Title
Creating, Collaborating, and Celebrating the Diversity of Research Data

Permalink
https://escholarship.org/uc/item/7cs811mj

Author
Borgman, Christine L.

Publication Date
2015-10-26

Copyright Information

October 26, 2015

Creating, Collaborating, and Celebrating the Diversity of Research Data

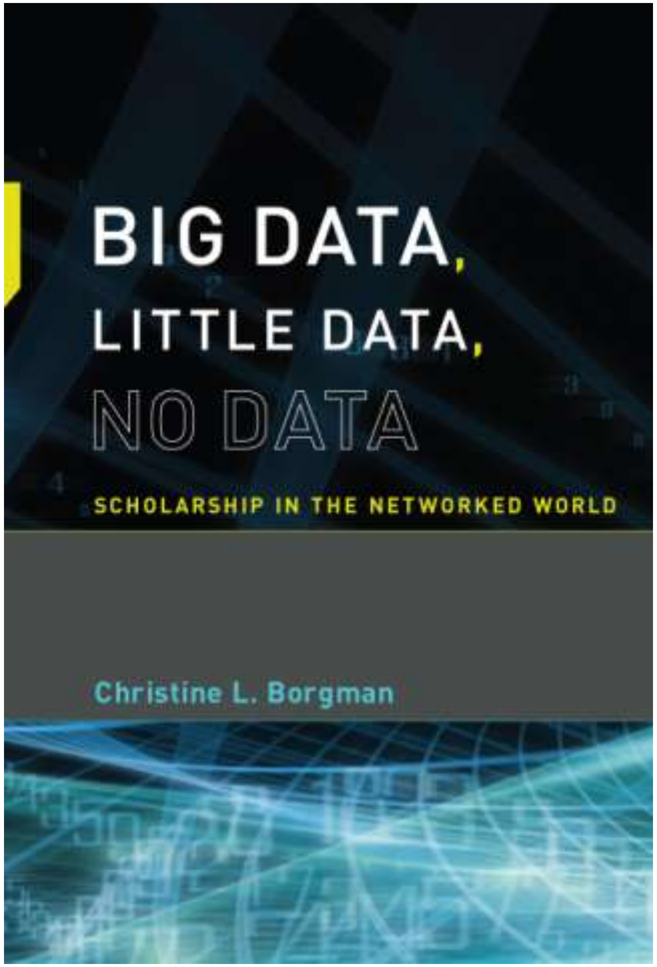
Christine L Borgman, *University of California, Los Angeles*

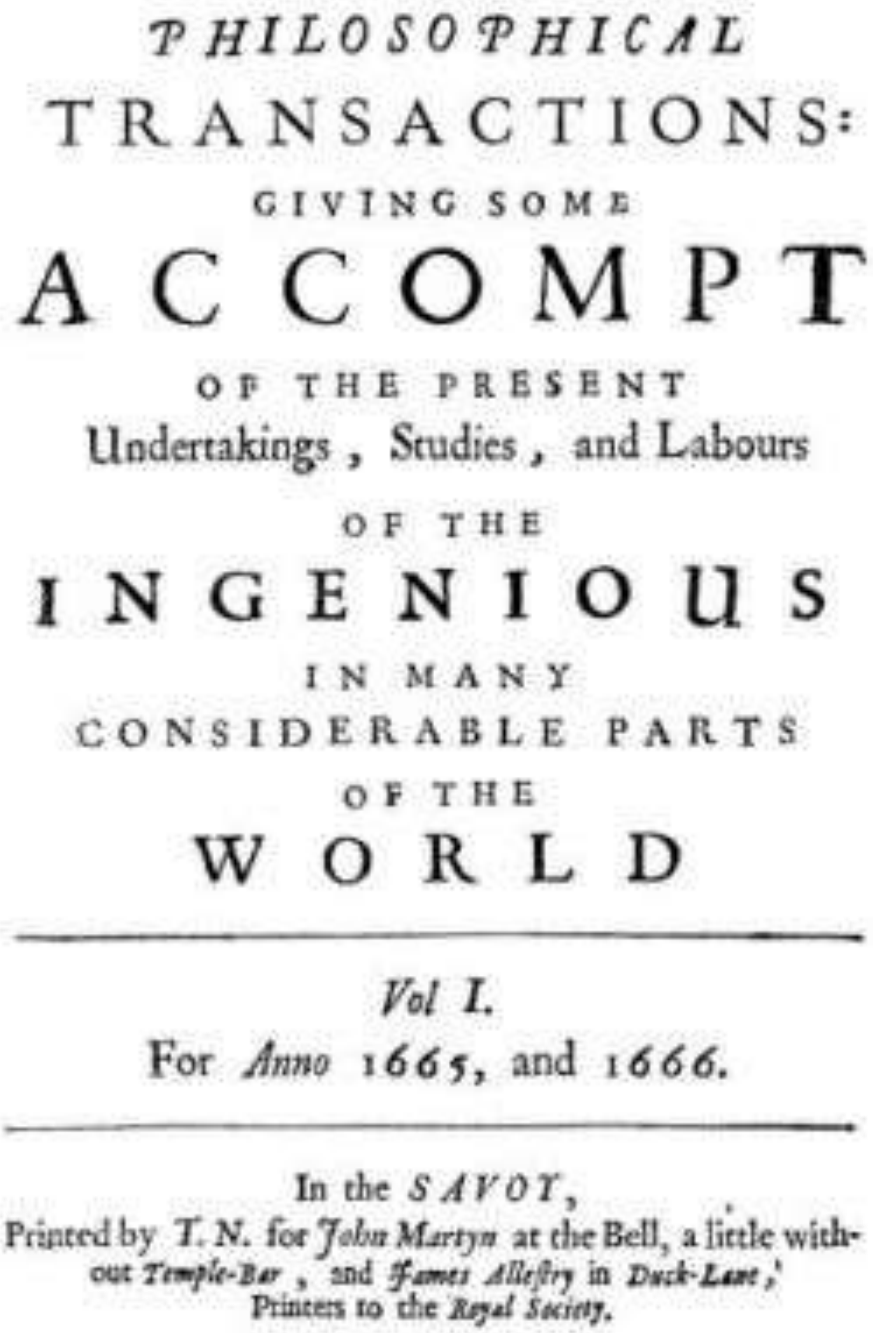
Creating, Collaborating, and Celebrating the Diversity of Research Data

Christine L. Borgman

Distinguished Professor
and Presidential Chair in Information Studies
University of California, Los Angeles
@SciTechProf

Seminar Presentation
Graduate School of Library, Information, and Media Studies
University of Tsukuba, Japan
October 26, 2015





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 without *Temple-Bar*, and *James Alleſtry* in *Duck-Lane*,
Printers to the *Royal Society*.

Big Data, Little Data, Open Data, and Libraries

Christine L. Borgman

Professor and Presidential Chair in Information Studies

University of California, Los Angeles

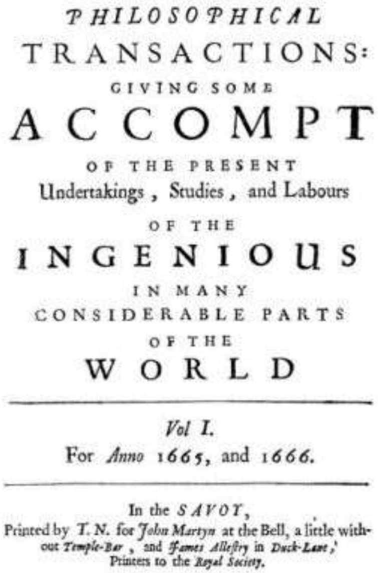
University of Göttingen

Inaugural Göttingen Lecture on Library Futures

23 March 2015

Data <–> Publications

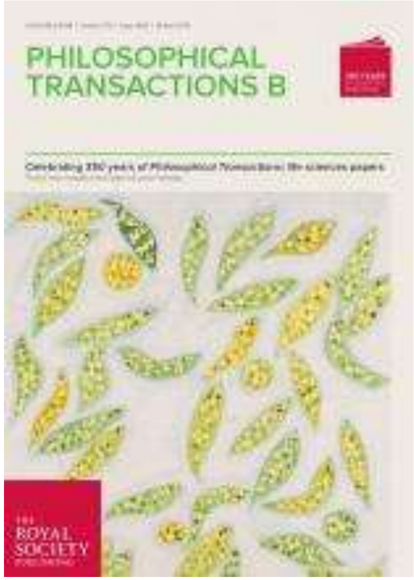
Publications are arguments made by authors, and data are the evidence used to support the arguments.



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 without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,
Printers to the *Royal Society*.


PHILOSOPHICAL
TRANSACTIONS B
Celebrating 350 years of Philosophical Transactions: life sciences papers
THE ROYAL SOCIETY

Theme issue 'Celebrating 350 years of Philosophical Transactions: life sciences papers' compiled and edited by Linda Partridge
19 April 2015; volume 370, issue 1666

THE WHITE HOUSE
WASHINGTON

Open access policies

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

- 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
  - 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

Australian Government
National Health and Medical Research Council

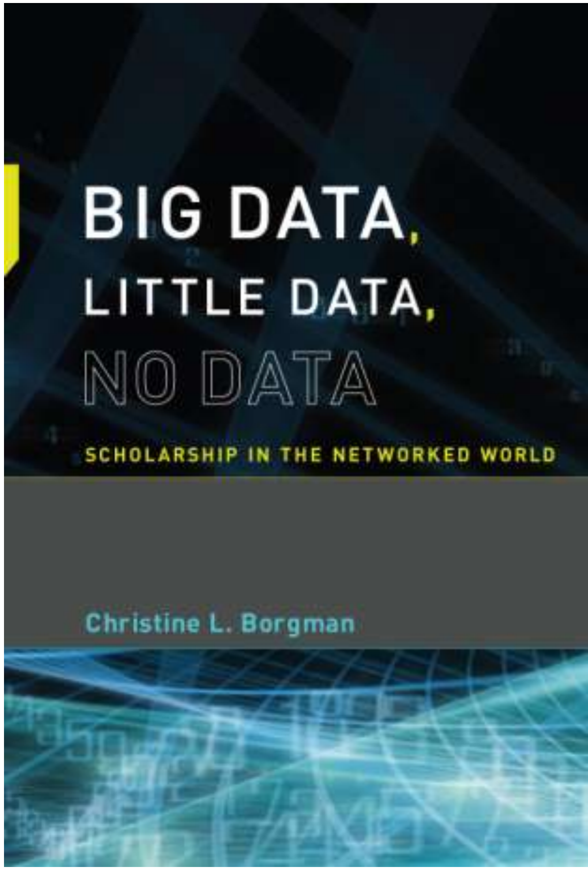
National Science Foundation
WHERE DISCOVERIES BEGIN

Supported by
wellcometrust

Policy RECommendations for Open Access to Research Data in Europe
RECODE

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

- Part I: Data and Scholarship
  - Ch 1: Provocations
  - Ch 2: What Are Data?
  - Ch 3: Data Scholarship
  - Ch 4: Data Diversity
- Part 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
- Part 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

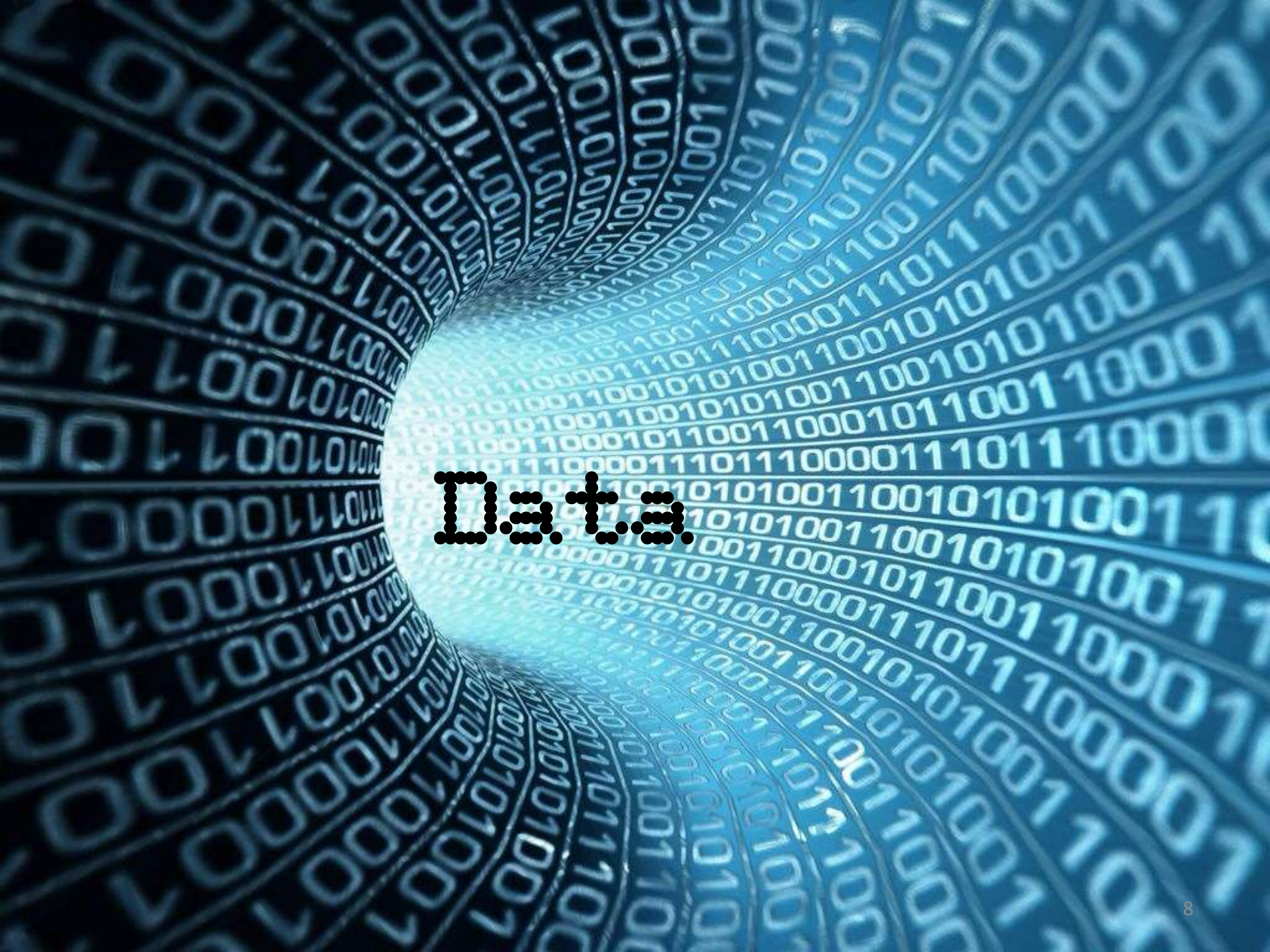


Celebrating the diversity of data

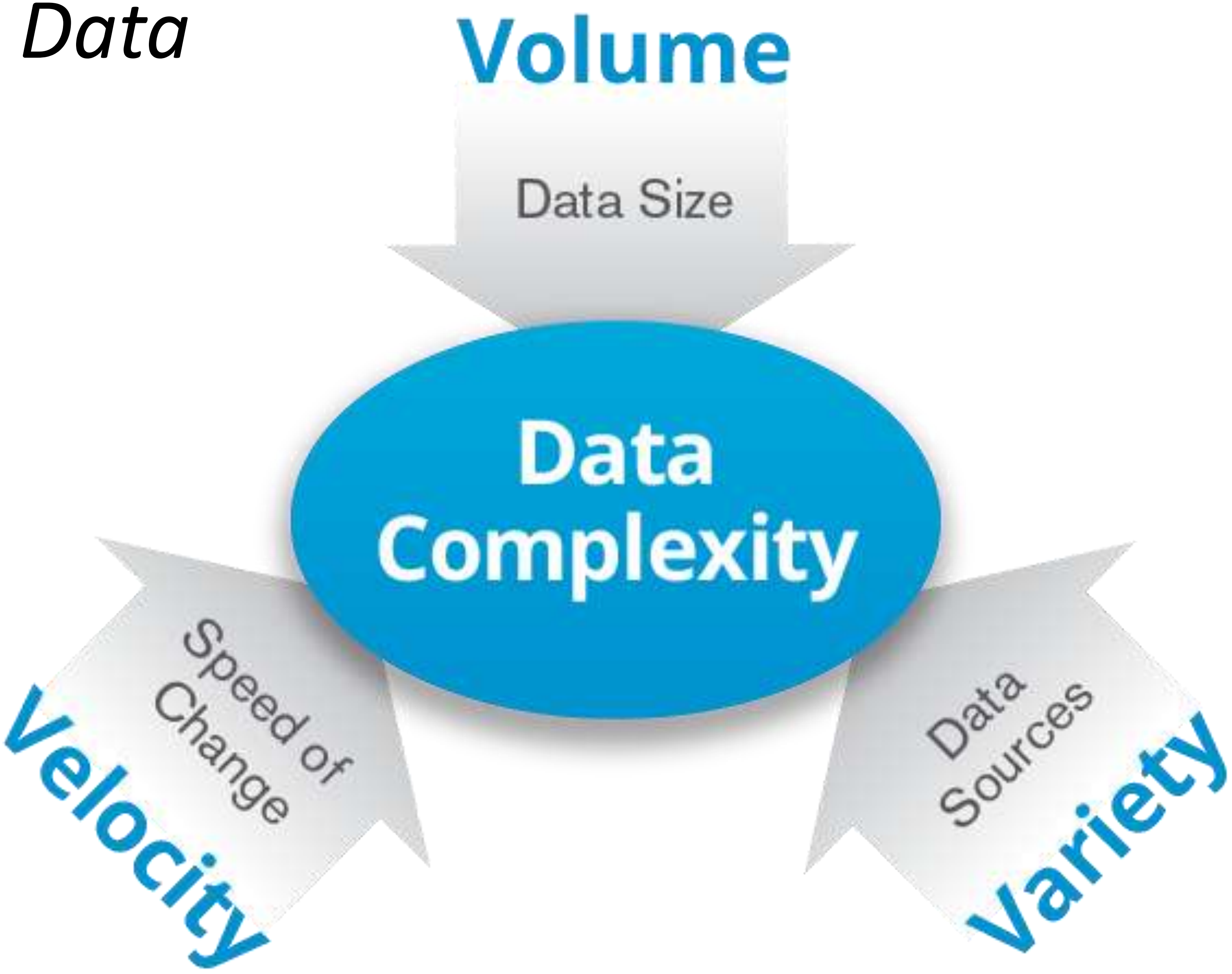
- Defining data
- Creating data
- Collaborating with data
- Consolidating data value

Persistent URL: photography.si.edu/SearchImage.aspx?id=5799
Repository: Smithsonian Institution Archives

Data

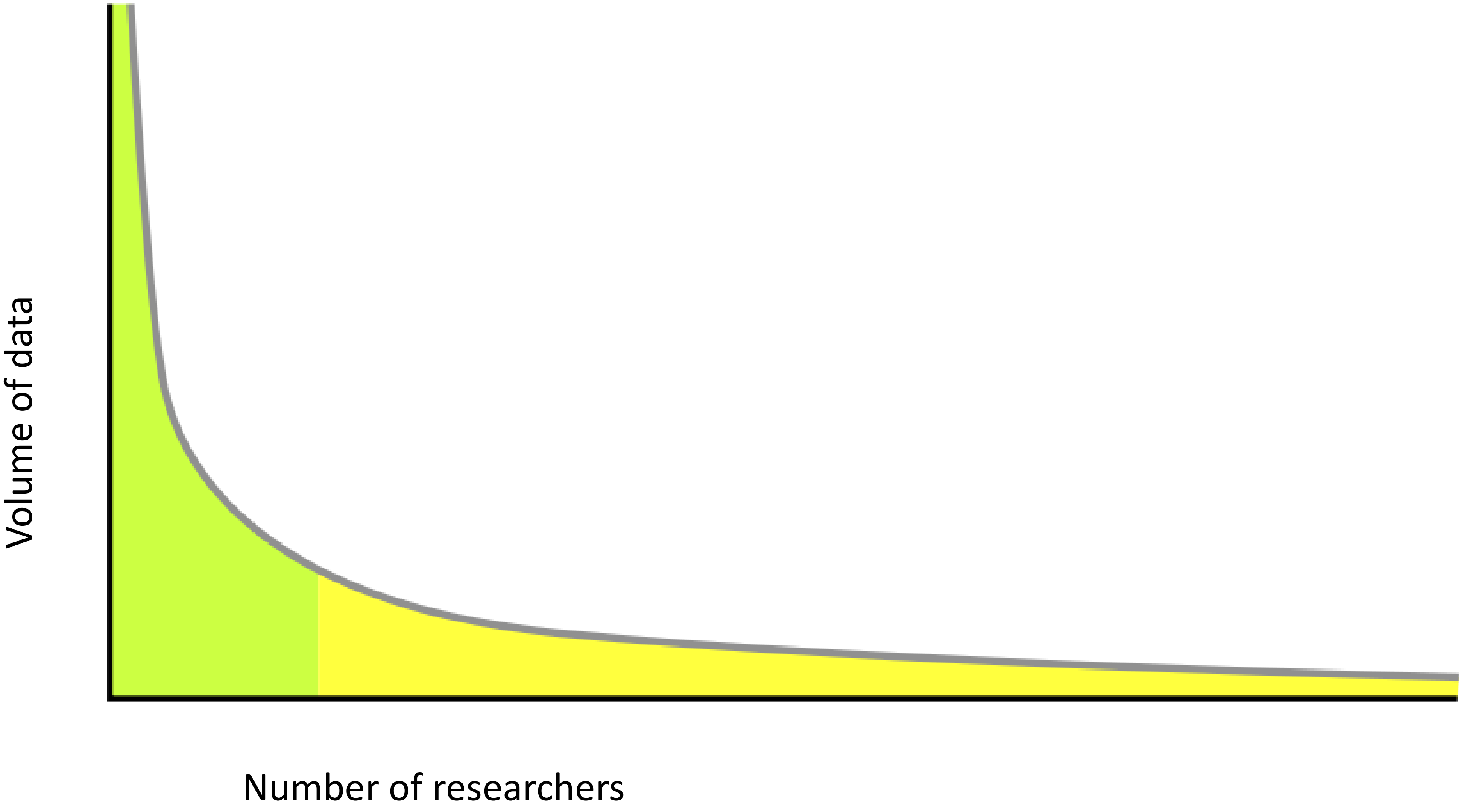


Big Data



http://www.datameer.com/product/hadoop.html

Long tail of data



Slide: The Institute for Empowering Long Tail Research

Open Data: Free

- 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

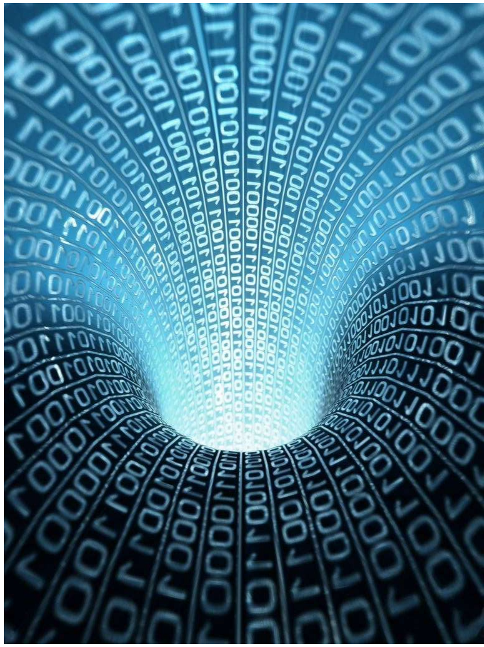
State Library and Archives of Florida, 1922.
Flickr commons photo

Open Data Commons. (2013).

```
/FontMatrix matrix def
/FontBBox[2048 -1164 1 index div -628 2 index div 4096 3 index div 2062 5 -1 rol
/sfnts [<
7472756500090000000000006376742000000000000000009C000007DA6670676D000000000000000878
019901AC01C101C501C901E101F601F601F602220222022802360023F0243024602670285028502940
01160125011800EA00EA00AE0000003E05BB008A04D70053003FFF8CFFD500150028002200990062
B200402F2B59B002602D2C21B0C051580C6423648BB81555621BB200802F2B59B002602D2C0C6423
B8FF80B30809341CB8FF8040E80809343609352F4A0959045809A7090626062B082C0B280C281328
D4401E091202550C0C047F18011840111202551840 1D1D02551810251E073C2C04002FED3FEDC42B
170003020913141500215902071B25090B1E0505262500180C0D0D065518021010065518 0CB8FFF8
1F1E4A62A1A4FB432E7900020044FFE40405058E001F002D024BB102024354584026 2F4010100255
2F2B2BDD2B2BC01112392F2BCDD0CD003FED5DC45D5D2B323FED12392F5D5D5DCD31301B4019125F
11242517012B50100110302A2912110608070A101B011BB80152B38F2D012DB802F6B2012A2BB801
025D5972103C103C11121739111239391112390111123939123939313043794 07A4B573A4524351C
2D602D702D04802D902D02B02D01002D102DC02DD02D042D6037A67F182B10F65D5D71723CFD3C10
FFEAB40C0C02550CB8FFE2B40D0D02550CB8FFD6B4101002550CB8FFDEB50F0F02550C05BD03E200
2F2F2F3FDDCD3F3F10ED10ED313001B00D4B5458BF0046FFE80045FFE8002FFFE80030FFE8B51A18
1010065529B8FFF2B70F0F06552929353411123 92F2B2B2BDD2B2B2BC01112392F2B2B2BCD2B2B2B
32363534272602 00D07E6B76CF7FCF7A677DCC53356B429F82617E694703AF9E87AF7BFC80A58BAD
961A9C1E9621982A9E2BA816A61AAB1CAD2BB916BE2BCD2BDA2BEC2BFB2B2A202D732573288F1397
201A01601A701A021A120B003FC45D5DED5D5D2F3FC45DED5D5D5D1217393F012F2B2BCD2F2BCDD4
3637029021267 75C4656201F5F92CBBD75546C2115170D21211C9E62455761FEDE2D2D9B7B364D33
393FDD5DCD31301B40350127600D5D362027602770 27B02704340B371F3A20481F4820051A08134F
FFE0400A1339082013391B201339012B2B2B2B002B012B2B59595913211523220615141713133635
C0B215391AB8FFF0401315393608153928301439293014391108163909B8FFE0401B163929401139
00000009000100000000000000000100000721FE4500571000FB74FADF10000001000000000000000000
003403630044036300040363B2242F1FBA034E006D0800400E1F7F027F037F047F050430440112BF
014A001F000D0126400B1F0DC61F0D571F0D371F0D410D019E0141000D00420141000D001E014100
021E0024455258B90024021E4459594BB8020153205C58B9010F00224544B1222245445958B90C00
2B2B2B2B2B2B2B2B2B2B2B2B2B000173750073730045694400737301737 42B2B2B2B2B732B00732B
00>] def
/CharStrings 27 dict dup begin
/.notdef 0 def
              /space 1 def
                          /comma 2 def
                                      /period 3 def
                                                   /W 4 def
                                                           /a 5 def
                                                                   /b 6 def
                                                                           /c 7
```

Open Data: Useful

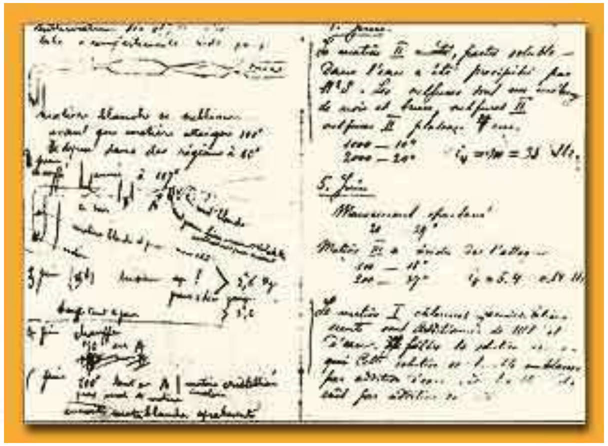
- Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, and sustainability.



Organization for Economic Cooperation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. http://www.oecd.org/dataoecd/9/61/38500813.pdf

What are data?

hudsonalpha.org



5. Juin

1000 — 10°
2000 — 20°

Marie Curie’s notebook aip.org

Pisa Griffin

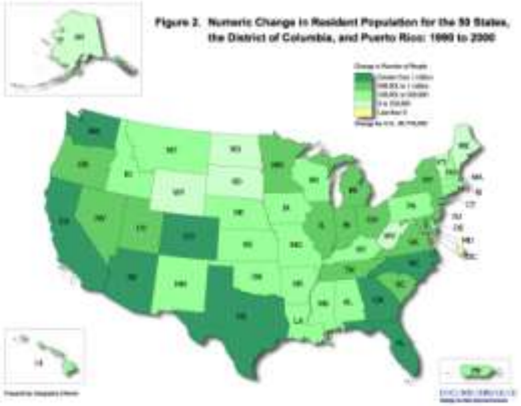
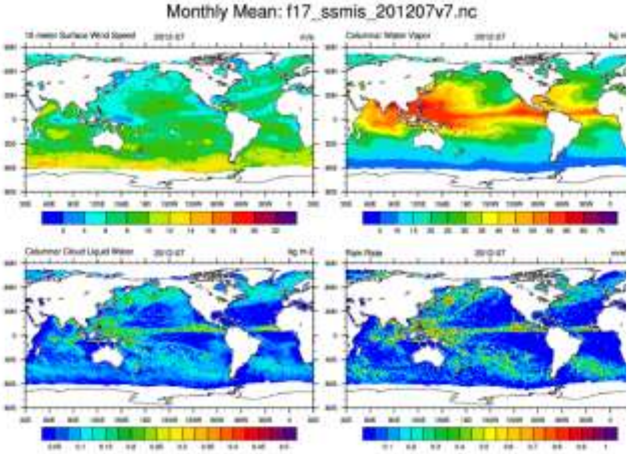
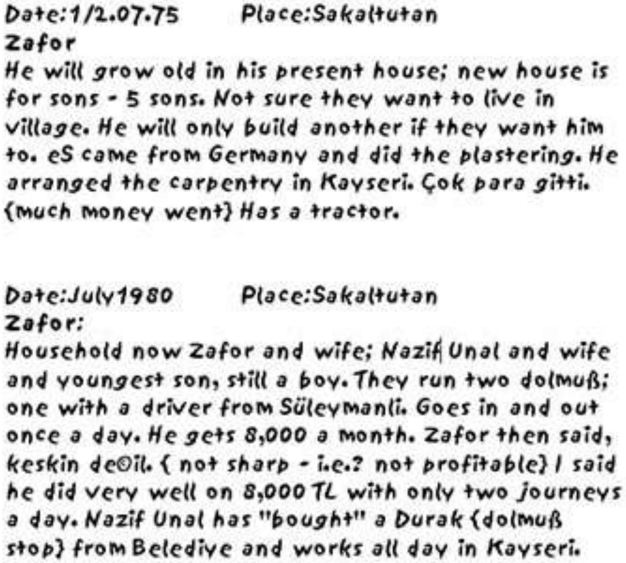


Figure 2. Numeric Change in Resident Population for the 50 States, the District of Columbia, and Puerto Rico: 1990 to 2000

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


Monthly Mean: f17_ssmis_201207v7.nc
10 meter Surface Wind Speed
2012-07
Columnar Water Vapor
2012-07
Columnar Cloud Liquid Water
2012-07
Rain Rate
2012-07

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.

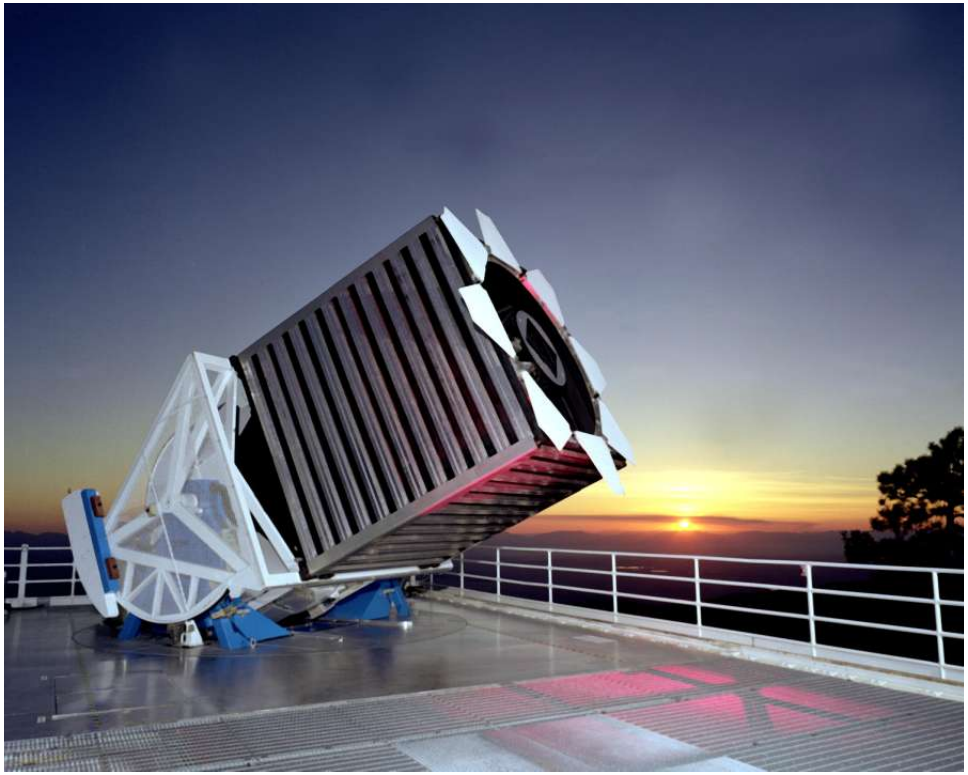
http://onlineqda.hud.ac.uk/Intro_QDA/Examples_of_Qualitative_Data.ph

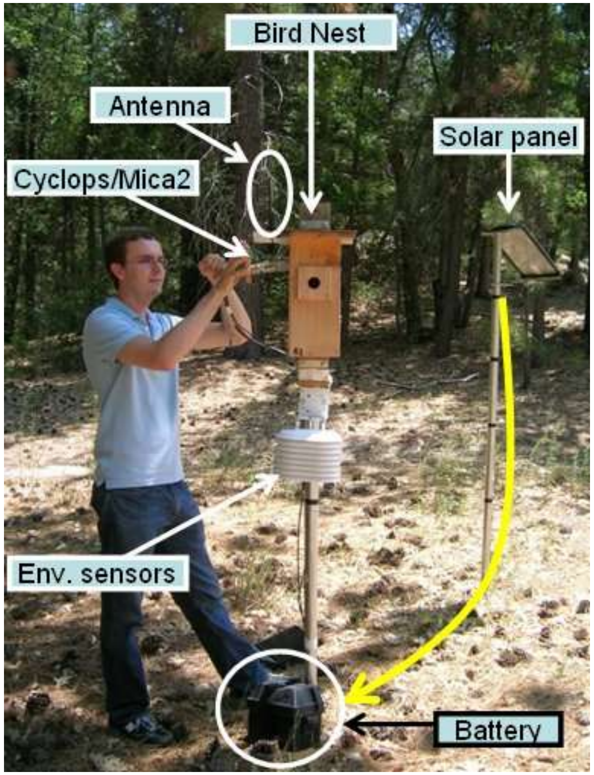
hudsonalpha.org

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

C.L. Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

Creating research data



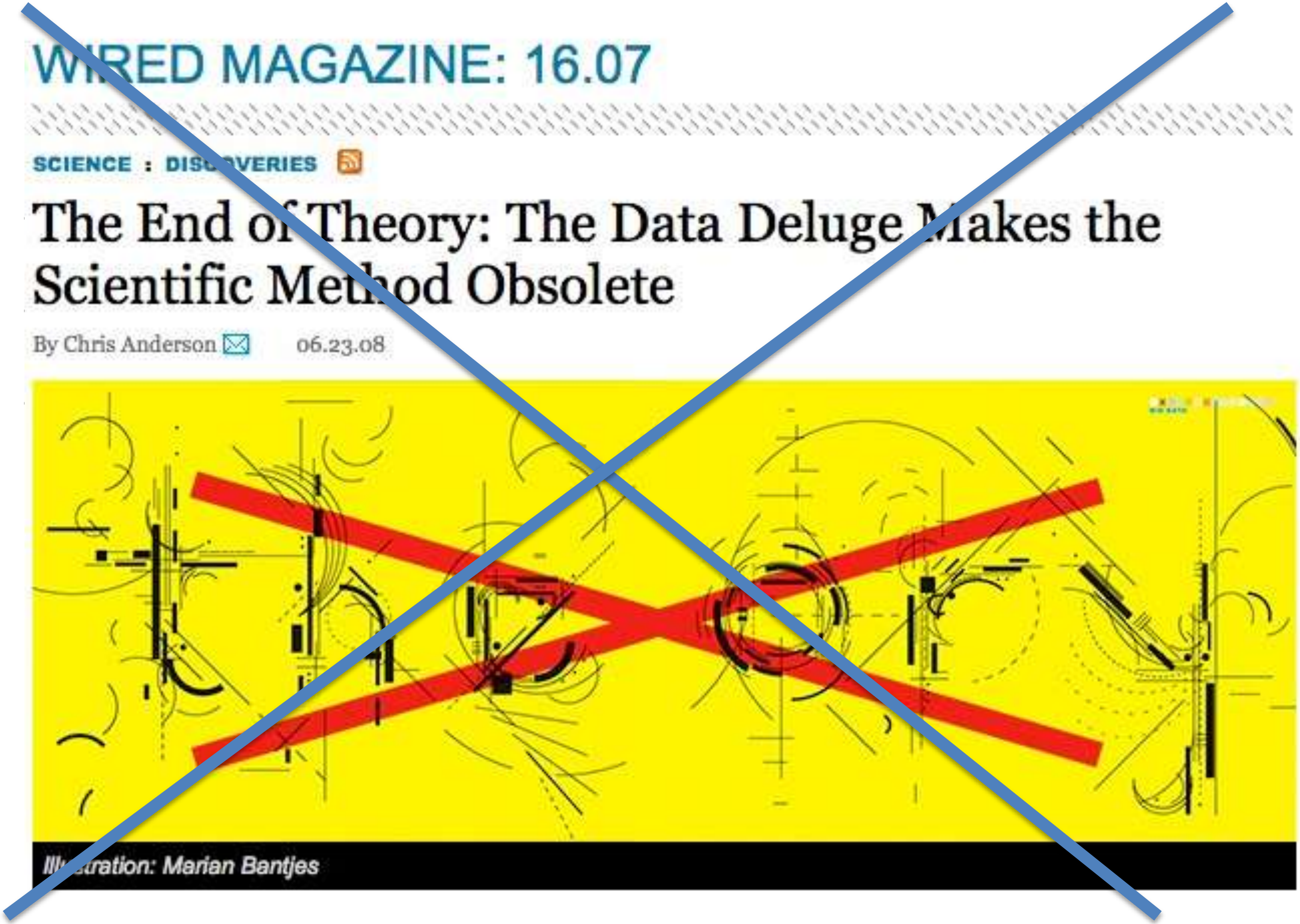


Sloan Digital Sky Survey Telescope,
Apache Point, New Mexico

Sensor networks

http://astro.uchicago.edu/~frieman/SDSS-telescope-photos/

http://enl.usc.edu/~jpaek/data/cyclops/bird_nest_2008/figures/nestbox2.jpg



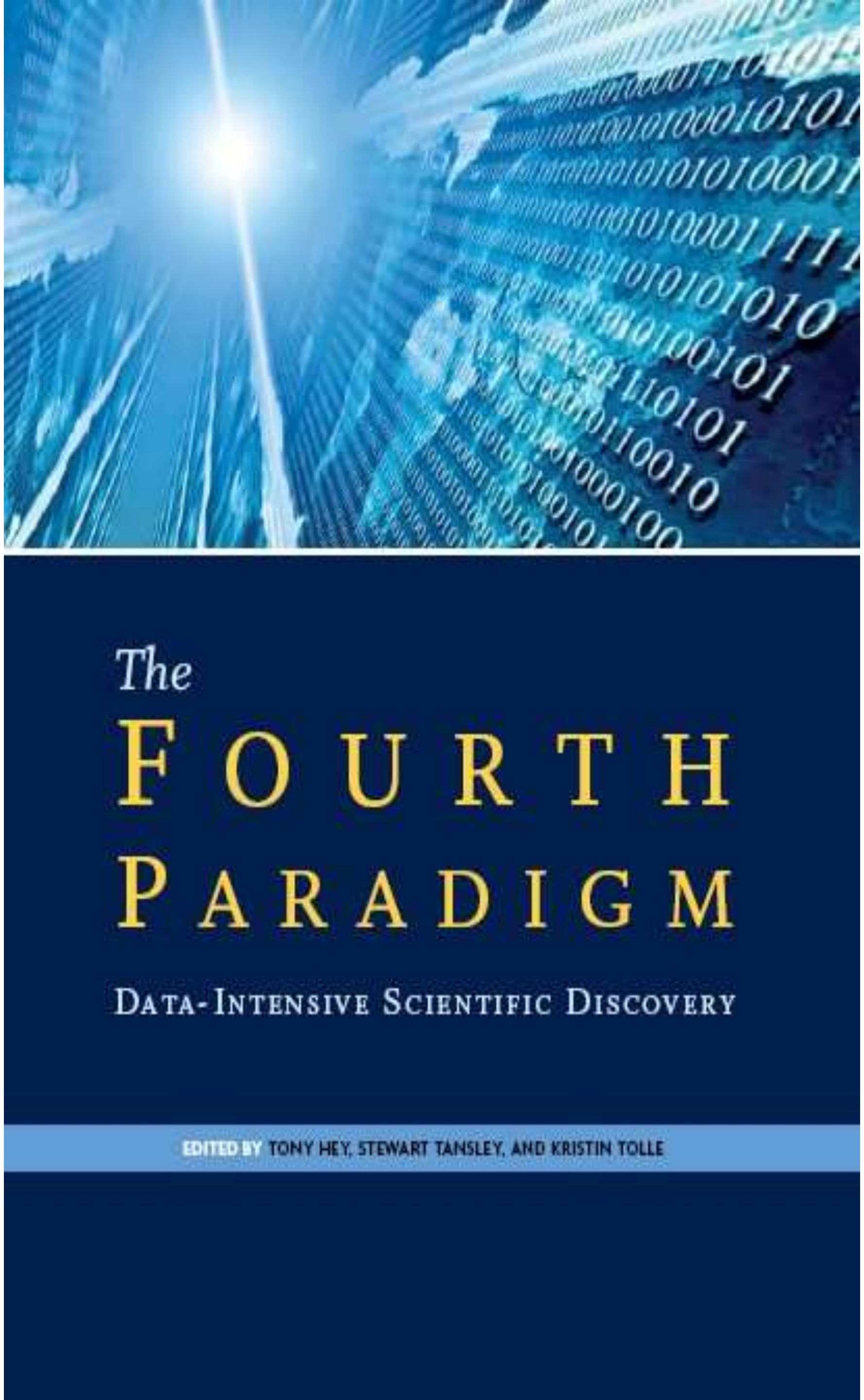
WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Illustration: Marian Bantjes

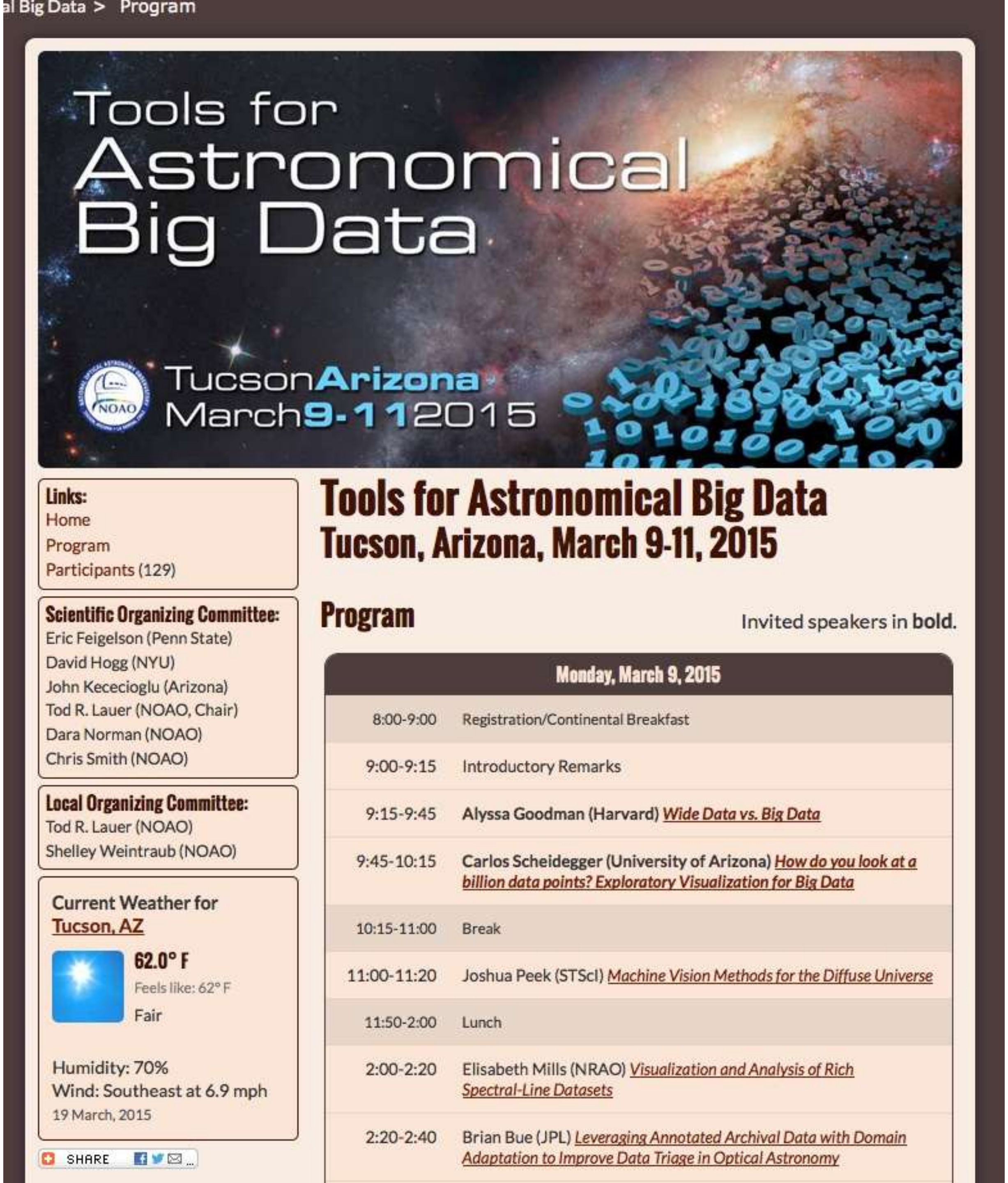


The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

Hey, Tansley & Tolle (eds.) (2009)



al Big Data > Program

Links:
Home
Program
Participants (129)

Scientific Organizing Committee:
Eric Feigelson (Penn State)
David Hogg (NYU)
John Kececioglu (Arizona)
Tod R. Lauer (NOAO, Chair)
Dara Norman (NOAO)
Chris Smith (NOAO)

Local Organizing Committee:
Tod R. Lauer (NOAO)
Shelley Weintraub (NOAO)

Current Weather for
Tucson, AZ

62.0° F
Feels like: 62° F
Fair

Humidity: 70%
Wind: Southeast at 6.9 mph
19 March, 2015

SHARE

Tools for Astronomical Big Data
Tucson, Arizona, March 9-11, 2015

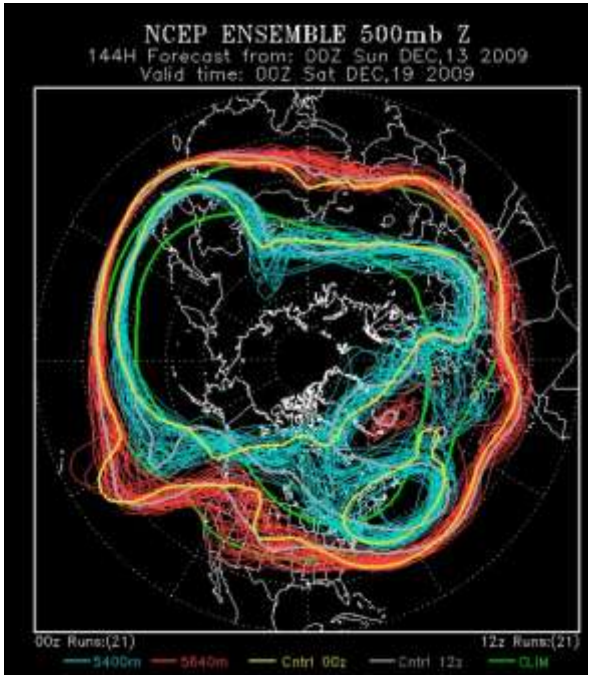
Program

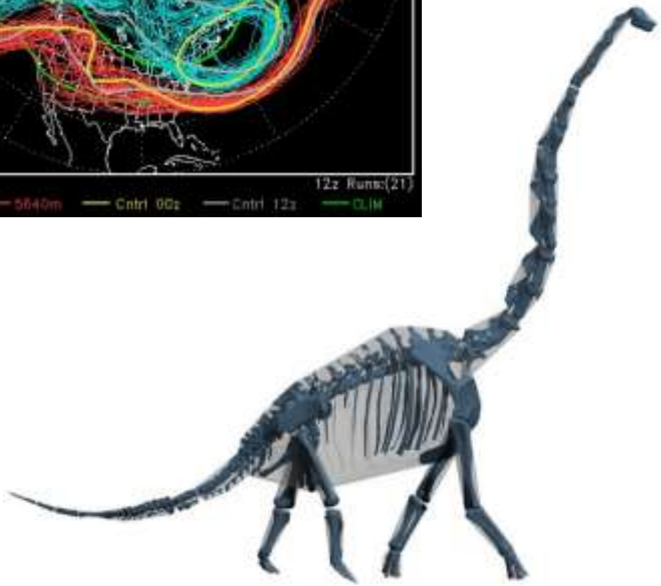
Invited speakers in **bold.**

Monday, March 9, 2015	
8:00-9:00	Registration/Continental Breakfast
9:00-9:15	Introductory Remarks
9:15-9:45	**Alyssa Goodman (Harvard)** *Wide Data vs. Big Data*
9:45-10:15	**Carlos Scheidegger (University of Arizona)** *How do you look at a billion data points? Exploratory Visualization for Big Data*
10:15-11:00	Break
11:00-11:20	Joshua Peek (STScI) *Machine Vision Methods for the Diffuse Universe*
11:50-2:00	Lunch
2:00-2:20	Elisabeth Mills (NRAO) *Visualization and Analysis of Rich Spectral-Line Datasets*
2:20-2:40	Brian Bue (JPL) *Leveraging Annotated Archival Data with Domain Adaptation to Improve Data Triage in Optical Astronomy*

Research process

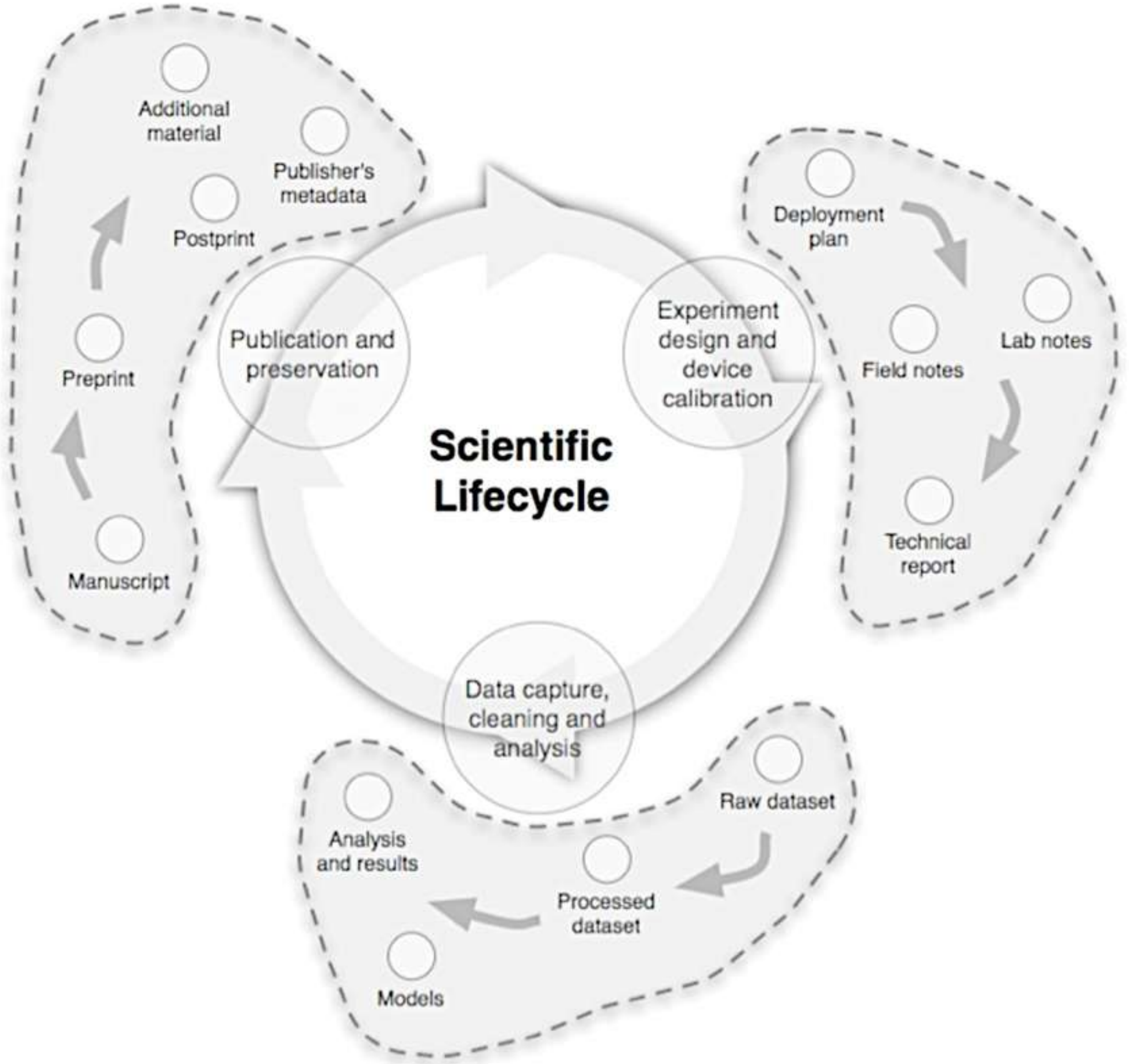
- Models and theories
- Research questions
- Methods
  - – Practices
  - – Data sources
  - – Software
  - – Instruments
  - – Infrastructure
  - – Domain expertise





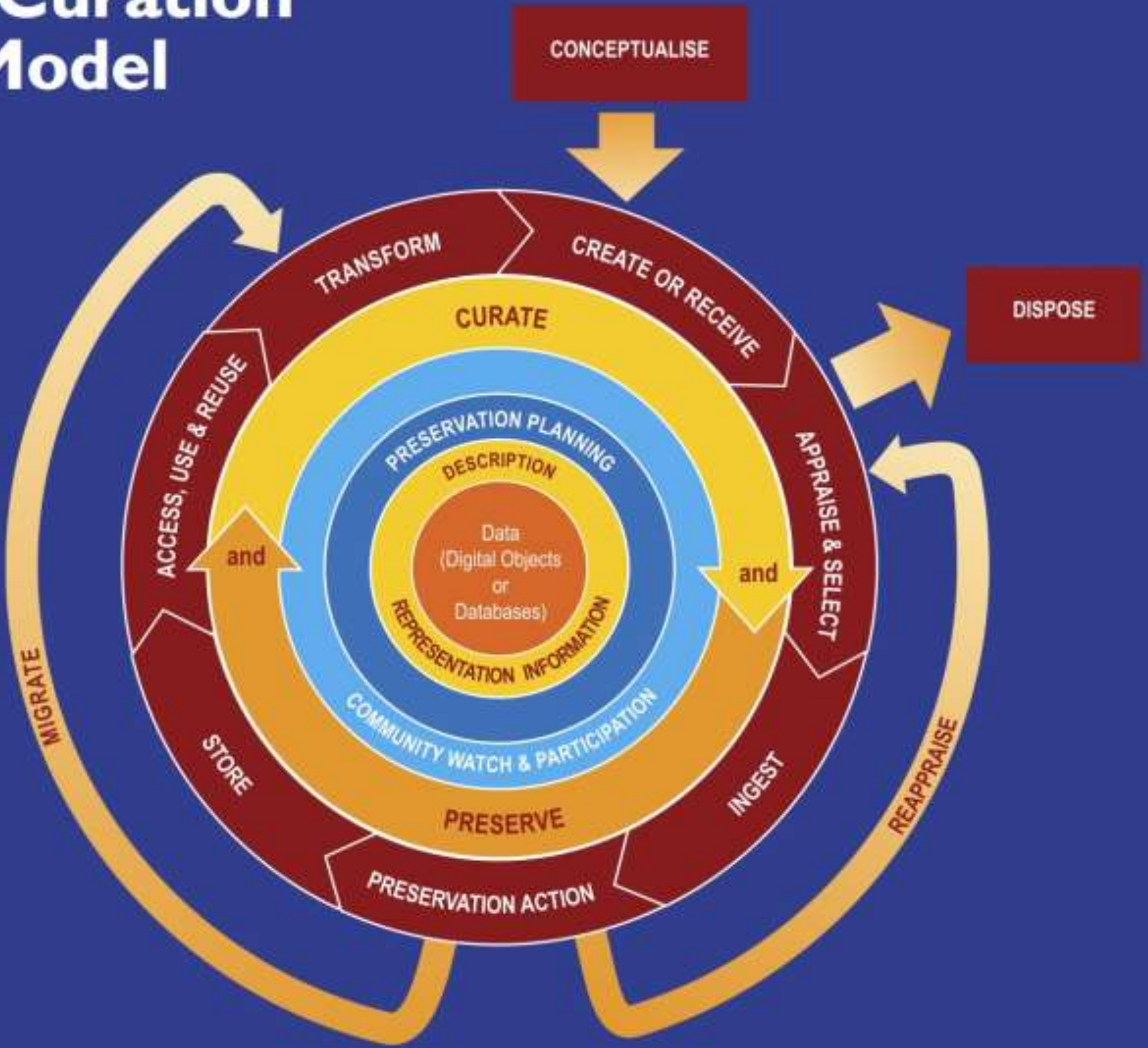
http://blogs.agu.org/wildwildscience/files/2009/12/Screen-shot-2009-12-13-at-04.10.146.jpg

http://www.livescience.com/20767-dinosaur-weight-estimates.html

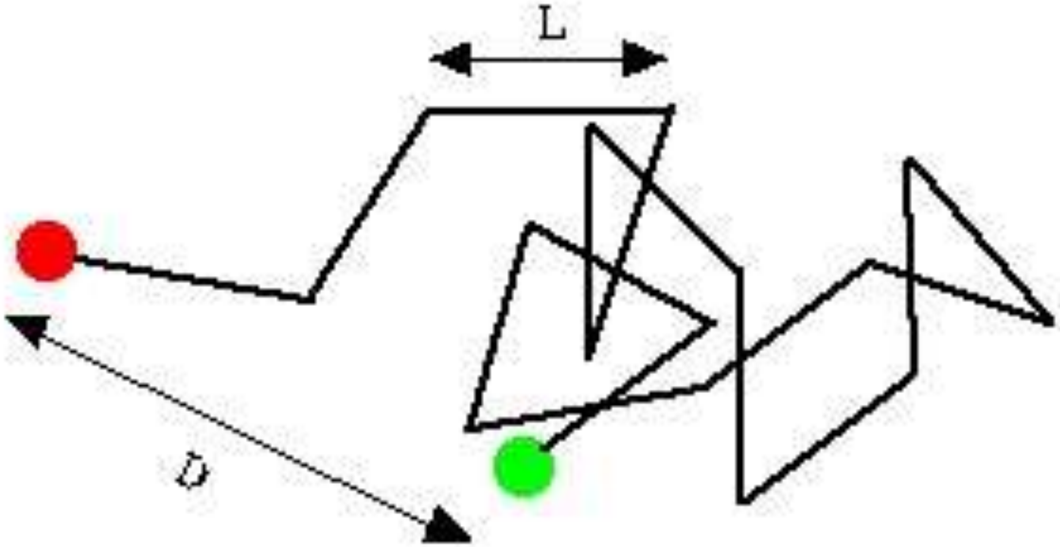


Pepe, A., Mayernik, M. S., Borgman, C. L. & Van de Sompel, H. (2010). From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. Journal of the American Society for Information Science and Technology, 61(3): 567–582.

The DCC Curation Lifecycle Model



Random walk



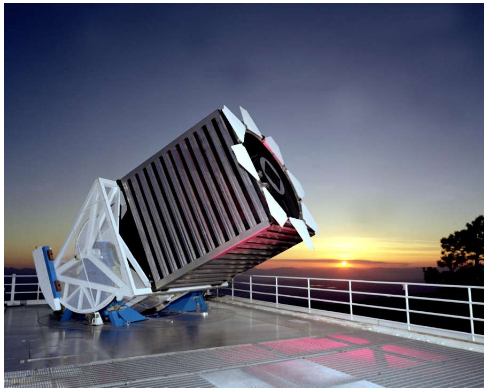
http://www2.ess.ucla.edu/~jewitt/oort2-random.html

Collaborating with data

http://digitalleaders.co.uk/wp-content/uploads/2014/09/collaboration.jpg

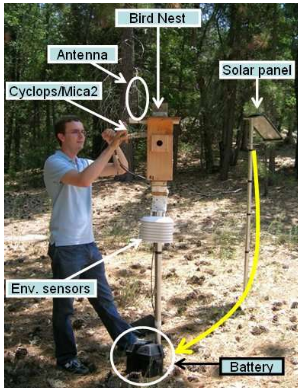
Big Science <–> Little Science

- Large instruments
- High cost
- Long duration
- Many collaborators
- Distributed work
- Domain expertise



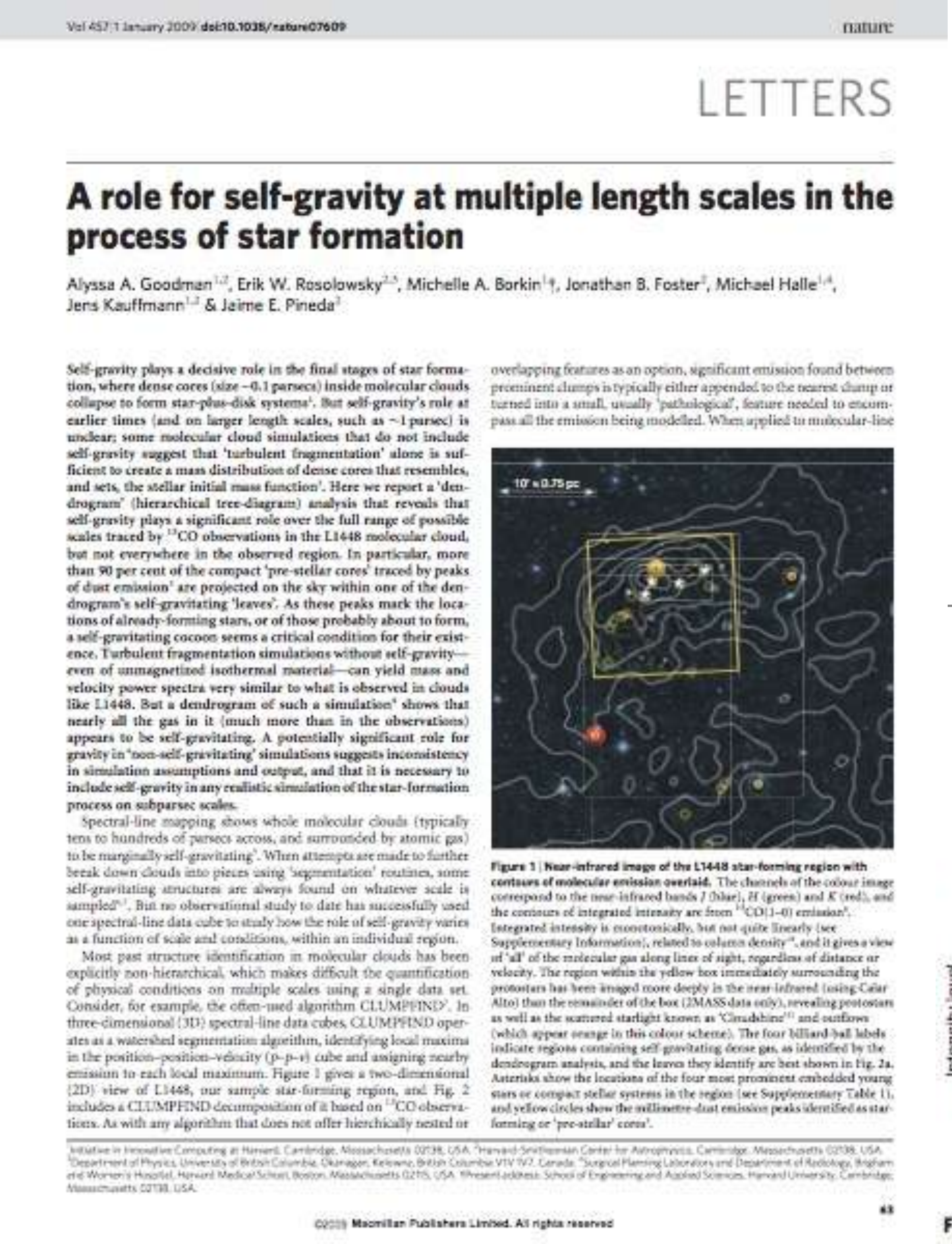
Sloan Digital Sky Survey

- Small instruments
- Low cost
- Short duration
- Small teams
- Local work
- Domain expertise



Sensor networks for science

Telescope for the Sloan Digital Sky Survey, Apache Point, New Mexico



Vol 457 | 1 January 2009 | doi:10.1038/nature07609

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]

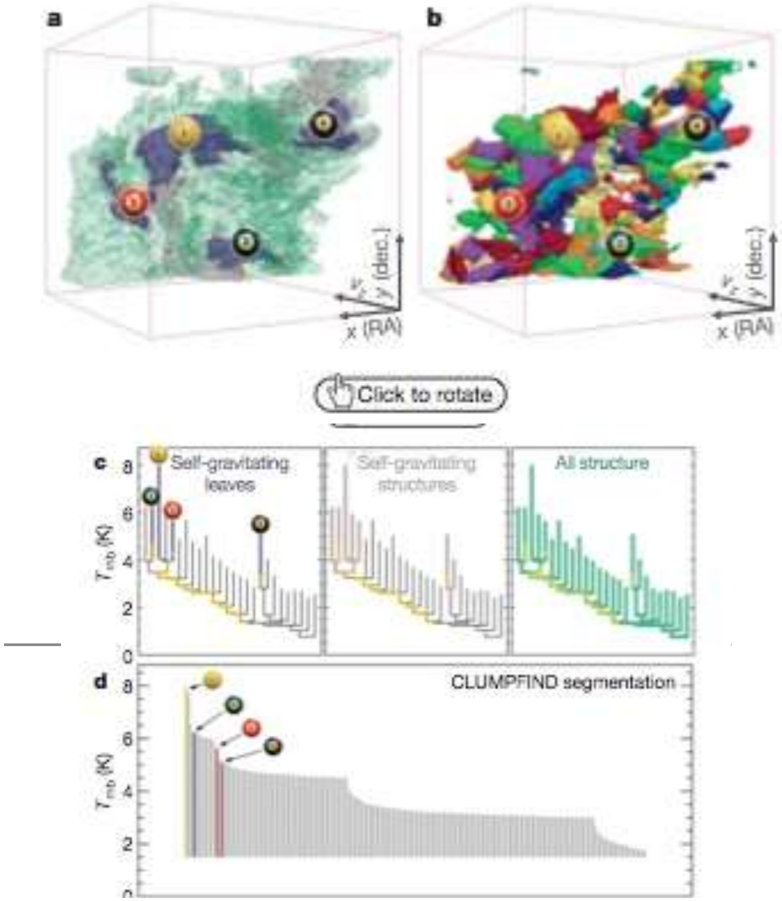
Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size ~0.1 parsecs) inside molecular clouds collapse to form star-plus-disk systems[1]. But self-gravity's role at earlier times (and on larger length scales, such as ~1 parsec) is unclear; some molecular cloud 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[9], 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'[10] 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 02115, USA. †Present address: School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.



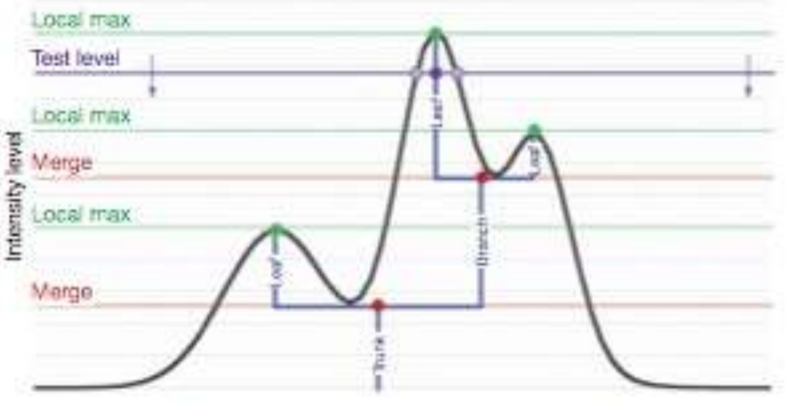
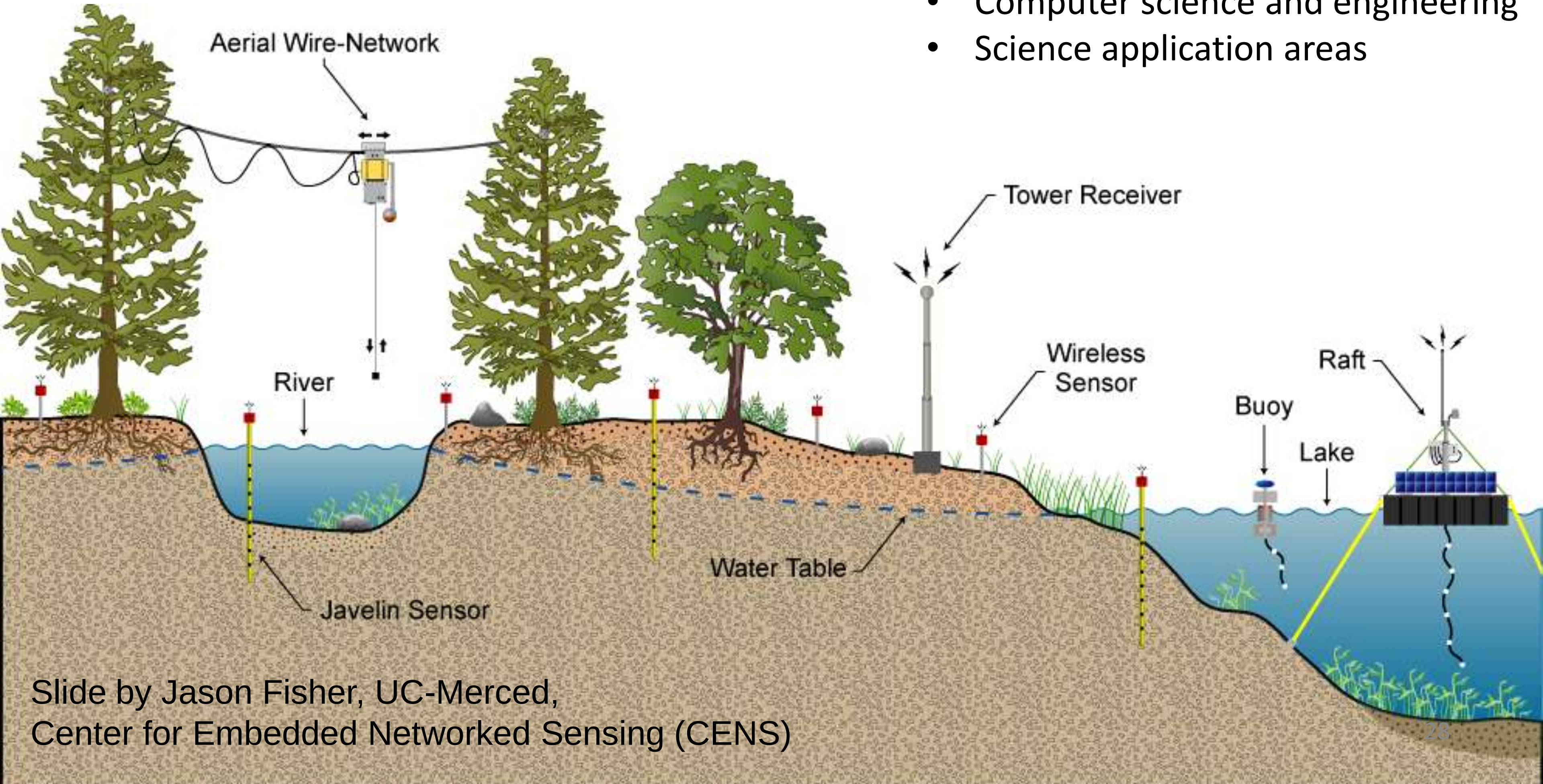


Figure 3 | Schematic illustration of the dendrogram process. Shown is the

Center for Embedded Networked Sensing

- NSF Science & Tech Ctr, 2002-2012
- 5 universities, plus partners
- 300 members
- Computer science and engineering
- Science application areas



Slide by Jason Fisher, UC-Merced,
Center for Embedded Networked Sensing (CENS)

Science <–> Data

Engineering researcher: ***"Temperature is temperature."***

CENS Robotics team

Biologist*: **"There are hundreds of ways to measure temperature.** 'The temperature is 98' is low-value compared to, 'the temperature of the surface, measured by the infrared thermopile, model number XYZ, is 98.' That means it is measuring a proxy for a temperature, rather than being in contact with a probe, and it is measuring from a distance. The accuracy is plus or minus .05 of a degree. I [also] want to know that it was taken outside versus inside a controlled environment, how long it had been in place, and the last time it was calibrated, which might tell me whether it has drifted.."*

Center for Dark Energy Biosphere Investigations

International Ocean Discovery Program
Iodp.tamu.org

Repository for seafloor cores. Photo: Peter Darch

- NSF Science & Tech Ctr, 2010-2020
- 20 universities, plus partners (35 institutions)
- 90 scientists
- Biological sciences
- Physical sciences

Self-descriptions C-DEBI scientists

Aquatic chemistry

Aquatic microbial ecology

Astrobiology

Biochemistry

Bioenergetics

Biogeochemistry

Biogeography

Bioinformatics

Biology

Chemical ecology

Chemical oceanography

Deep sea biogeochemistry

Deep sea microbiology

Ecology

Ecophysiology

Environmental chemistry

Environmental microbiology

Genomics

Geobiology

Geochemistry

Geomicrobiology

Geophysics

Hydrogeology

Hydrology

Hydrothermal microbiology

Inorganic chemistry

Marine ecology

Marine geology

Marine microbial biogeochemistry

Marine microbial ecology

Marine microbiology

Metabolomics

Metagenomics

Microbial biogeochemistry

Microbial biogeography

Microbial ecology

Microbial oceanography

Microbial physiology

Microbiology

Mineralogy

Molecular biogeochemistry

Molecular biology

Molecular microbial ecology

Molecular microbiology

Molecular physiology

Paleoceanography

Paleoclimatology

Paleogeomicrobiology

Petrology

Physiology

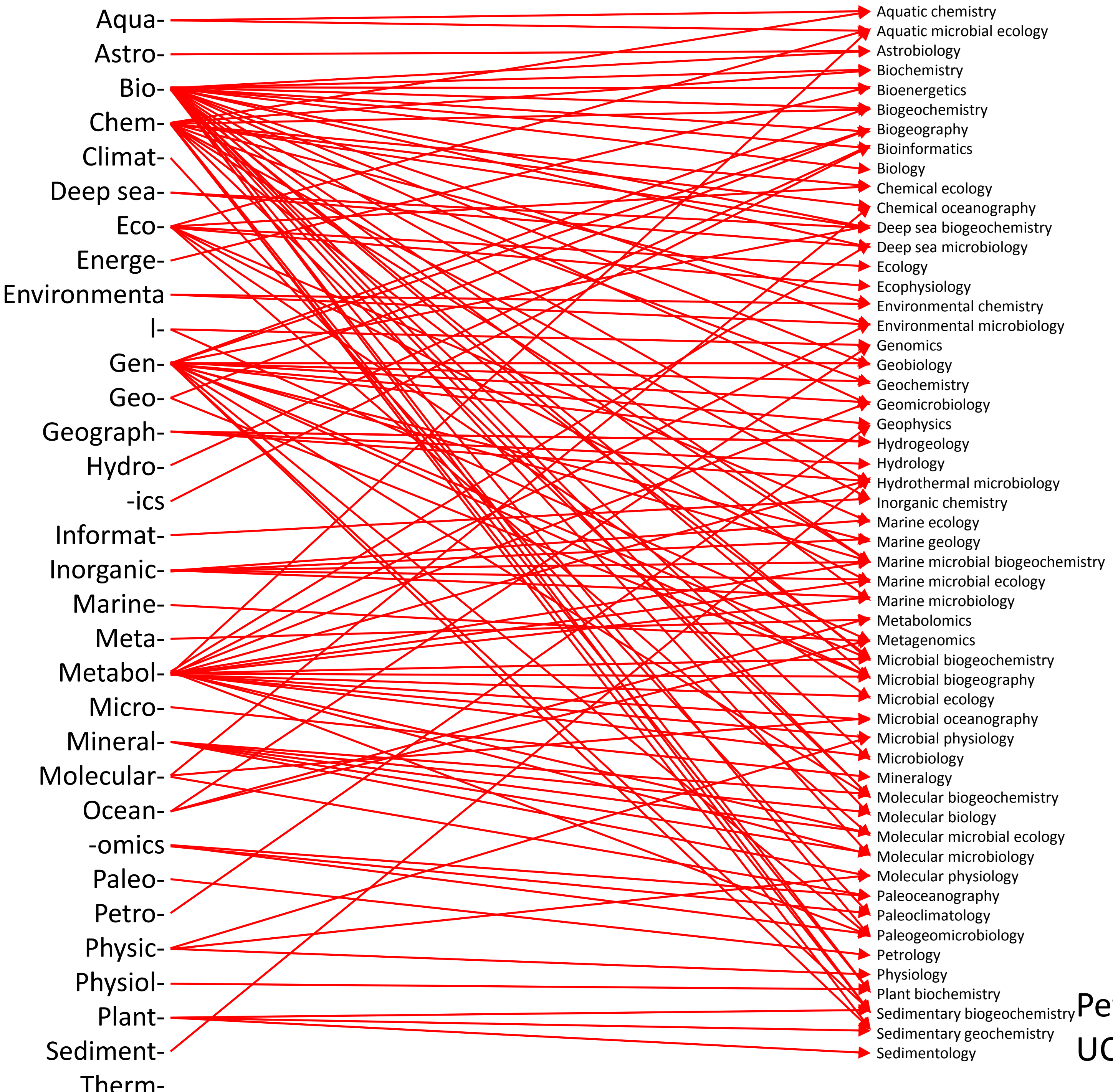
Plant biochemistry

Sedimentary biogeochemistry

Sedimentary geochemistry

Sedimentology

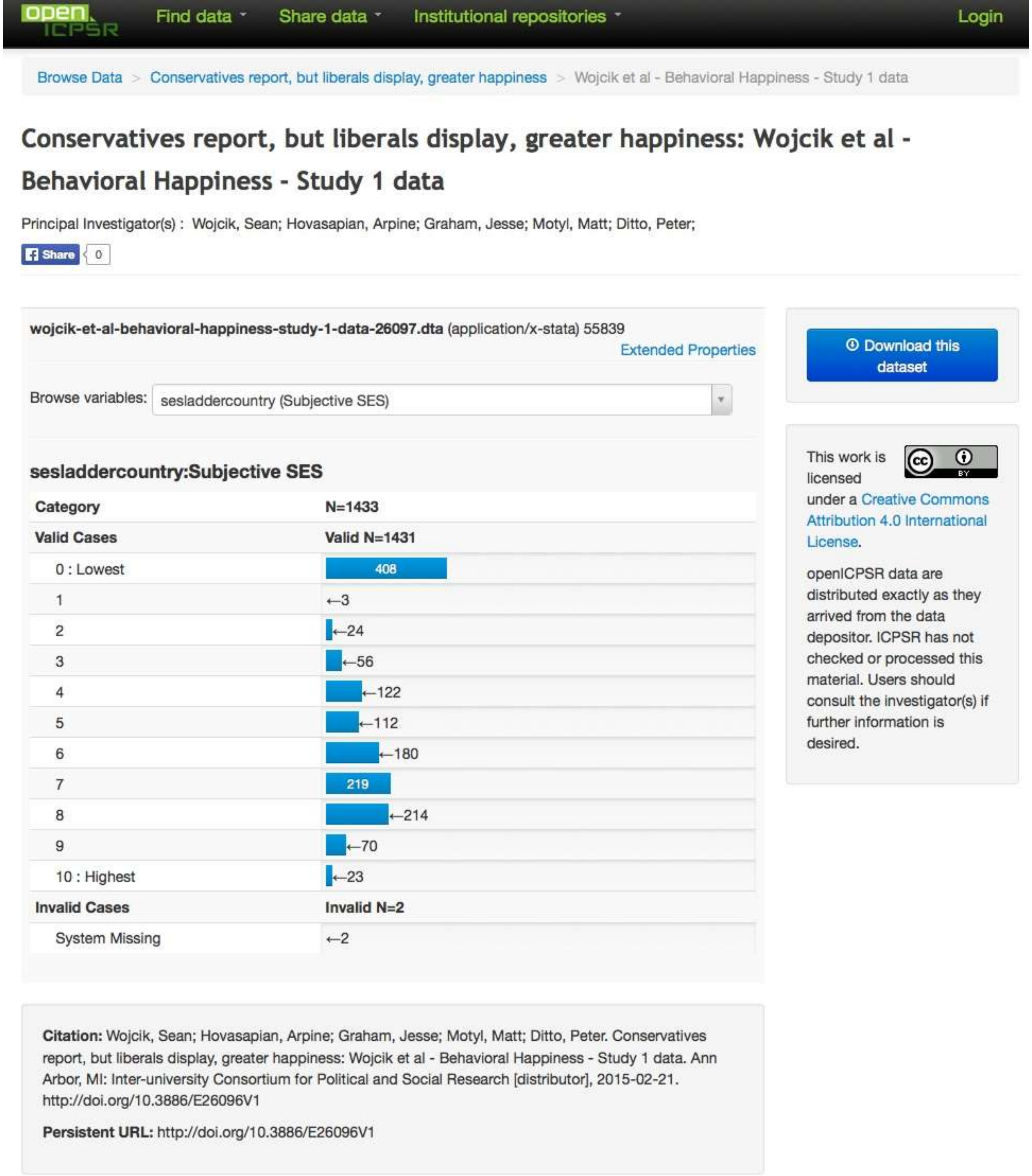
Peter Darch,
UCLA



Aqua-
Astro-
Bio-
Chem-
Climat-
Deep sea-
Eco-
Energe-
Environmental-
Gen-
Geo-
Geograph-
Hydro-
-ics
Informat-
Inorganic-
Marine-
Meta-
Metabol-
Micro-
Mineral-
Molecular-
Ocean-
-omics
Paleo-
Petro-
Physic-
Physiol-
Plant-
Sediment-
Therm-

Aquatic chemistry
Aquatic microbial ecology
Astrobiology
Biochemistry
Bioenergetics
Biogeochemistry
Biogeography
Bioinformatics
Biology
Chemical ecology
Chemical oceanography
Deep sea biogeochemistry
Deep sea microbiology
Ecology
Ecophysiology
Environmental chemistry
Environmental microbiology
Genomics
Geobiology
Geochemistry
Geomicrobiology
Geophysics
Hydrogeology
Hydrology
Hydrothermal microbiology
Inorganic chemistry
Marine ecology
Marine geology
Marine microbial biogeochemistry
Marine microbial ecology
Marine microbiology
Metabolomics
Metagenomics
Microbial biogeochemistry
Microbial biogeography
Microbial ecology
Microbial oceanography
Microbial physiology
Microbiology
Mineralogy
Molecular biogeochemistry
Molecular biology
Molecular microbial ecology
Molecular microbiology
Molecular physiology
Paleoceanography
Paleoclimatology
Paleogeomicrobiology
Petrology
Physiology
Plant biochemistry
Sedimentary biogeochemistry
Sedimentary geochemistry
Sedimentology

Peter Darch, UCLA


openICPSR
Find data
Share data
Institutional repositories
Login
Browse Data > Conservatives report, but liberals display, greater happiness > Wojcik et al - Behavioral Happiness - Study 1 data
Share 0
Download this dataset
Extended Properties

Conservatives report, but liberals display, greater happiness: Wojcik et al - Behavioral Happiness - Study 1 data

Principal Investigator(s) : Wojcik, Sean; Hovasapian, Arpine; Graham, Jesse; Motyl, Matt; Ditto, Peter;

wojcik-et-al-behavioral-happiness-study-1-data-26097.dta (application/x-stata) 55839

Browse variables: sesladdercountry (Subjective SES)

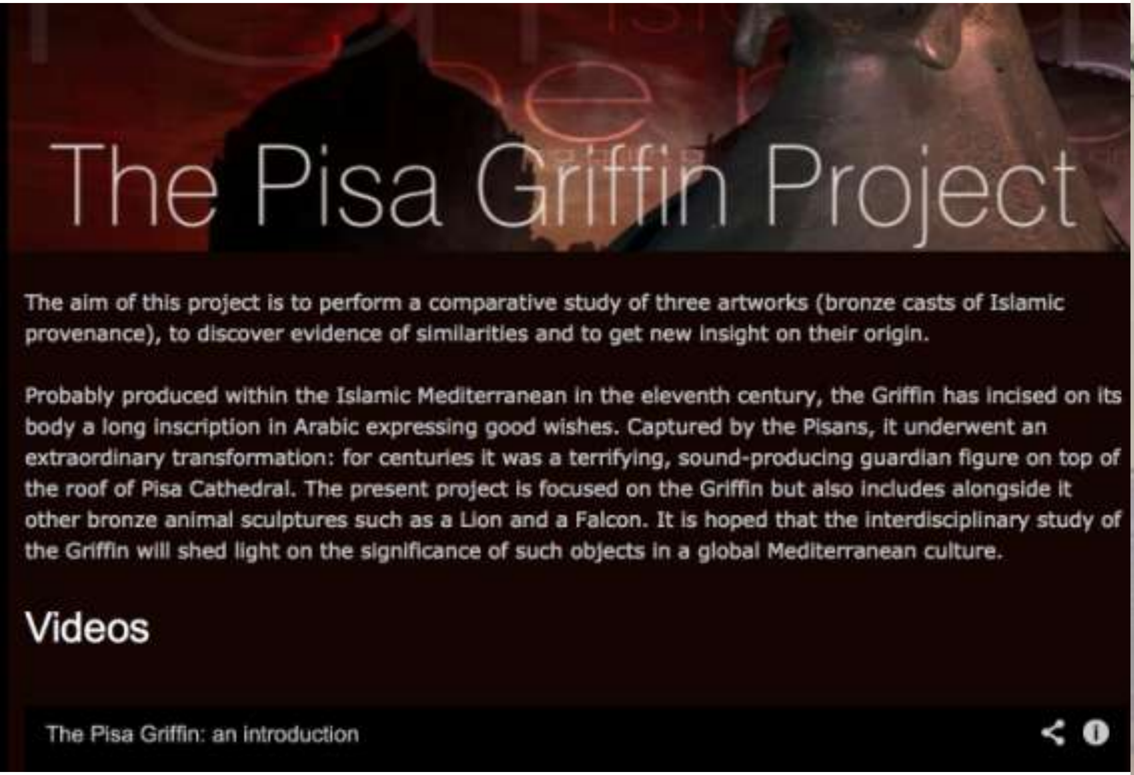
sesladdercountry:Subjective SES

Category	**N=1433**
Valid Cases	**Valid N=1431**
0 : Lowest	408
1	←3
2	←24
3	←56
4	←122
5	←112
6	←180
7	219
8	←214
9	←70
10 : Highest	←23
Invalid Cases	**Invalid N=2**
System Missing	←2

openICPSR data are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users should consult the investigator(s) if further information is desired.

Citation: Wojcik, Sean; Hovasapian, Arpine; Graham, Jesse; Motyl, Matt; Ditto, Peter. Conservatives report, but liberals display, greater happiness: Wojcik et al - Behavioral Happiness - Study 1 data. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-02-21. http://doi.org/10.3886/E26096V1

Persistent URL: http://doi.org/10.3886/E26096V1



The Pisa Griffin Project

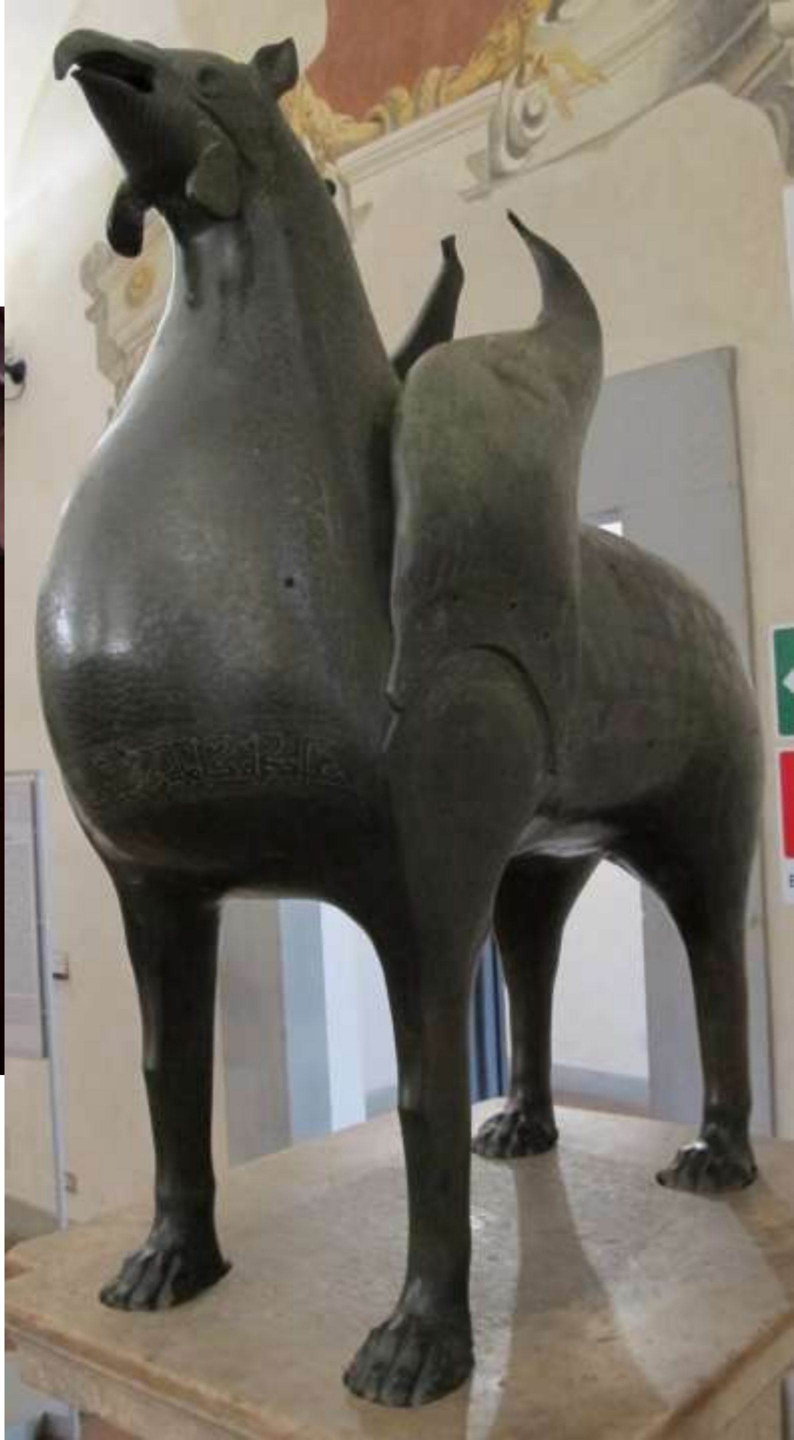
The aim of this project is to perform a comparative study of three artworks (bronze casts of Islamic provenance), to discover evidence of similarities and to get new insight on their origin.

Probably produced within the Islamic Mediterranean in the eleventh century, the Griffin has incised on its body a long inscription in Arabic expressing good wishes. Captured by the Pisans, it underwent an extraordinary transformation: for centuries it was a terrifying, sound-producing guardian figure on top of the roof of Pisa Cathedral. The present project is focused on the Griffin but also includes alongside it other bronze animal sculptures such as a Lion and a Falcon. It is hoped that the interdisciplinary study of the Griffin will shed light on the significance of such objects in a global Mediterranean culture.

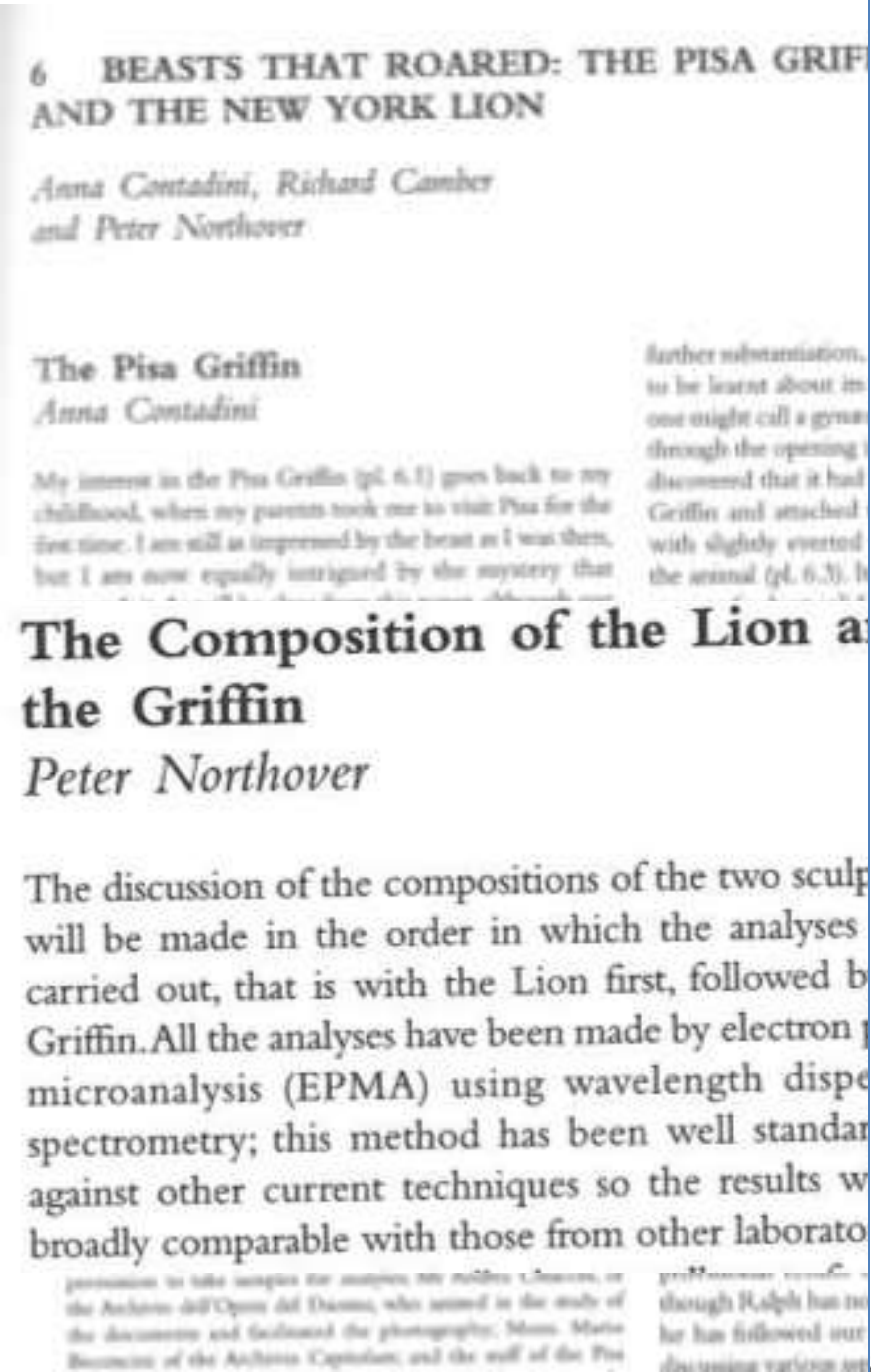
Videos

The Pisa Griffin: an introduction

http://vcg.isti.cnr.it/griffin/



Arte islamica, ippogrifo, XI sec 03, own work



6 BEASTS THAT ROARED: THE PISA GRIF AND THE NEW YORK LION

Anna Contadini, Richard Camber and Peter Northover

The Pisa Griffin

Anna Contadini

My interest in the Pisa Griffin (pl. 6.1) goes back to my childhood, when my parents took me to visit Pisa for the first time. I am still as impressed by the beast as I was then, but I am now equally intrigued by the mystery that

further substantiation, to be learnt about its one might call a gynae through the opening discovered that it had Griffin and attached with slightly everted the animal (pl. 6.3). It

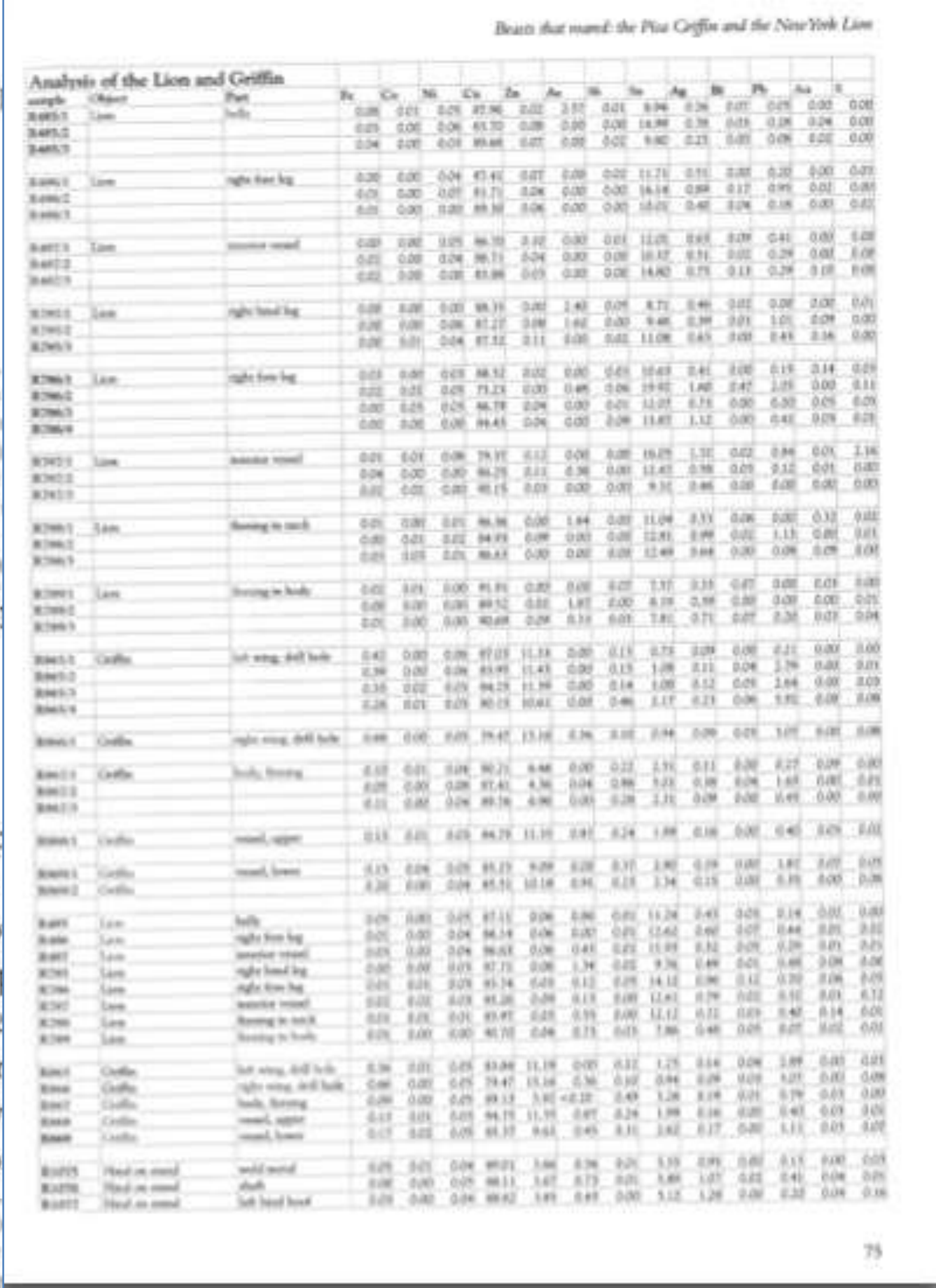
The Composition of the Lion a the Griffin

Peter Northover

The discussion of the compositions of the two sculp will be made in the order in which the analyses carried out, that is with the Lion first, followed b Griffin. All the analyses have been made by electron microanalysis (EPMA) using wavelength dispe spectrometry; this method has been well standar against other current techniques so the results w broadly comparable with those from other laborato

permission to take samples for analysis; the Archivio dell'Opera del Duomo, who assisted in the study of the documents and facilitated the photography; Mons. Mario Bocuccini of the Archivio Capitolare; and the staff of the Pisa

though Ralph has no he has followed our discussing various asp



Beasts that roared: the Pisa Griffin and the New York Lion

Analysis of the Lion and Griffin

sample	Object	Part	Fe	Co	Ni	Cu	Zn	As	Sb	Sn	Ag	Bi	Pb	Au	S
B4853.1	Lion	belly	0.08	0.01	0.05	87.96	0.02	2.57	0.01	8.94	0.26	0.07	0.05	0.00	0.00
B4853.2			0.03	0.00	0.06	85.70	0.08	0.00	0.00	14.89	0.39	0.03	0.18	0.24	0.00
B4853.3			0.04	0.07	0.03	89.66	0.07	0.00	0.02	9.80	0.23	0.05	0.06	0.02	0.00
B4861.1	Lion	right fore leg	0.20	0.00	0.04	87.41	0.07	0.09	0.02	11.71	0.51	0.05	0.20	0.00	0.03
B4861.2			0.01	0.00	0.07	81.71	0.26	0.00	0.00	16.14	0.89	0.17	0.95	0.02	0.00
B4861.3			0.01	0.00	0.00	89.50	0.06	0.00	0.00	10.01	0.40	0.06	0.18	0.00	0.02
B4872.1	Lion	interior vessel	0.00	0.00	0.05	86.70	3.02	0.00	0.01	12.05	0.65	0.09	0.41	0.00	1.00
B4872.2			0.01	0.00	0.04	86.71	0.24	0.00	0.00	10.17	0.51	0.02	0.29	0.00	1.08
B4872.3			0.02	0.00	0.00	83.86	0.03	0.00	0.00	14.80	0.75	0.11	0.29	0.10	0.08
R2905.1	Lion	right hind leg	0.08	0.00	0.00	86.35	0.00	1.40	0.05	8.71	0.46	0.02	0.08	0.00	0.01
R2905.2			0.00	0.00	0.06	87.27	0.08	1.62	0.00	9.46	0.39	0.01	1.01	0.09	0.00
R2905.3			0.00	0.01	0.04	87.32	0.11	0.00	0.02	11.08	0.65	0.00	0.45	0.16	0.00
R2964.1	Lion	right fore leg	0.03	0.00	0.05	88.52	0.02	0.00	0.05	10.63	0.41	0.00	0.15	0.14	0.09
R2964.2			0.02	0.05	0.05	73.21	0.00	0.46	0.06	19.92	1.60	0.47	2.05	0.00	0.11
R2964.3			0.00	0.05	0.05	86.79	0.24	0.00	0.01	12.05	0.71	0.00	0.50	0.05	0.05
R2964.4			0.00	0.00	0.00	86.45	0.06	0.00	2.06	11.87	1.12	0.00	0.41	0.05	0.05
R2972.1	Lion	interior vessel	0.01	0.01	0.06	79.37	0.12	0.00	0.05	16.05	1.31	0.02	0.84	0.01	1.16
R2972.2			0.04	0.00	0.00	86.25	0.11	0.36	0.00	11.43	0.56	0.05	0.12	0.01	0.00
R2972.3			0.02	0.02	0.00	90.15	0.03	0.00	0.00	9.31	0.46	0.00	0.00	0.00	0.00
R2966.1	Lion	flowing in neck	0.01	0.00	0.01	86.36	0.00	1.04	0.00	11.09	0.53	0.06	0.00	0.12	0.02
R2966.2			0.00	0.01	0.02	84.93	0.09	0.00	0.00	12.81	0.99	0.02	1.15	0.00	0.01
R2966.3			0.05	0.05	0.01	86.63	0.00	0.00	0.00	12.49	0.66	0.00	0.06	0.05	0.00
R2991.1	Lion	flowing in body	0.02	0.01	0.00	91.91	0.00	0.00	0.02	7.37	0.35	0.07	0.00	0.03	0.00
R2991.2			0.00	0.00	0.00	89.52	0.02	1.07	0.00	8.19	0.36	0.00	0.00	0.00	0.01
R2991.3			0.01	0.00	0.00	90.69	0.24	0.33	0.01	7.81	0.71	0.07	0.20	0.07	0.04
B4865.1	Griffin	left wing, drill hole	0.42	0.00	0.06	87.03	11.18	0.00	0.15	0.73	0.09	0.00	0.21	0.00	0.00
B4865.2			0.39	0.00	0.06	83.95	11.43	0.00	0.15	1.08	0.11	0.04	2.39	0.00	0.01
B4865.3			0.30	0.02	0.05	84.25	11.39	0.00	0.16	1.05	0.12	0.05	2.64	0.00	0.03
B4865.4			0.28	0.01	0.05	80.15	10.61	0.00	0.46	1.17	0.13	0.06	5.92	0.08	0.08
B4866.1	Griffin	right wing, drill hole	0.08	0.00	0.05	79.47	15.16	0.36	0.02	0.84	0.09	0.03	3.07	0.00	0.08
B4867.1	Griffin	body, flowing	0.15	0.01	0.08	90.21	4.46	0.00	0.22	2.31	0.11	0.00	2.27	0.09	0.00
B4867.2			0.05	0.00	0.08	87.41	4.36	0.04	0.86	3.23	0.18	0.08	1.65	0.00	0.01
B4867.3			0.11	0.02	0.04	89.36	4.90	0.00	0.28	2.31	0.09	0.00	0.45	0.00	0.00
B4868.1	Griffin	vessel, upper	0.13	0.01	0.03	84.75	11.35	0.85	0.24	1.88	0.16	0.00	0.40	0.05	0.02
B4869.1	Griffin	vessel, lower	0.15	0.04	0.05	85.23	9.26	0.00	0.37	2.80	0.19	0.00	1.82	0.07	0.05
B4869.2	Griffin		0.20	0.00	0.04	85.51	10.18	0.96	0.23	1.34	0.15	0.00	0.39	0.00	0.06
B4853	Lion	belly	0.05	0.03	0.05	87.11	0.06	0.86	0.01	11.24	0.43	0.05	0.16	0.02	0.00
B4861	Lion	right fore leg	0.01	0.00	0.04	86.14	0.06	0.00	0.01	12.62	0.60	0.07	0.44	0.01	0.02
B4872	Lion	interior vessel	0.03	0.00	0.04	86.63	0.06	0.41	0.02	11.93	0.52	0.05	0.28	0.01	0.05
R2905	Lion	right hind leg	0.00	0.00	0.05	87.71	0.06	1.34	0.02	9.76	0.49	0.01	0.48	0.08	0.00
R2964	Lion	right fore leg	0.01	0.05	0.05	85.34	0.03	0.12	0.55	14.12	0.96	0.12	0.55	0.06	0.05
R2972	Lion	interior vessel	0.02	0.01	0.03	85.26	0.09	0.12	0.05	12.61	0.79	0.02	0.32	0.01	0.32
R2966	Lion	flowing in neck	0.01	0.02	0.01	85.97	0.03	0.35	0.00	12.12	0.72	0.03	0.40	0.14	0.01
R2991	Lion	flowing in body	0.01	0.00	0.00	90.70	0.06	0.33	0.01	7.86	0.48	0.05	0.07	0.02	0.02
B4865	Griffin	left wing, drill hole	0.36	0.01	0.05	83.84	11.15	0.00	0.22	1.25	0.16	0.04	2.89	0.03	0.03
B4866	Griffin	right wing, drill hole	0.08	0.00	0.05	79.47	15.16	0.36	0.02	0.84	0.09	0.03	3.07	0.00	0.08
B4867	Griffin	body, flowing	0.09	0.00	0.05	89.13	3.81	<0.12	0.49	1.28	0.19	0.03	0.79	0.03	0.00
B4868	Griffin	vessel, upper	0.13	0.01	0.03	84.75	11.35	0.87	0.24	1.88	0.16	0.00	0.40	0.05	0.02
B4869	Griffin	vessel, lower	0.17	0.02	0.05	85.37	9.63	0.45	0.31	2.62	0.17	0.00	1.11	0.03	0.07
B1055	Hind on stand	solid metal	0.05	0.01	0.04	89.01	3.66	0.36	0.01	5.53	0.09	0.02	0.13	0.00	0.05
B1056	Hind on stand	shaft	0.08	0.00	0.05	88.11	3.67	0.73	0.01	5.88	1.07	0.02	0.41	0.04	0.05
B1057	Hind on stand	left hind hoof	0.01	0.00	0.04	88.62	3.85	0.45	0.00	5.12	1.28	0.00	0.22	0.04	0.16

75

RDA

RESEARCH DATA ALLIANCE

Research Data Sharing
without barriers

Precondition:

Researchers share data

Lack of incentives to share data

- Labor to document data
- Benefits to unknown others
- Competition
- Control
- Confidentiality…

Image source: www.buildingsrus.co.uk/.../ target1.htm

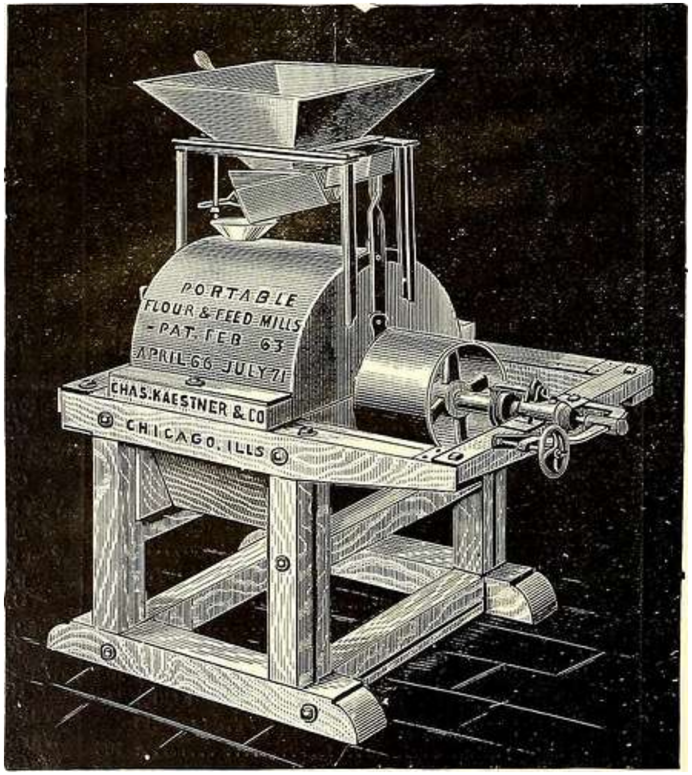
Lack of incentives to reuse data

- Identify useful data
  - Documentation
  - Interpretation
  - Software
- Cleaning
- Trust
- Credit
- Licensing…

http://fyi.uiowa.edu/wp-content/uploads/2011/10/utopia_in_four_movements_filmstill5_utopiasign.jpg

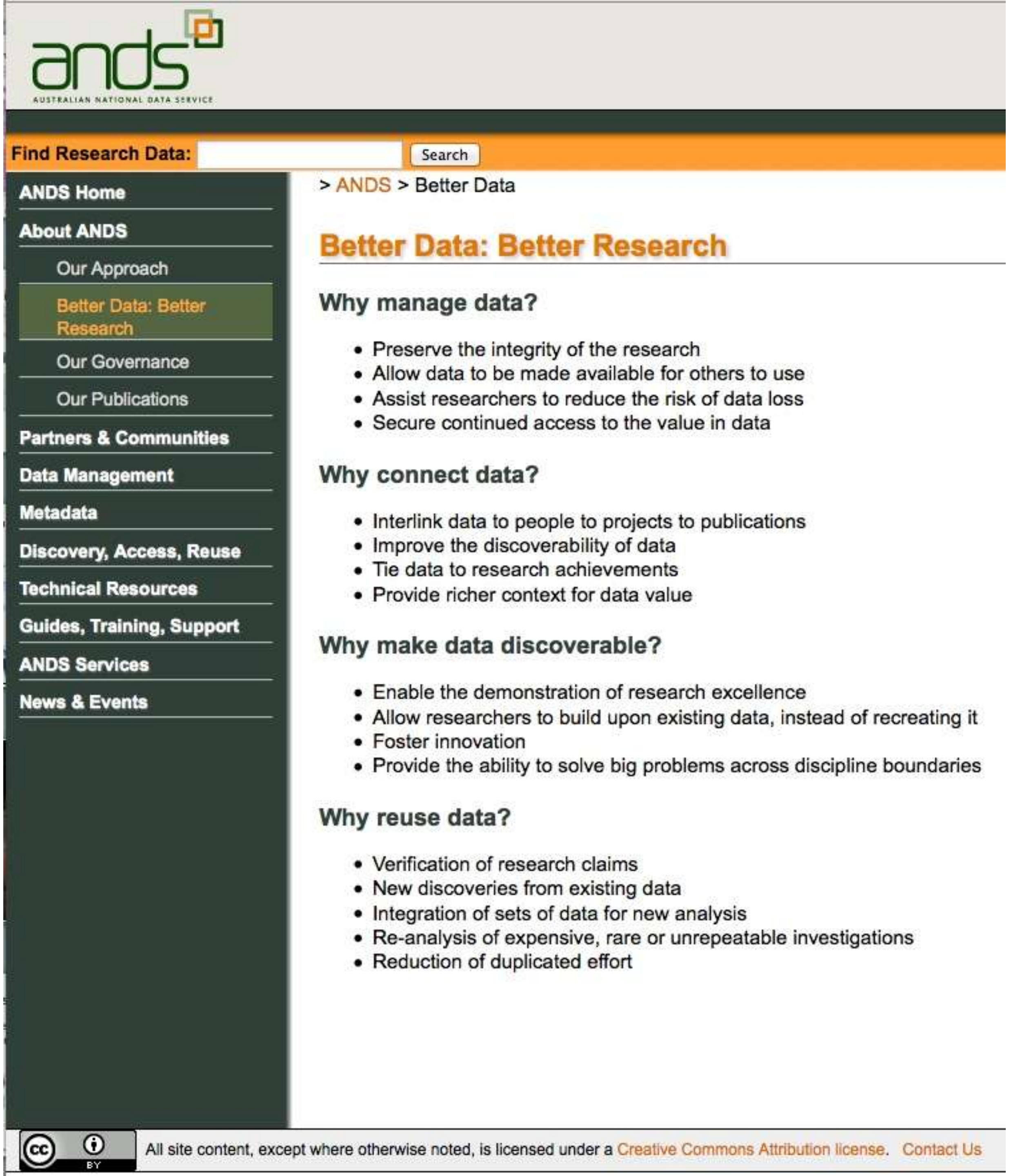
Consolidating value in data

July 19, 1922. State Library and Archives of Florida.
Flickr commons



Page 105 of "The Street railway journal" (1884);
Flickr Commons

Australian National Data Service



> 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?

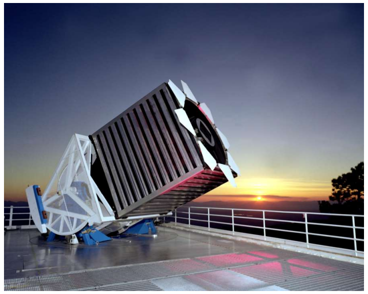
- Enable the demonstration of research excellence
- Allow researchers to build upon existing data, instead of recreating it
- Foster innovation
- Provide the ability to solve big problems across discipline boundaries

Why reuse data?

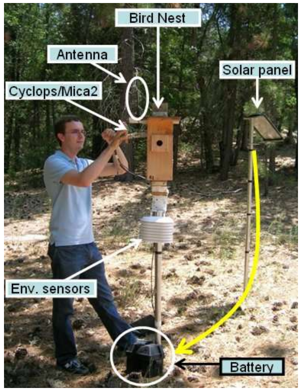
- Verification of research claims
- New discoveries from existing data
- Integration of sets of data for new analysis
- Re-analysis of expensive, rare or unrepeatable investigations
- Reduction of duplicated effort

Ways to pool data

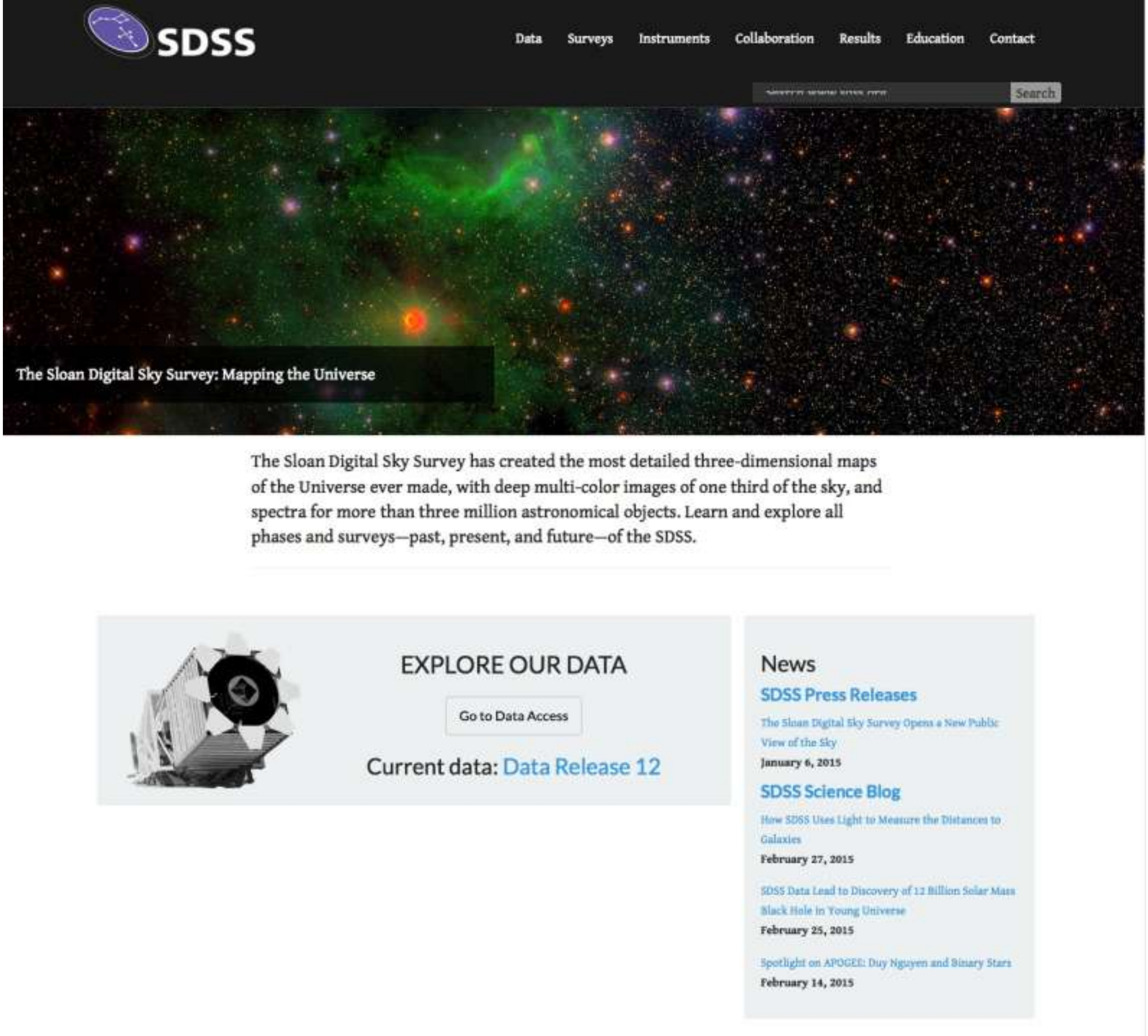
- Centralized data production
  - Top down investments in data
  - Pooled data resources for the community



- Decentralized data production
  - Bottom up investments in data
  - Local data resources pooled later



Sloan Digital Sky Survey



The Sloan Digital Sky Survey: Mapping the Universe

The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects. Learn and explore all phases and surveys—past, present, and future—of the SDSS.

EXPLORE OUR DATA

Go to Data Access

Current data: Data Release 12

News

SDSS Press Releases

The Sloan Digital Sky Survey Opens a New Public View of the Sky
January 6, 2015

SDSS Science Blog

How SDSS Uses Light to Measure the Distances to Galaxies
February 27, 2015

SDSS Data Lead to Discovery of 12 Billion Solar Mass Black Hole in Young Universe
February 25, 2015

Spotlight on APOGEE: Duy Nguyen and Binary Stars
February 14, 2015

Social Science Surveys

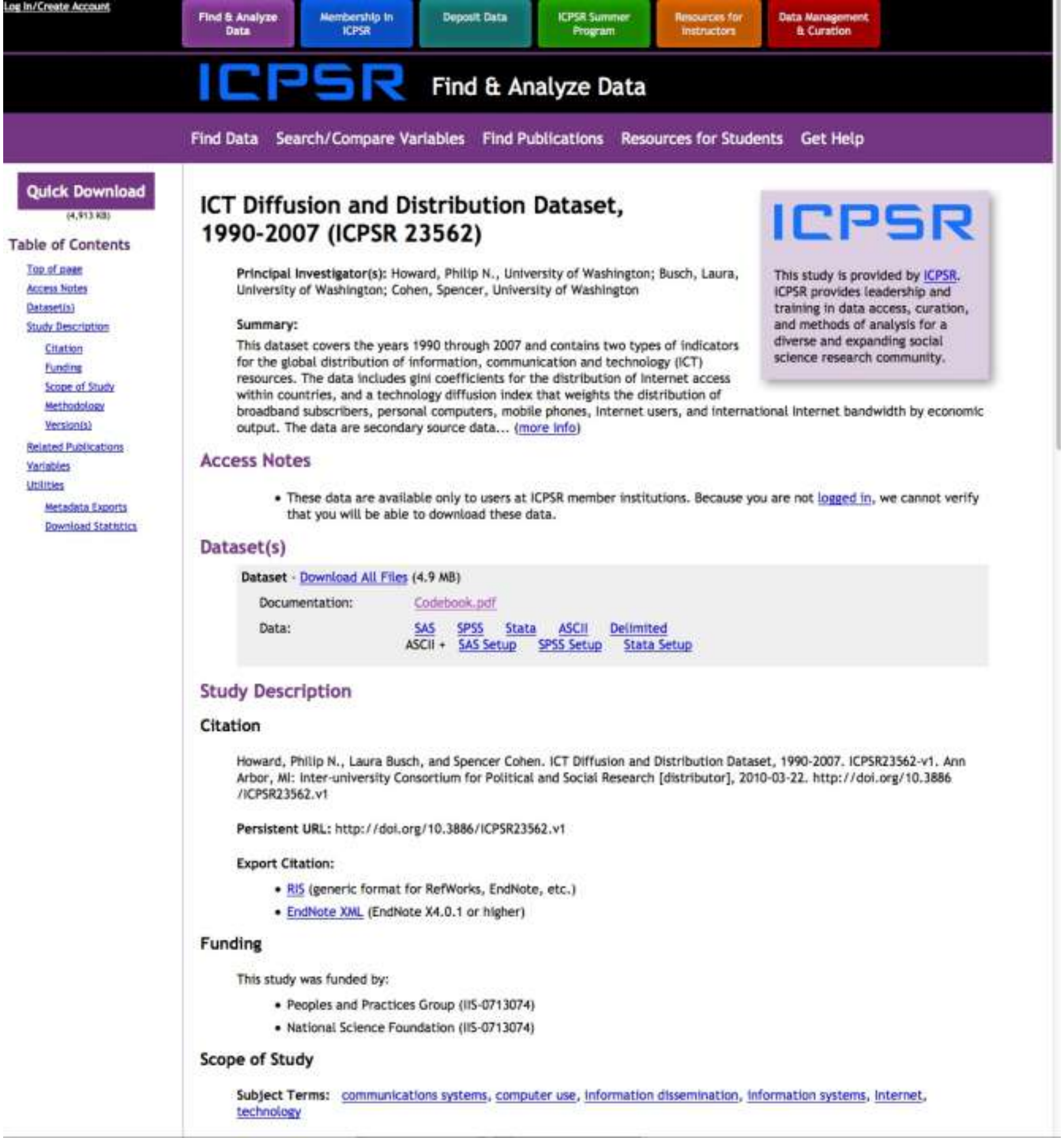


Table of Contents

- Top of page
- Access Notes
- Dataset(s)
- Study Description
  - Citation
  - Funding
  - Scope of Study
  - Methodology
  - Version(s)
- Related Publications
- Variables
- Utilities
  - Metadata Exports
  - Download Statistics

ICT Diffusion and Distribution Dataset, 1990-2007 (ICPSR 23562)

Principal Investigator(s): Howard, Philip N., University of Washington; Busch, Laura, University of Washington; Cohen, Spencer, University of Washington

Summary:

This dataset covers the years 1990 through 2007 and contains two types of indicators for the global distribution of information, communication and technology (ICT) resources. The data includes gini coefficients for the distribution of Internet access within countries, and a technology diffusion index that weights the distribution of broadband subscribers, personal computers, mobile phones, Internet users, and international Internet bandwidth by economic output. The data are secondary source data... (more info)

ICPSR

This study is provided by ICPSR. ICPSR provides leadership and training in data access, curation, and methods of analysis for a diverse and expanding social science research community.

Access Notes

- These data are available only to users at ICPSR member institutions. Because you are not logged in, we cannot verify that you will be able to download these data.

Dataset(s)

Dataset - Download All Files (4.9 MB)

Documentation:	Codebook.pdf
Data:	SAS SPSS Stata ASCII Delimited
ASCII +	SAS Setup SPSS Setup Stata Setup

Study Description

Citation

Howard, Philip N., Laura Busch, and Spencer Cohen. ICT Diffusion and Distribution Dataset, 1990-2007. ICPSR23562-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-03-22. http://doi.org/10.3886/ICPSR23562.v1

Persistent URL: http://doi.org/10.3886/ICPSR23562.v1

Export Citation:

- RIS (generic format for RefWorks, EndNote, etc.)
- EndNote XML (EndNote X4.0.1 or higher)

Funding

This study was funded by:

- Peoples and Practices Group (IIS-0713074)
- National Science Foundation (IIS-0713074)

Scope of Study

Subject Terms: communications systems, computer use, information dissemination, information systems, Internet, technology

Discovery and Interpretation

- Identify the form and content
- Identify related objects
- Interpret
- Evaluate
- Open
- Read
- Compute upon
- Reuse
- Combine
- Describe
- Annotate…

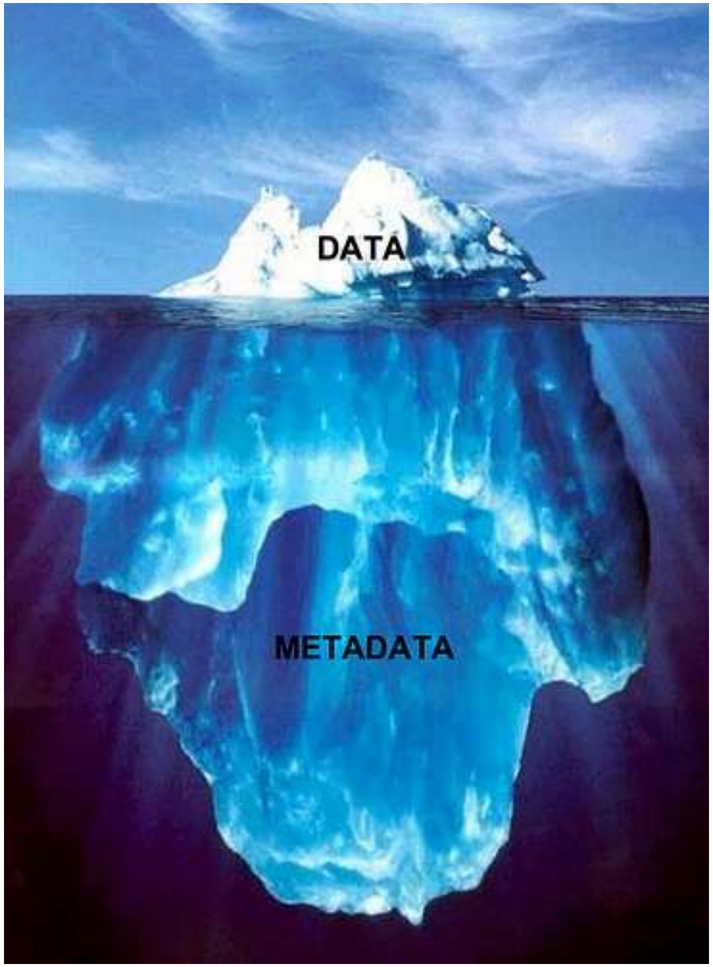
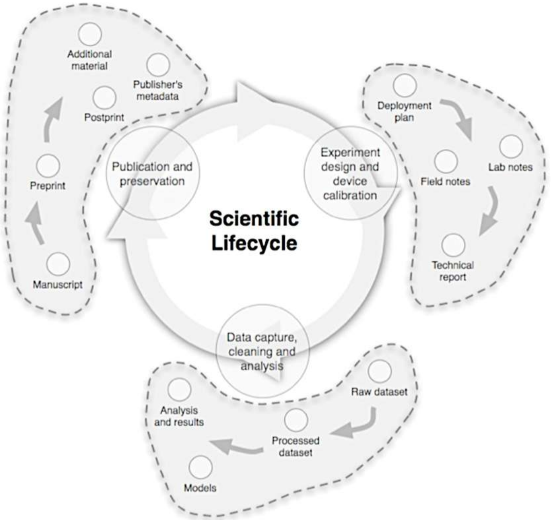


Image from Soumitri Varadarajan blog. Iceberg image © Ralph A. Clevenger. Flickr photo

Describing and attributing data

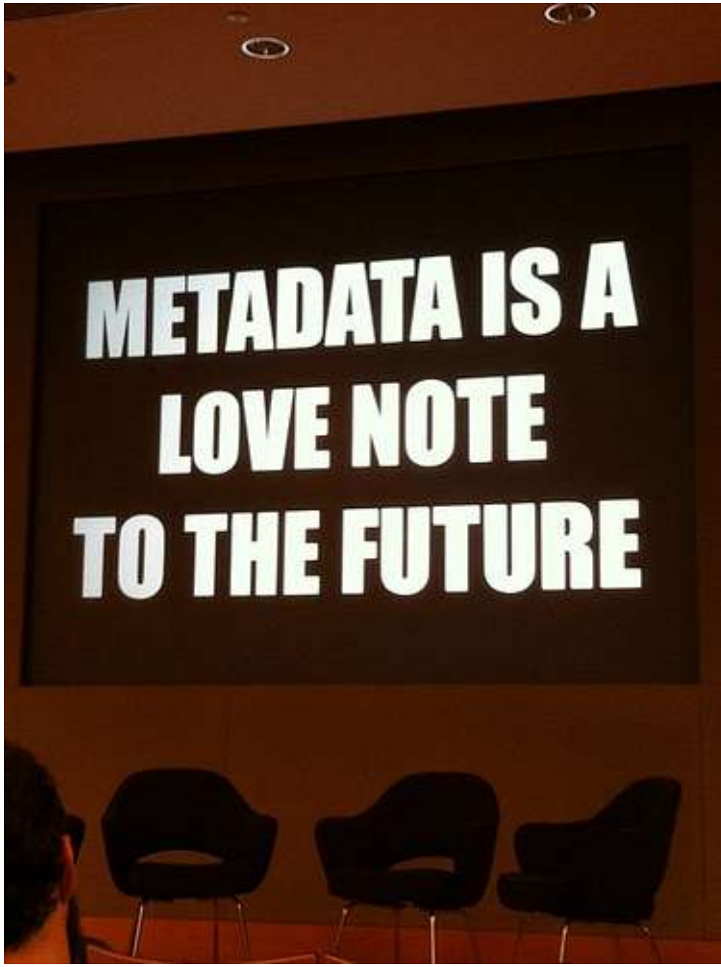
- Compound objects
  - Observations
  - Software
  - Protocols…
- Attribution
  - Investigators
  - Data collectors
  - Analysts…
- Ownership, responsibility



Mary Jane Rathbun (1860-1943), working with crab specimens

Metadata

- Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.
  - descriptive
  - structural
  - administrative



National Information Standards Organization 2004

photo by @kissane

Provenance

- Libraries: Origin or source
- Museums: Chain of custody
- Internet: Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. (World Wide Web Consortium (W3C) Provenance working group)

British Library, provenance record: Bestiary - caption: 'Owl mobbed by smaller birds'

Reuse across place and time

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

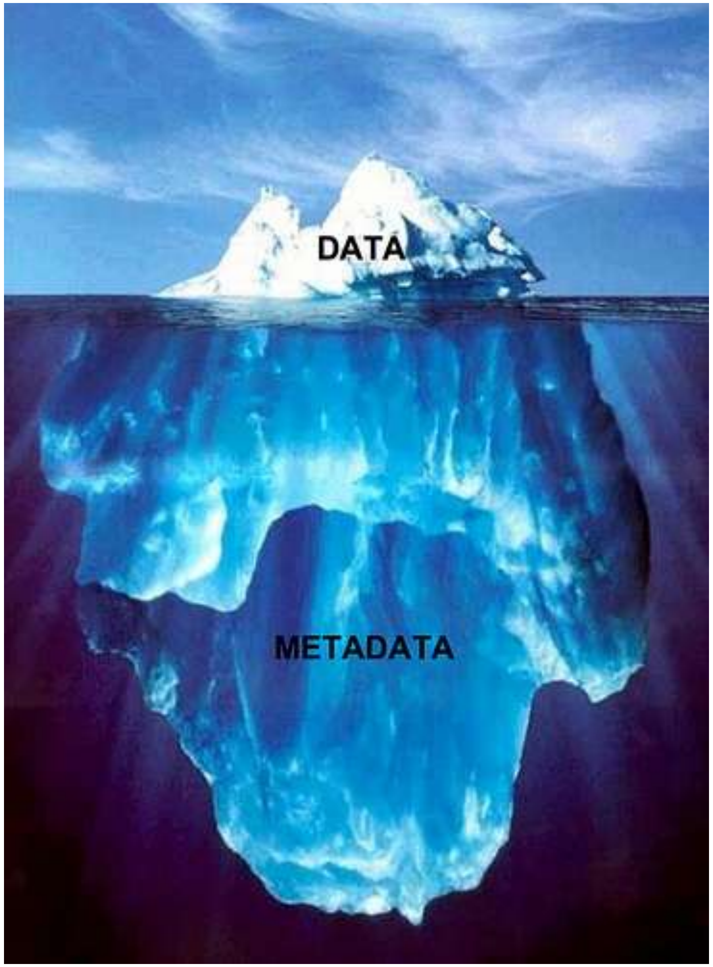
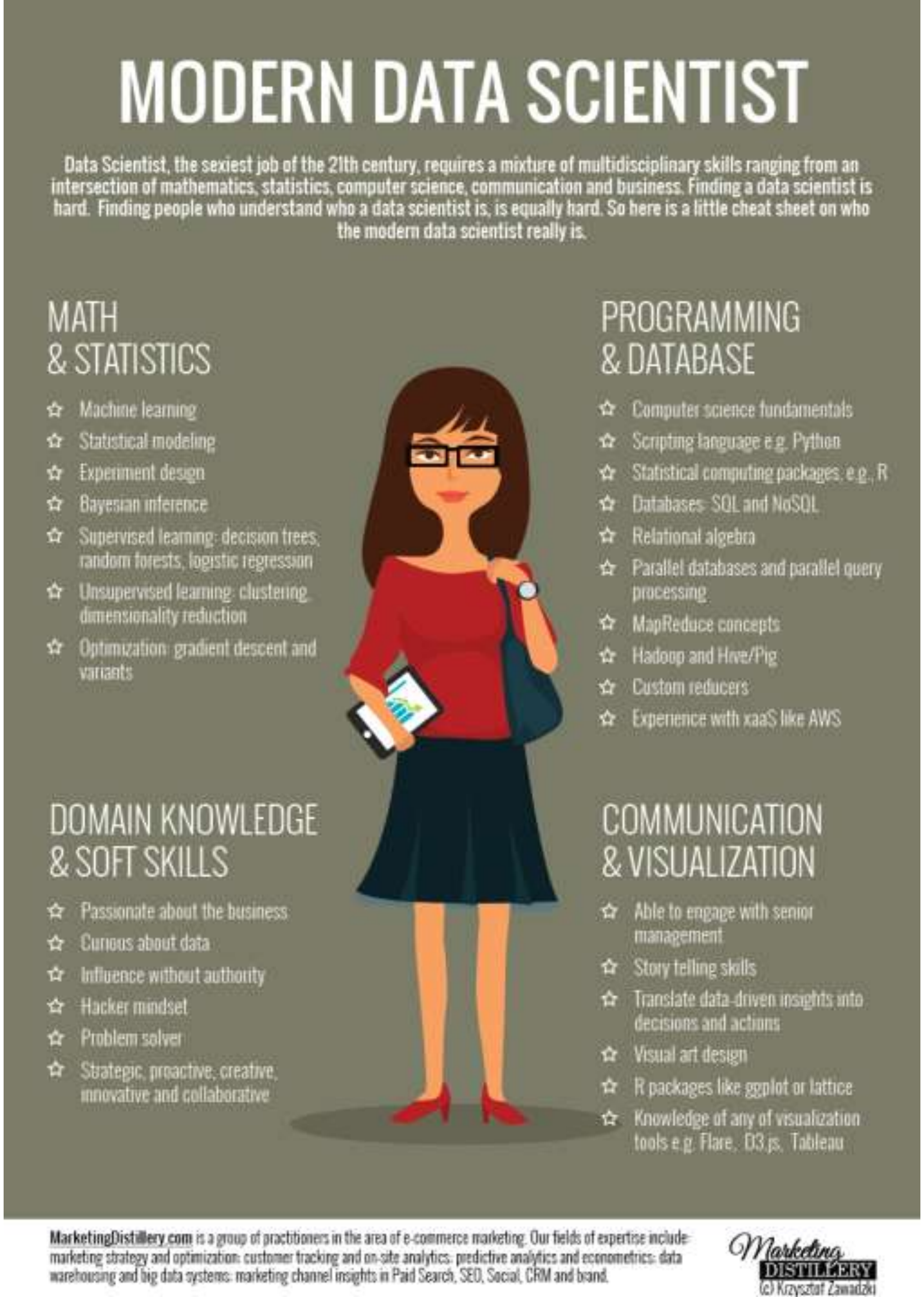


Image from Soumitri Varadarajan blog. Iceberg image © Ralph A. Clevenger. Flickr photo



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing DISTILLERY
(c) Krzysztof Zawadzki

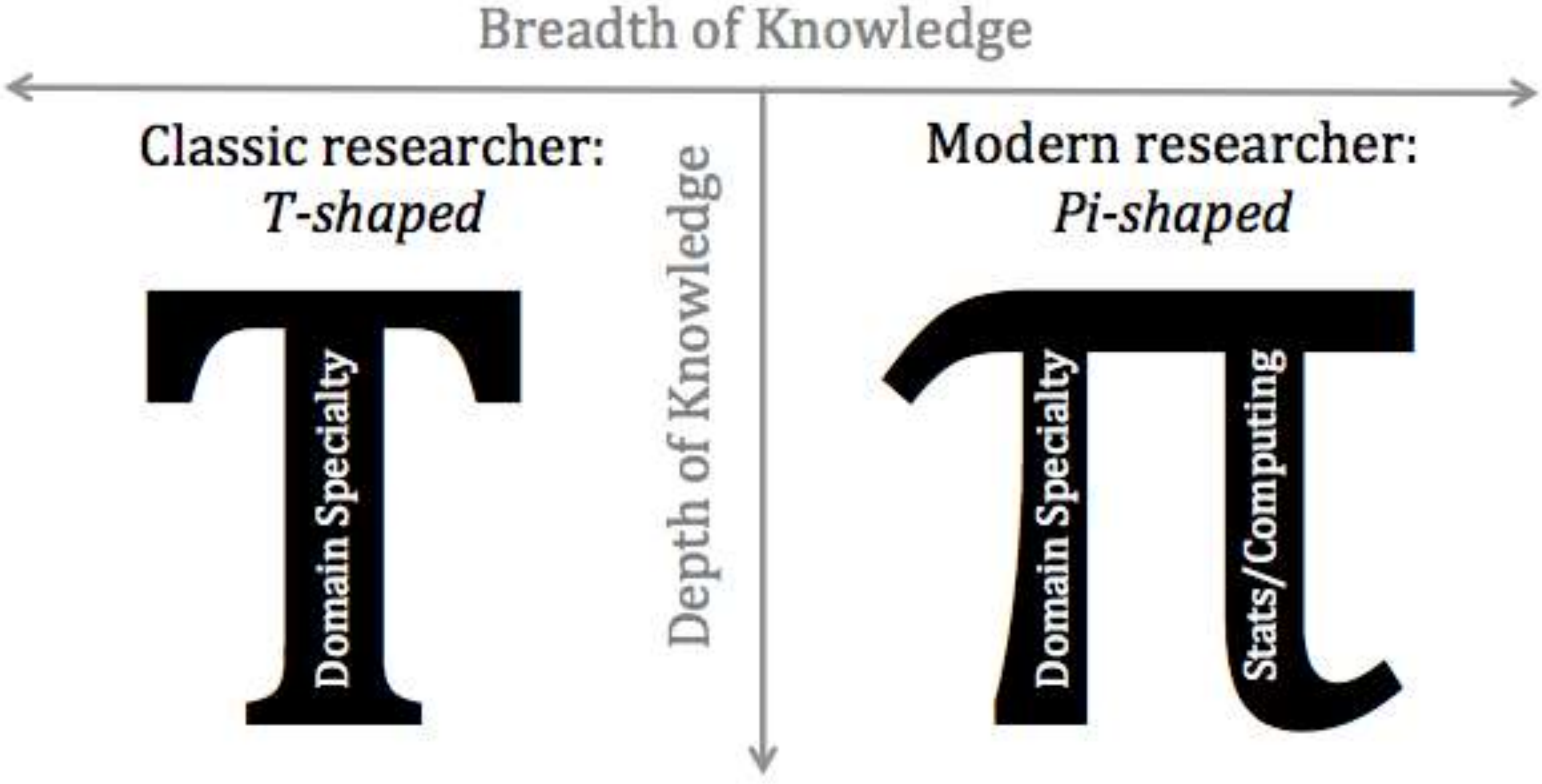
https://github.com/okulbilisim/awesome-datascience

Data Curation and Stewardship

- Services and tools
- Data management planning
- Selection and appraisal
- Metadata, provenance
- Migration
- Economics
- Infrastructure

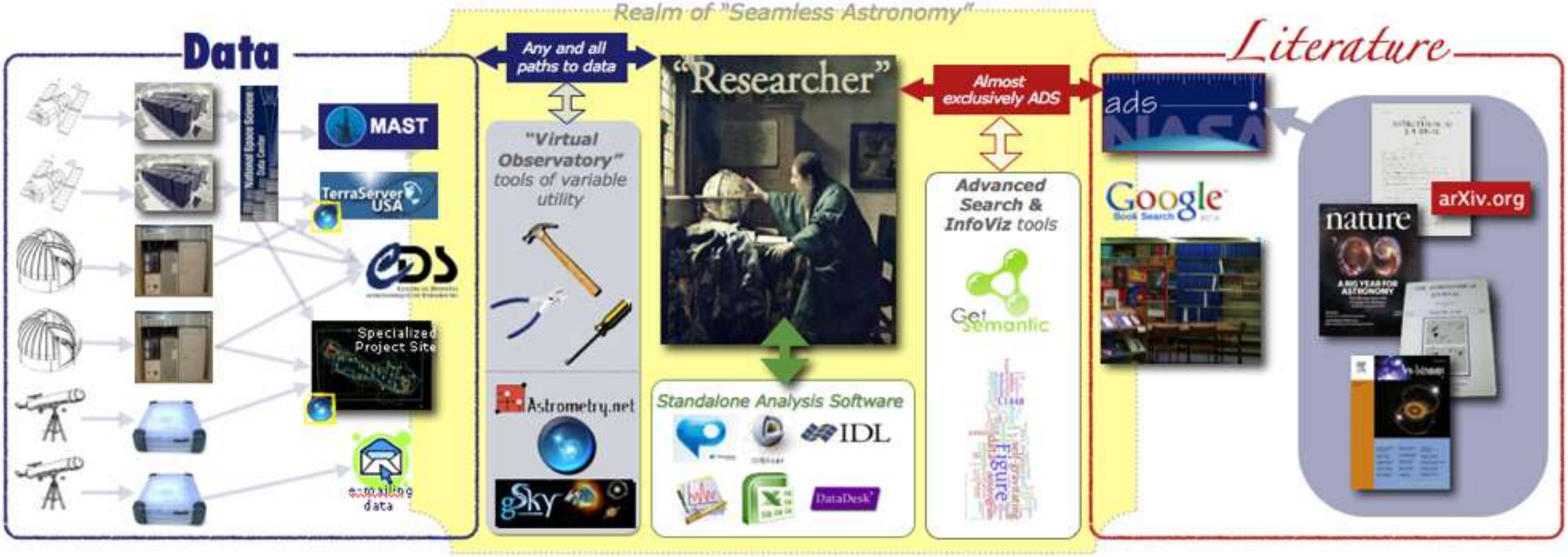
http://www.librarygirl.net/2013/08/putting-your-best-foot-forward-tl.html

Research workforce



https://jakevdp.github.io/blog/2014/08/22/hacking-academia/

Knowledge Infrastructures



Realm of "Seamless Astronomy"
Data
National Space Science Data Center
MAST
TerraServer USA
CDS
Specialized Project Site
emailing data
Any and all paths to data
"Virtual Observatory" tools of variable utility
Astrometry.net
gSky
"Researcher"
Standalone Analysis Software
IDL
DataDesk
Almost exclusively ADS
Advanced Search & InfoViz tools
GetSemantic
Figure
Literature
ads NASA
Google Book Search
nature
arXiv.org

Image: Alyssa Goodman, Harvard Astronomy

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

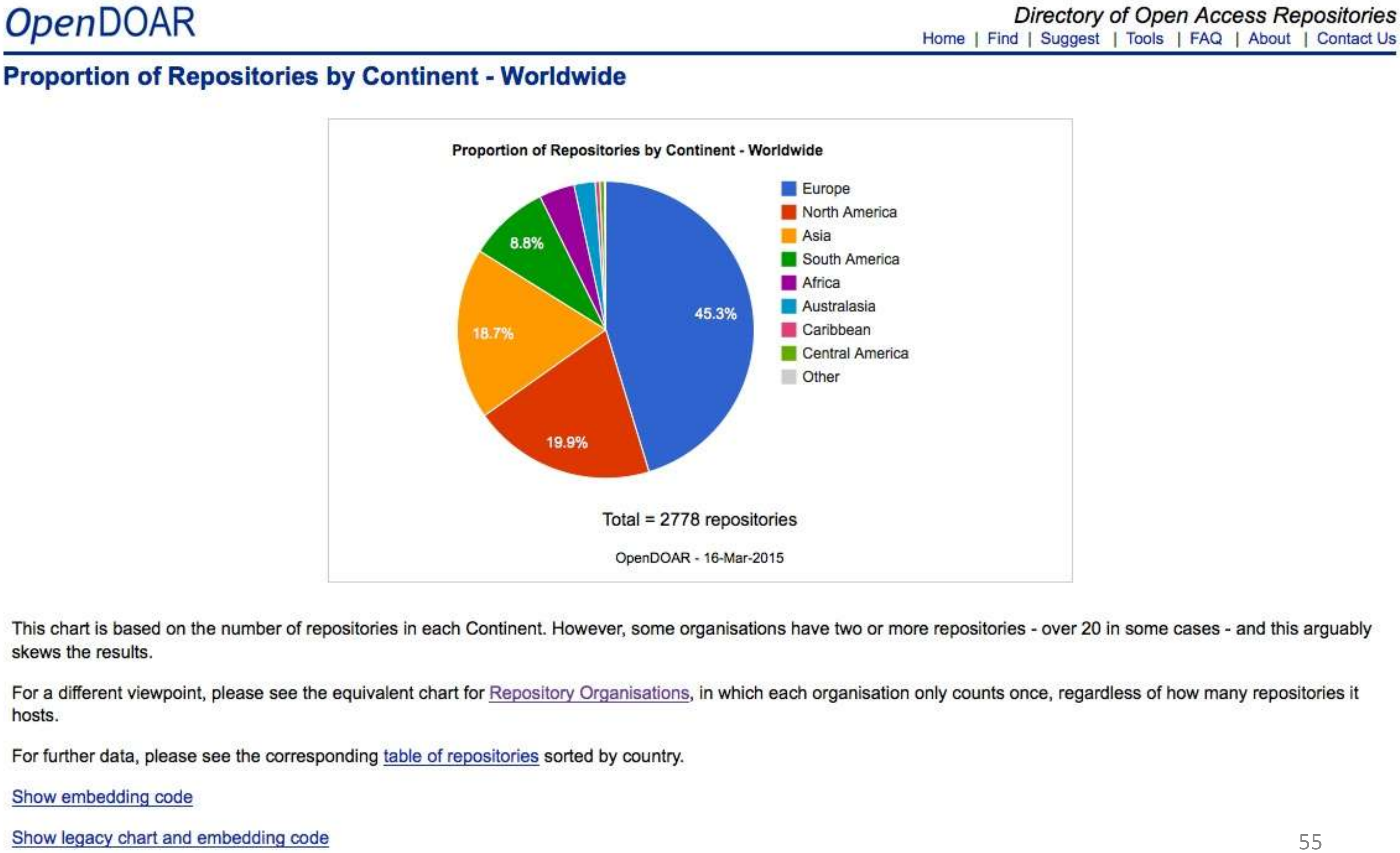
http://knowledgeinfrastructures.org

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*. MIT Press.

Data Repositories



OpenDOAR

Directory of Open Access Repositories

Proportion of Repositories by Continent - Worldwide

Proportion of Repositories by Continent - Worldwide

Total = 2778 repositories

OpenDOAR - 16-Mar-2015

This chart is based on the number of repositories in each Continent. However, some organisations have two or more repositories - over 20 in some cases - and this arguably skews the results.

For a different viewpoint, please see the equivalent chart for Repository Organisations, in which each organisation only counts once, regardless of how many repositories it hosts.

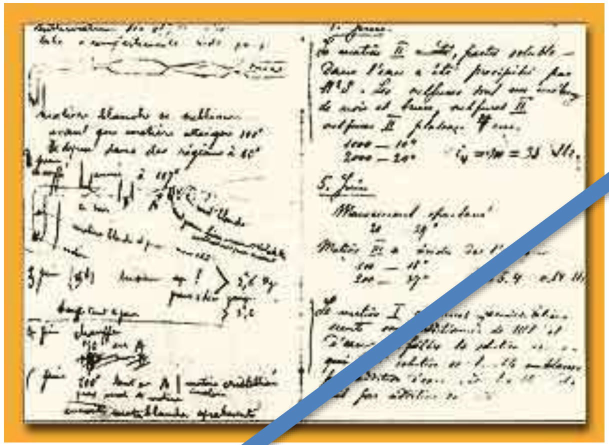
For further data, please see the corresponding table of repositories sorted by country.

Show embedding code

Show legacy chart and embedding code

No Data

- Data not available
- Data not released
- Data not usable



Marie Curie's notebook aip.org

Pisa Griffin

hudsonalpha.org

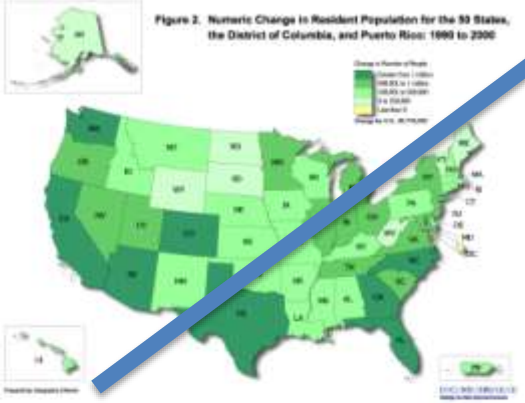
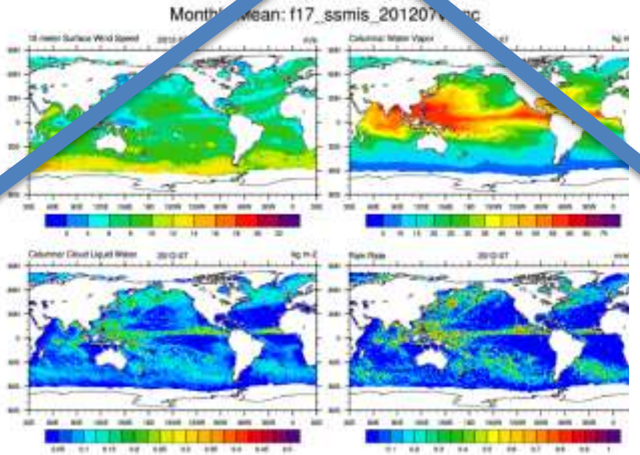


Figure 2. Numeric Change in Resident Population for the 50 States, the District of Columbia, and Puerto Rico: 1990 to 2000

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:July 1980 Place:Sakaltutan

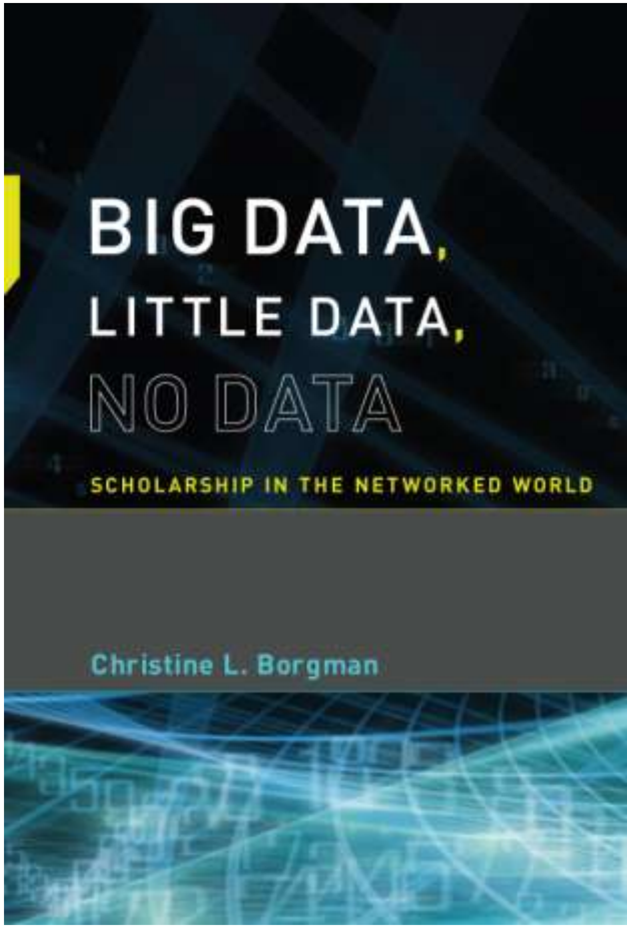
Zafor:

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 - ie.? 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.ph

Conclusions

- Defining data
  - Representations used as evidence
  - One person’s signal is another’s noise
- Creating data
  - Models, questions, methods
  - Domain expertise
  - Data science expertise
- Collaborating with data
  - Documentation, description, identity, linking
  - Incentives for release and reuse
  - Curation and stewardship expertise
- Consolidating data value
  - Infrastructure and workforce investments
  - Value propositions
  - Trust fabric



BIG DATA,
LITTLE DATA,
NO DATA

SCHOLARSHIP IN THE NETWORKED WORLD

Christine L. Borgman

Acknowledgements

UCLA KI Team

- Peter Darch, Milena Golshan, Irene Pasquetto, , Bernie Randles, Ashley Sands

- Former members: Rebekah Cummings, David Fearon, Ariel Hernandez, Elaine Levia, Jaklyn Nunga, Rachel Mandel, Matthew Mayernik, Alberto Pepe, Kalpana Shankar, Katie Shilton, [illegible] Traweek, Jillia[illegible] Wallis, Laura W[illegible] K[illegible] Zhang

- Research funding: Alfred P. Sloan Foundation, National Science Foundation, Microsoft Research, DANS-Netherlands

- University of Oxford: Balliol College, Oliver Smithies [illegible]ord [illegible]