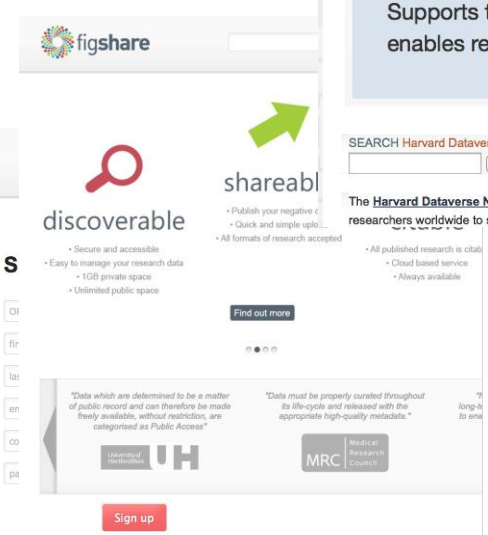
- **Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.**

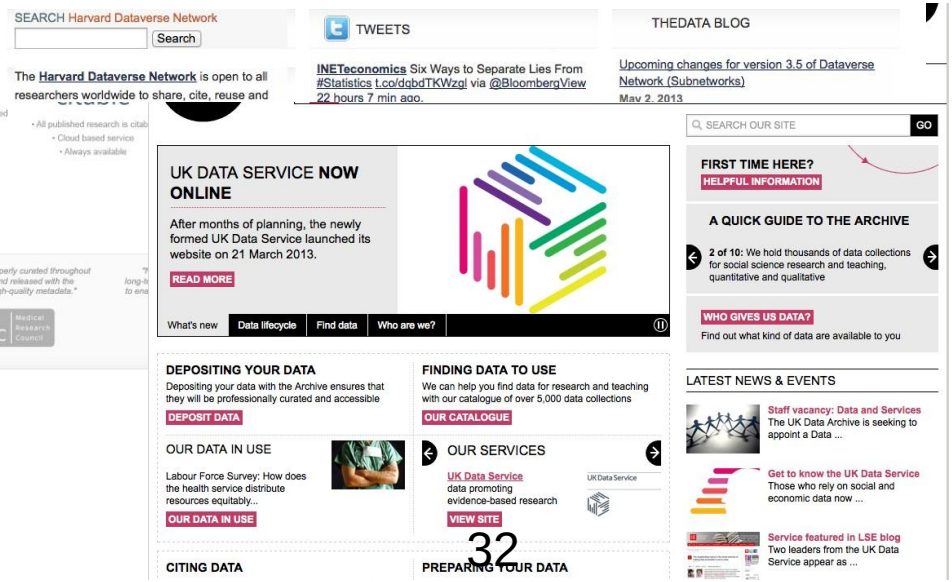
Representations of research objects











Description and Classification

The Open Biological and Biomedical Ontologies

Ontologies | Resources | Participate | About

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The 🐾s link to the term request trackers for the listed ontologies.

OBO Foundry ontologies

Title	Domain	Prefix	File	Last changed
Biological process	biological process	GO	go.obo	
Cellular component	anatomy	GO	go.obo	
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	
Molecular function	biological function	GO	go.obo	
Phenotypic quality	phenotype	PATO	quality.obo	
PRotein Ontology (PRO)	proteins	PR	pro.obo	
Xenopus anatomy and development	anatomy	XAO	xenopus_anatomy_edit.obo	
Zebrafish anatomy and development	anatomy	ZFA	zebrafish_anatomy.obo	2013/04/12

OBO Foundry candidate ontologies and other ontologies of interest

Title	Domain	Prefix	File	Last changed
Adverse Event Reporting Ontology	health	AERO	aero.owl	
Amphibian gross anatomy	anatomy	AAO	AAO_v2_edit.obo	
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Anatomical Entity Ontology	anatomy	AEO	aeo.obo	2012/06/01
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2013/05/02
Basic Formal Ontology	upper	BFO	1.1	
Bilateria anatomy	anatomy	BILA	bilateria_mrca.obo	
Biological imaging methods	experiments	FBbi	image.obo	2011/05/24
BRENDA tissue / enzyme source	anatomy	BTO	BrendaTissueOBO	
C. elegans development	anatomy	WBls	worm_development.obo	

Quick Links

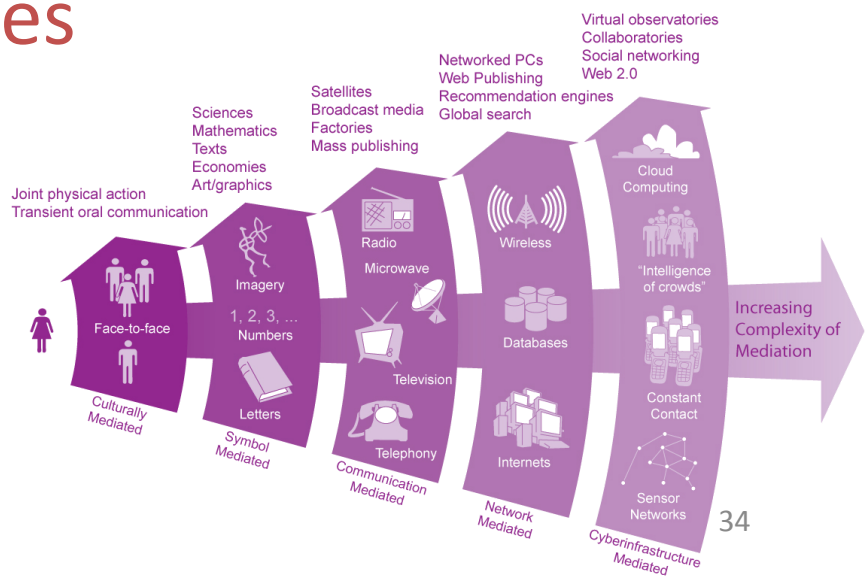
- Mappings between ontologies
- Download alternate formats
- About the OBO Foundry
- Current events
- How to join
- OBO Foundry paper in *Nature Biote*
  November 2007

Other Ontology Lists

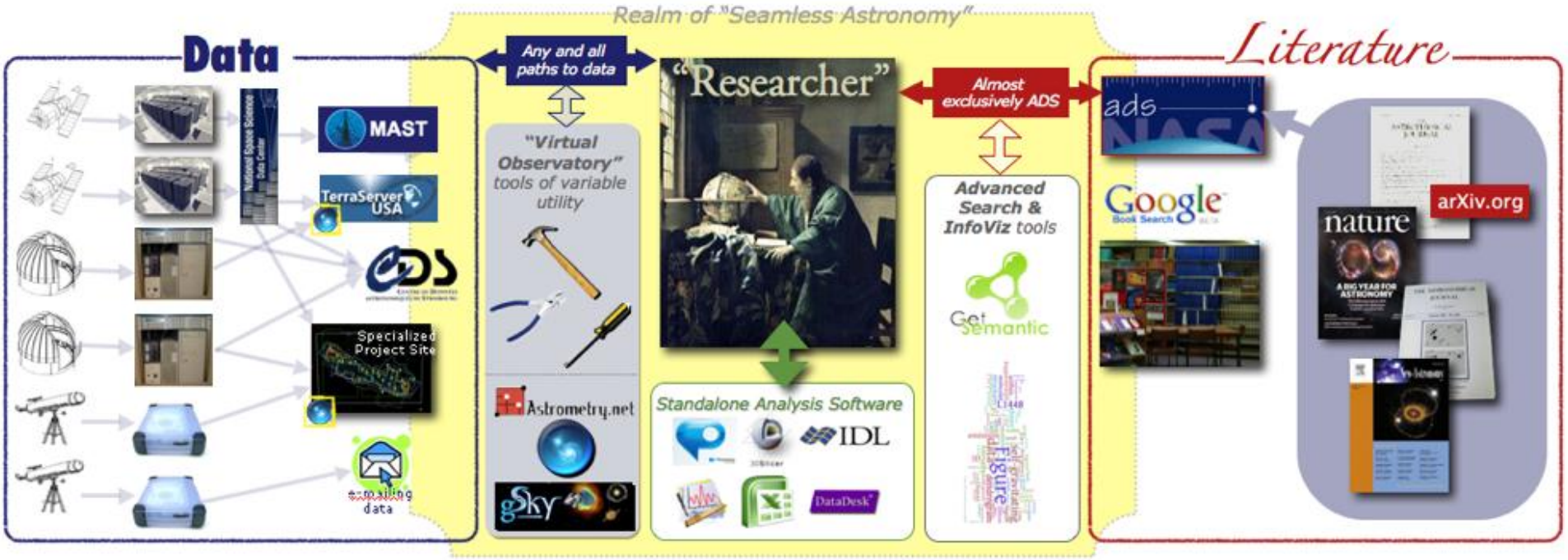
- BioPortal (NCBO's ontology repo
- Ontology Lookup Service (OLS)
  term lookup)

Libraries, librarians, and data

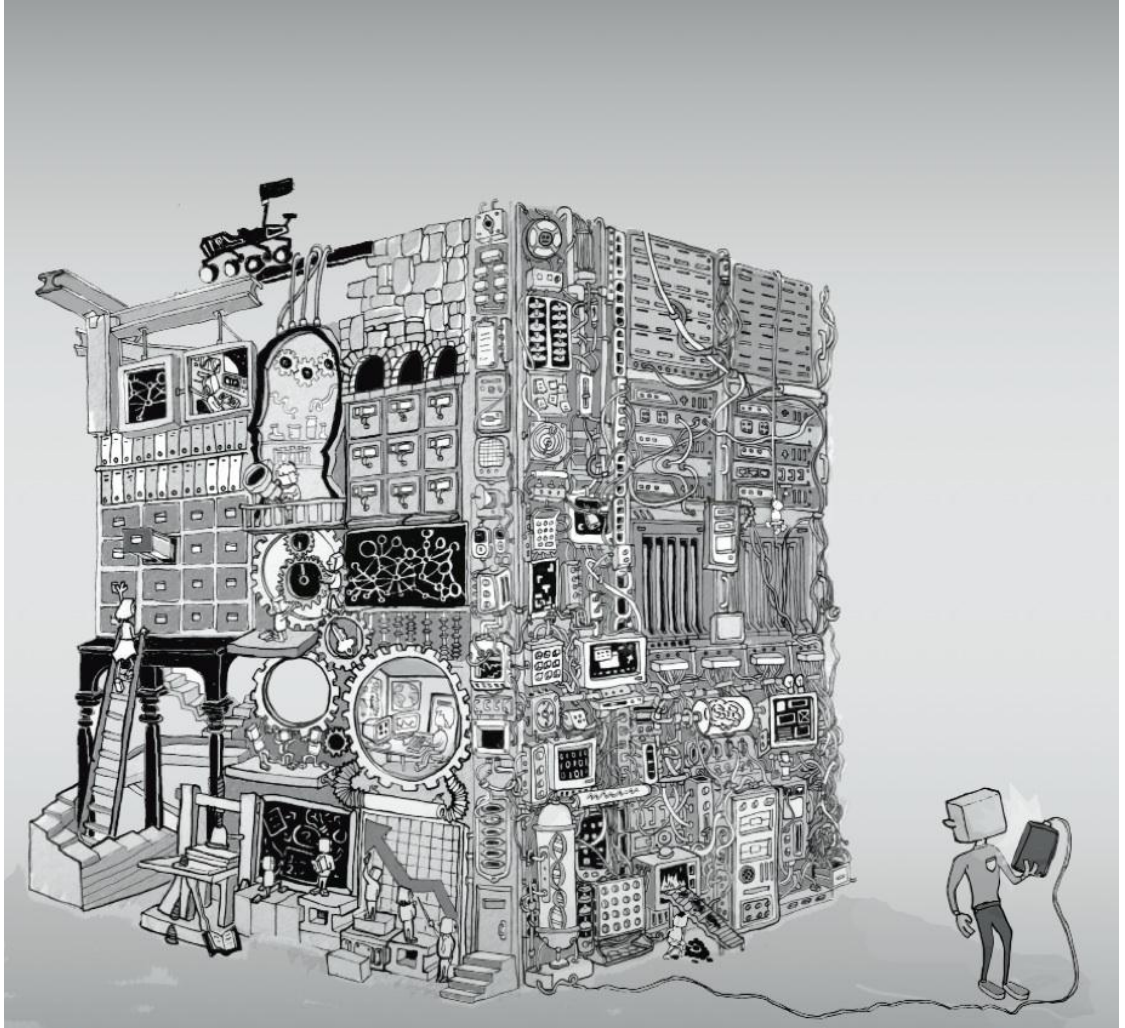
1. Data are not publications
2. Data are not natural objects
3. Data are representations
4. Data sharing and reuse depend on knowledge infrastructures



Knowledge Infrastructures



Slide: Alyssa Goodman, Harvard

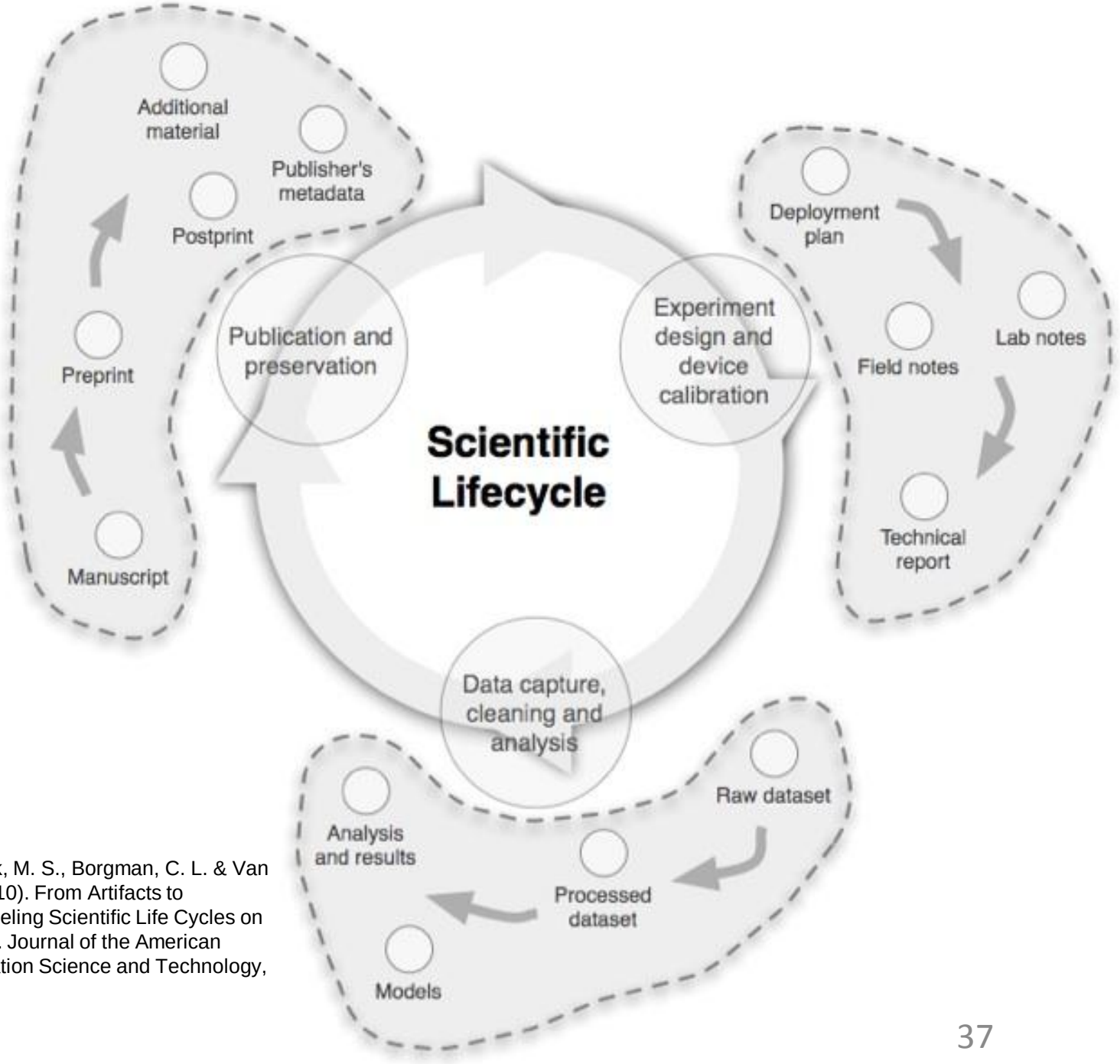


Knowledge Infrastructures: Intellectual Frameworks and Research Challenges

Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation

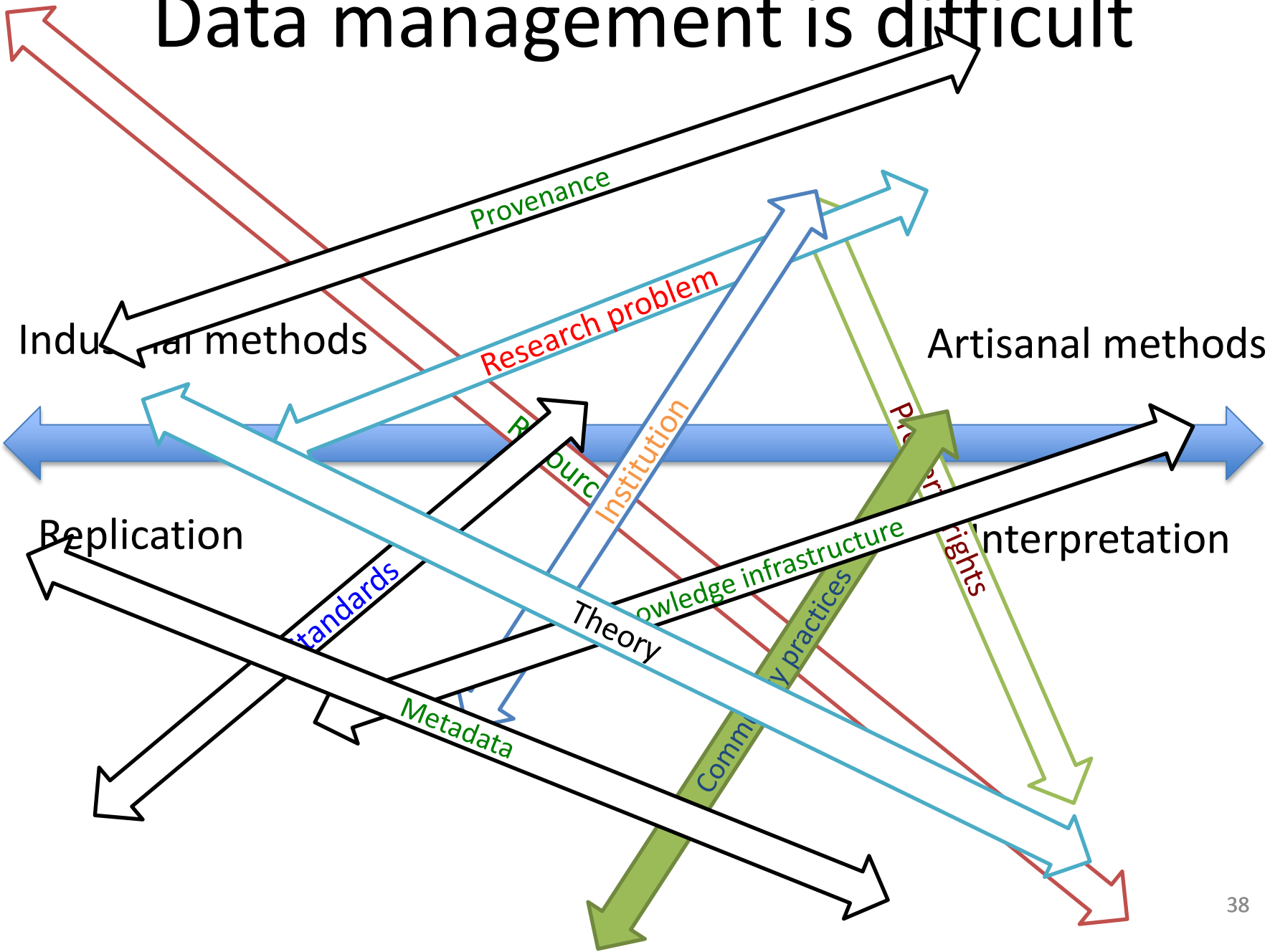
University of Michigan School of Information, 25-28 May 2012

http://knowledgeinfrastructures.org



Pepe, A., Mayernik, M. S., Borgman, C. L. & Van de Sompel, H. (2010). From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. Journal of the American Society for Information Science and Technology, 61(3): 567–582.

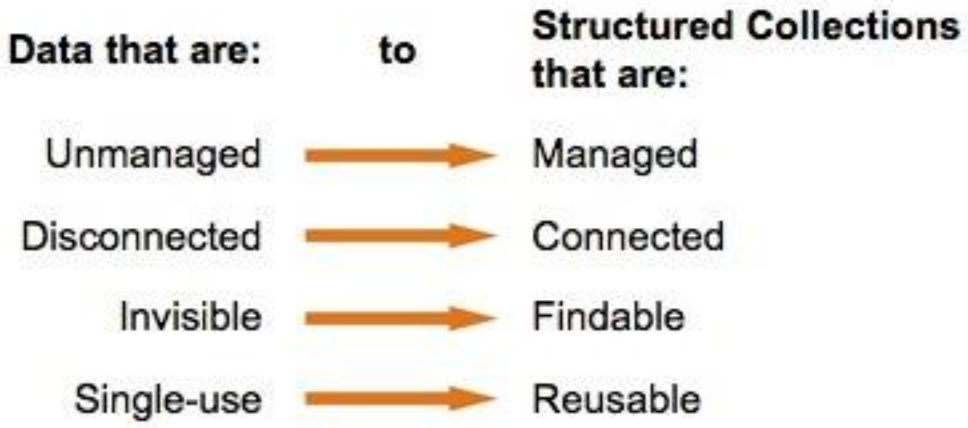
Data management is difficult



Ways to Release Data

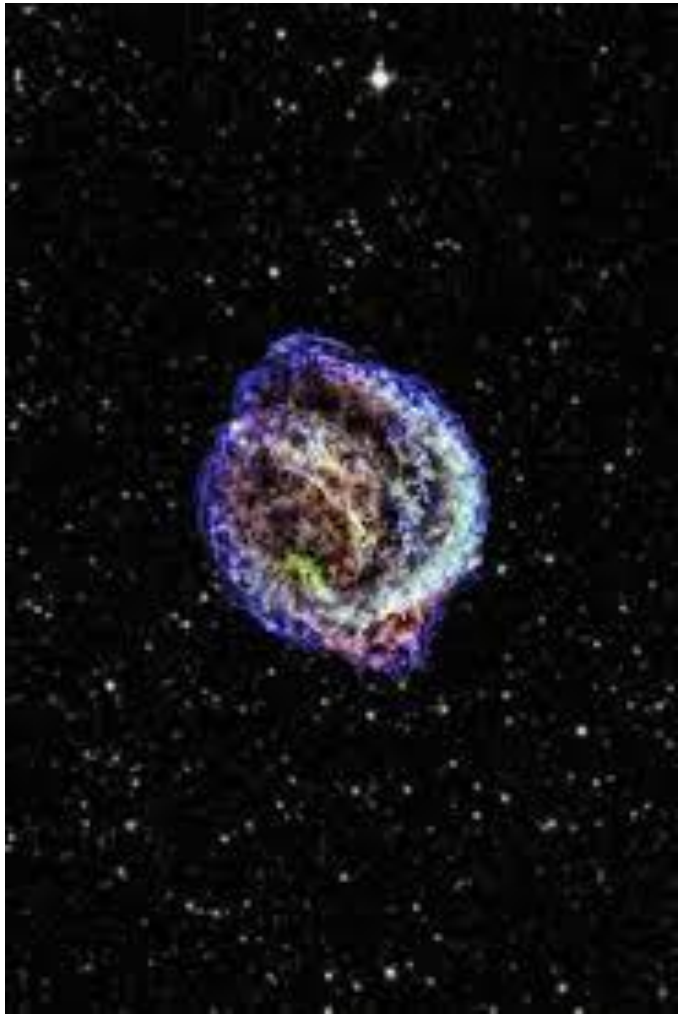
- Contribute to archive
- Attach to journal article
- Post on local website
- License on request
- Release on request
- ….

ANDS goals



Degrees of reuse

- Reuse by investigator
- Reuse by collaborators
- Reuse by colleagues
- Reuse by unaffiliated others
- Reuse at later times
  - Months
  - Years
  - Decades
  - Centuries



http://chandra.harvard.edu/photo/2013/kepler/kepler_525.jpg

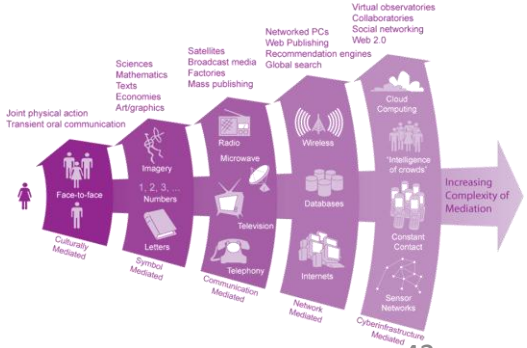
Economics of the Knowledge Commons

	Subtractability / Rivalry		
		Low	High
Exclusion	Difficult	**Public Goods** General knowledge Public domain data	**Common-pool resources** Libraries Data archives
	Easy	**Toll or Club Goods** Subscription journals Subscription data	**Private Goods** Printed books Raw or competitive data

Adapted from C. Hess & E. Ostrom (Eds.), *Understanding knowledge as a commons: from theory to practice*. Cambridge, Mass.: MIT Press.

Conclusions

- Stakeholders
- Economics of research data
- Sustainability
- Knowledge infrastructures and libraries



Stakeholders in research data

- Scholars
  - Grants and funding
  - Credit and promotion
  - Publishing
  - Managing data

http://2.bp.blogspot.com/-fzIYX9zpcUU/T0kAuG83FGI/AAAAAAAAAKg/hgaotz3E--A/s1600/infrastructure.jpg

- Others
  - Scholars
  - Students
  - Readers / public at large
  - Universities
  - Funding agencies
  - Repositories
  - Governments
  - Private enterprise
  - Libraries
  - Archives
  - Museums…

Economics of Research Data

- Private goods
- Common-pool resources
- Toll or club goods
- Public goods

http://2.bp.blogspot.com/-fzIYX9zpcUU/T0kAuG83FGI/AAAAAAAAAKg/hgaotz3E--A/s1600/infrastructure.jpg

Sustainability

- What to keep
- Why to keep them
- How to keep them
- How long to keep them
- Who will govern them
- What kinds of expertise are required

http://2.bp.blogspot.com/-fzIYX9zpcUU/T0kAuG83FGI/AAAAAAAAAKg/hgaotz3E--A/s1600/infrastructure.jpg

Knowledge Infrastructure and Libraries

- Technical fabric
  - Data archives
  - Metadata, provenance, classification, collections
  - Tools, services, support
- Social fabric
  - Data curation workforce
  - Data management training
- Trust fabric
  - Memory institutions
  - Stewardship

http://2.bp.blogspot.com/-fzIYX9zpcUU/T0kAuG83FGI/AAAAAAAAAKg/hgaotz3E--A/s1600/infrastructure.jpg

Acknowledgements

- UCLA Data Practices team
  - Rebekah Cummings, Peter Darch, David Fearon, Ariel Hernandez, Matthew Mayernik, Alberto Pepe, Ashley Sands, Katie Shilton, Sharon Traweek, Jillian Wallis, Laura Wynholds, Kan Zhang
- Research funding
  - National Science Foundation
  - Alfred P. Sloan Foundation
  - Microsoft External Research
- University of Oxford
  - Balliol College
  - Oliver Smithies Fellowship
  - Oxford Internet Institute
  - Oxford eResearch Center
  - Bodleian Library

Provocations

1. How do rights, responsibilities, and risks associated with research data vary by discipline and by stakeholder?
2. How can data be exchanged across domains, contexts, and over time?
3. How do publications and data differ?
4. What are scholars' motivations to share their data?
5. What kinds of expertise are required to manage, curate, and sustain access to research data?
6. How can knowledge infrastructures adapt to the needs of generations of scholars and the demands of competing stakeholders?

10 Simple Rules for the Care and Feeding of Scientific Data*

1. Love your data, and help others love it too.
2. Share your data online, with a permanent identifier.
3. Conduct science with a particular level of reuse in mind.
4. Publish workflow as context
5. Link your data to your publications as early as possible.
6. Publish your code (even the small bits).
7. Say how you want to get credit for your data (and software).
8. Foster and use data repositories.
9. Reward colleagues who share their data properly.
10. Be a booster for data science.

**Goodman, A.; Pepe, A.; Blocker, A.; Borgman, C.L., et al, (forthcoming), PLOS Computational Biology*
https://www.authorea.com/users/3/articles/3410/_show_article