Title
Data Scholarship in the Humanities

Author
Borgman, Christine L.

Publication Date
2014-10-09

Data Scholarship in the Humanities

Christine L. Borgman

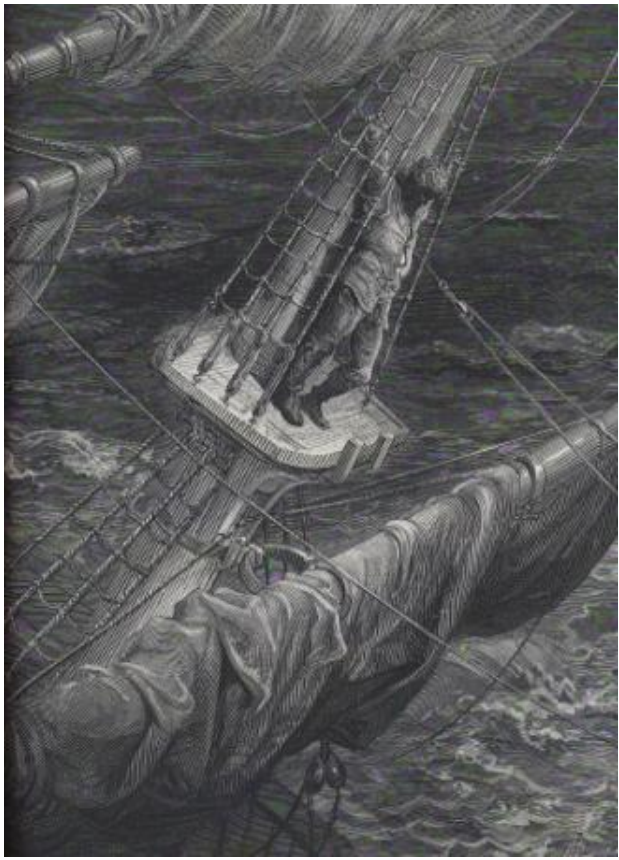
Professor and Presidential Chair in Information Studies

University of California, Los Angeles

Visiting Professor KNAW, DANS – Digital Archiving and Data Services

New Trends in eHumanities, Meertens Institute, Amsterdam (eHumanities Group Research Meeting)

Royal Netherlands Academy of Arts and Sciences 9 October 2014

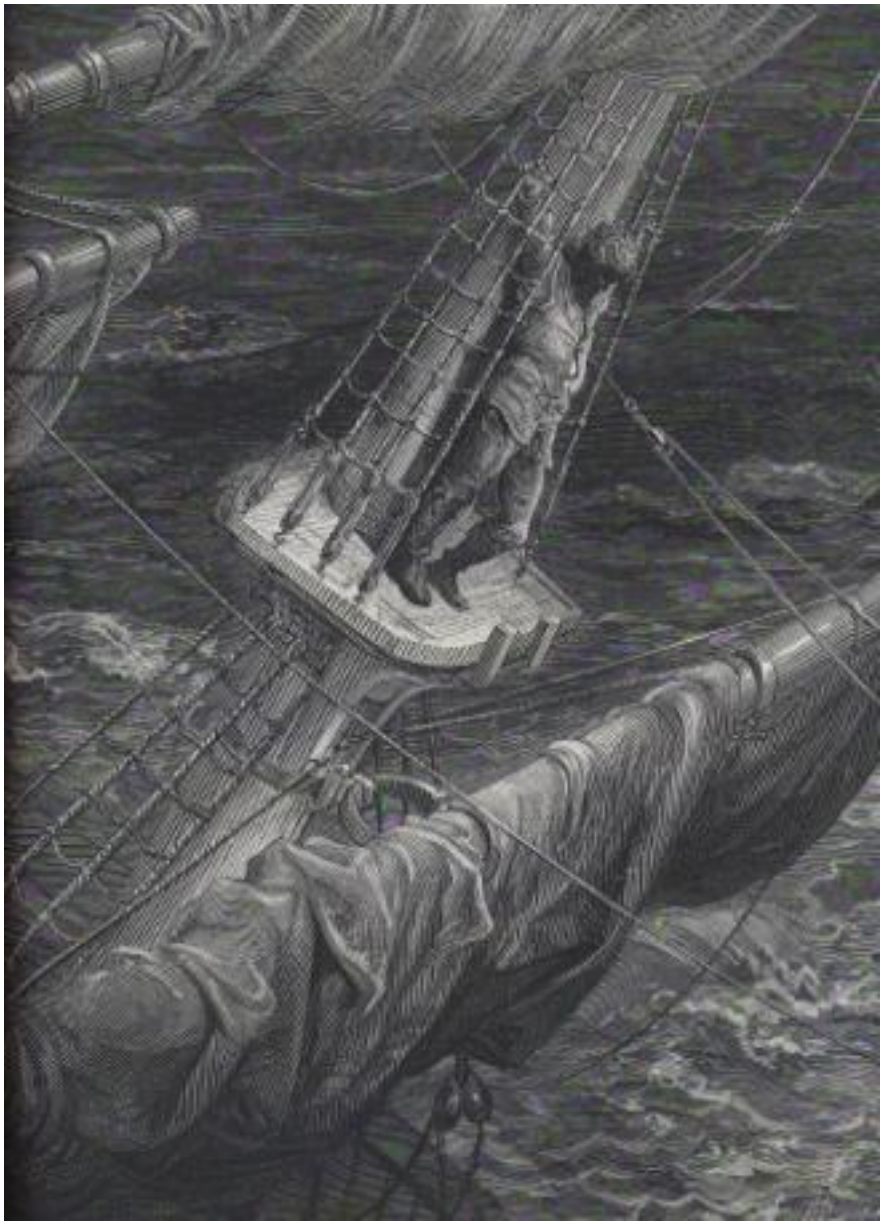


Gustave Dore, *Rime of the Ancient Mariner*, Woodcut, 1798

Day after day, day after day,
We stuck, nor breath nor motion;
As idle as a painted ship
Upon a painted ocean.

Water, water, every where,
And all the boards did shrink;
Water, water, every where,
Nor any drop to drink.

Stanzas from
The Rime of the Ancient Mariner
Samuel Taylor Coleridge, 1798



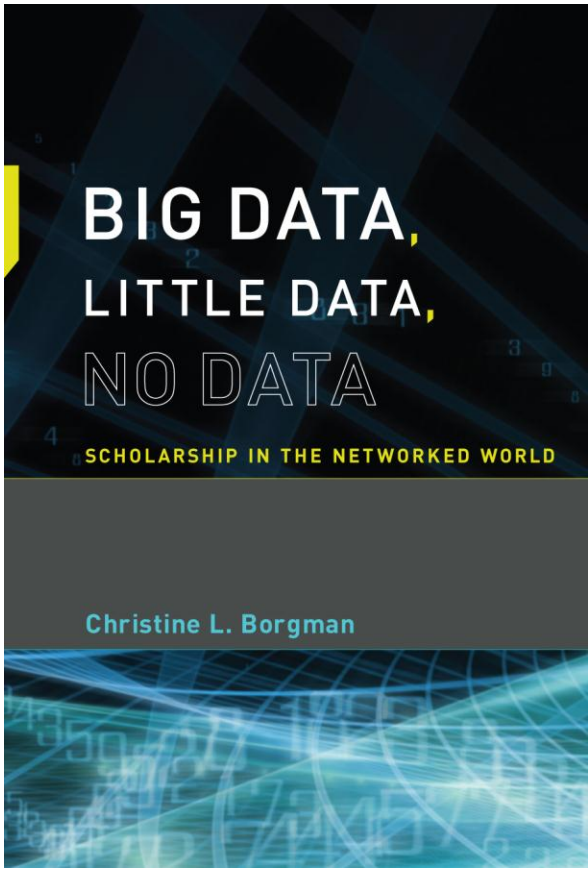
Gustave Dore, Ancient Mariner Illustration, 1798


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

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



*C. L. Borgman (2015, January) MIT Press

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

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
  - 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

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

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Chinese science gets mass transformation

Teamwork at centre of Chinese Academy of Sciences reform.

David Cyranoski

23 September 2014

BEIJING

Pang Xinglei/Xinhua Press/Corbis

Research Data Sharing
without barriers

Precondition:

Researchers share data

Data sharing

- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

Data sharing

- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

Rewards

- Publications
- Publications
- Publications
- Publications
- Publications
- Publications
- Grants
- Awards and honors
- Teaching
- Service
- Data

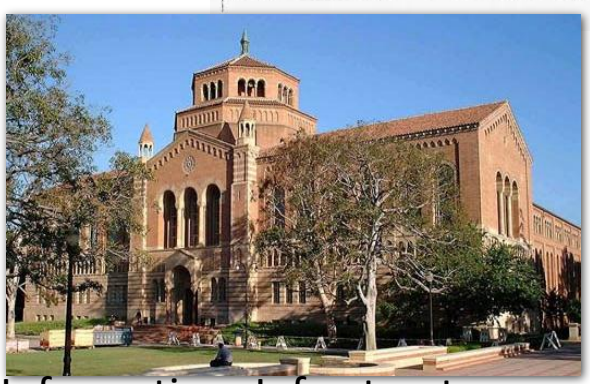
http://blog.startfreshtoday.com/Portals/170402/images/improve-credit-score1.jpg

Functions of Scholarly Publications

- Legitimization
  - Authority, quality
  - Priority, trustworthiness
- Dissemination
  - Awareness
  - Diffusion
  - Publicity
- Access, preservation, curation
  - Availability
  - Discovery
  - Retrieval
  - Persistence

Internet Archive Book Images
Image from page 477 of "The five great monarchies of the ancient eastern world; or, The history, geography, and antiquites of Chaldaea, Assyria, Babylon, Media, and Persia" (1862)

Public Schools Athletic League; Bain News Service [between ca. 1910 and ca. 1915]



Borgman, C.L. (2007). Scholarship in the Digital Age: Information, Infrastructure, and the Internet. MIT Press.

Data sharing

- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

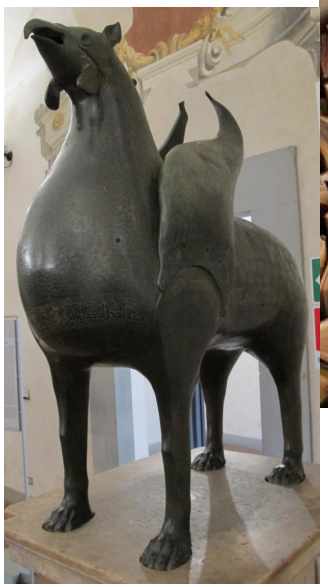
Responsibility

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

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

Responsibility

- Publications
  - Independent units
  - Authorship is negotiated
- Data
  - Compound objects
  - Ownership is rarely clear
  - Attribution
    - Long term responsibility: Investigators
    - Expertise for interpretation: Data collectors and analysts



hudsonalpha.org

Pisa Griffin Image: Arte islamica, ippogrifo, XI sec 03, own work

Attribution of data

- Legal responsibility
  - Licensed data
  - Specific attribution required
- Scholarly credit: contributorship
  - “Author” of data
  - Contributor of data to this publication
  - Colleague who shared data
  - Software developer
  - Data collector
  - Instrument builder
  - Data curator
  - Data manager
  - Data scientist
  - Field site staff
  - Data calibration
  - Data analysis, visualization
  - Funding source
  - Data repository
  - Lab director
  - Principal investigator
  - University research office
  - Research subjects
  - Research workers, e.g., citizen science…

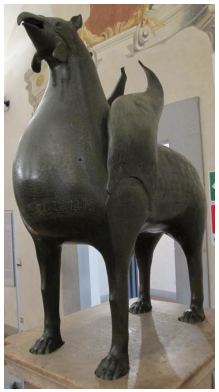
For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop. Washington, D.C.: The National Academies Press. 2012

Data sharing

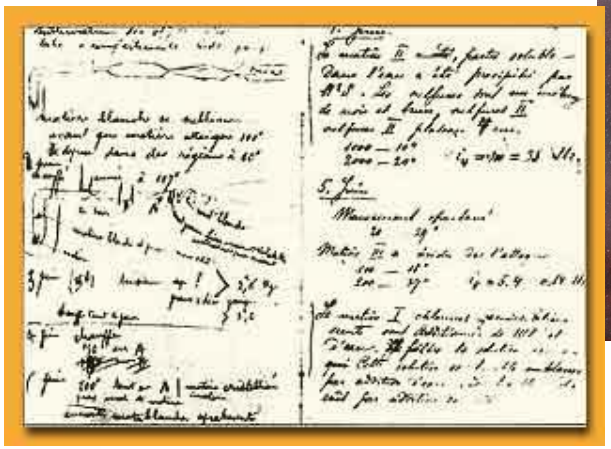
- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

What are data?



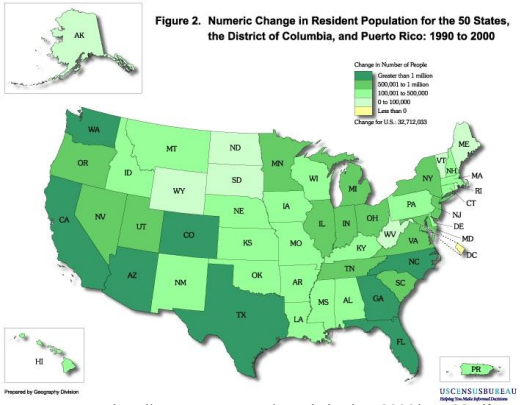
Pisa Griffin
Arte islamica, ippogrifo, XI sec 03, own work



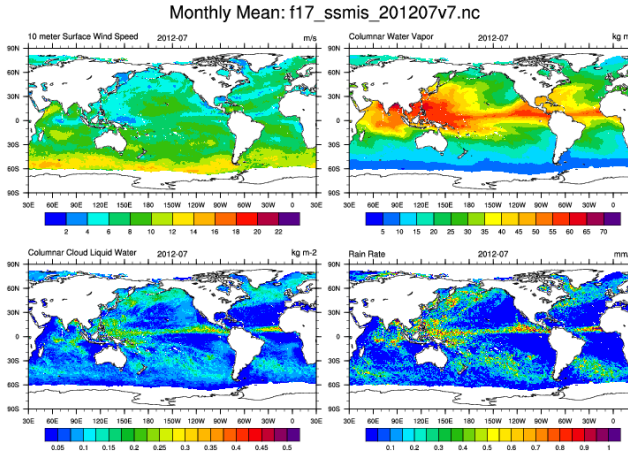
Marie Curie's notebook aip.org

NASA Astronomy Picture of the Day

hudsonalpha.org



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.

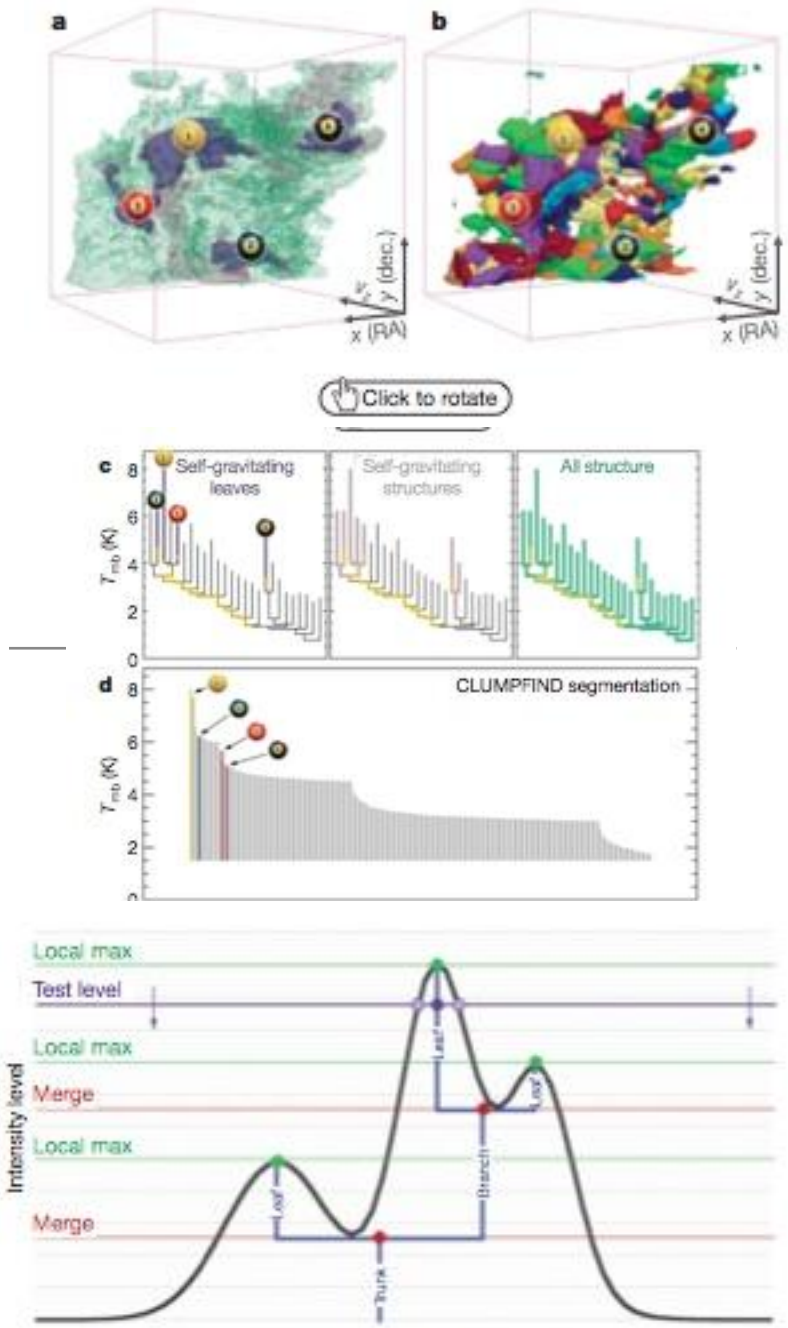
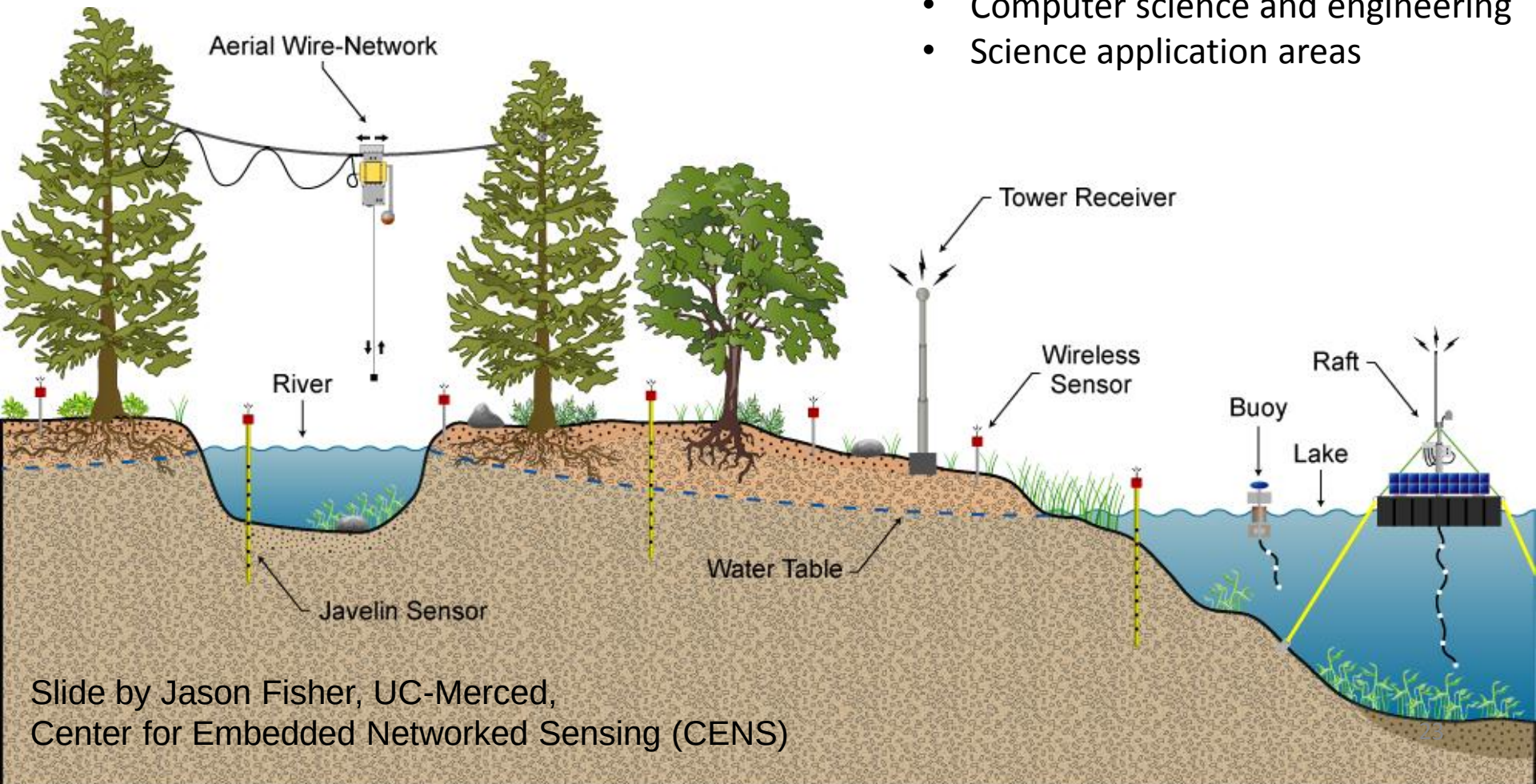


Figure 3 | Schematic illustration of the dendrogram process. Shown is the

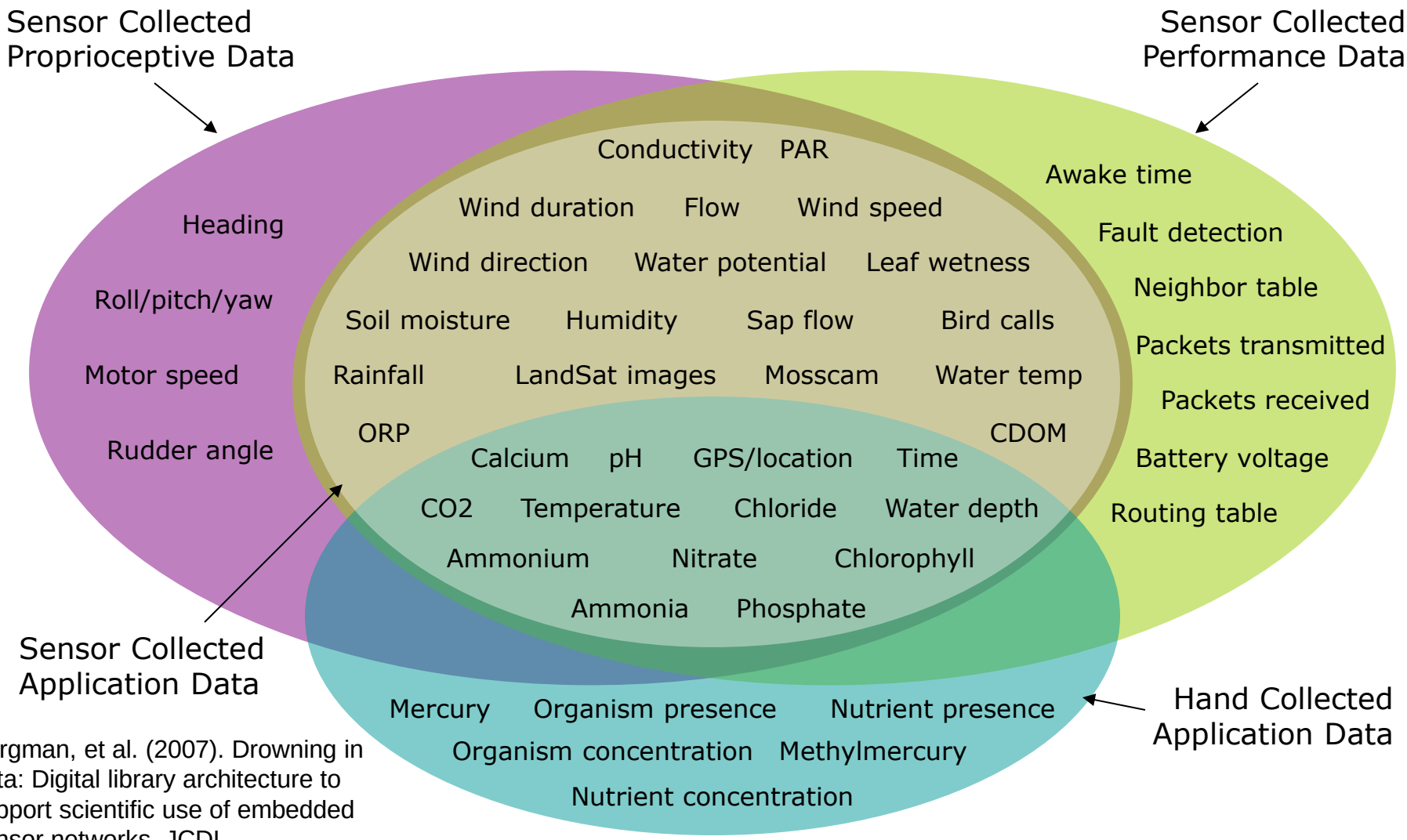
Center for Embedded Networked Sensing

- 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)

CENS data variation



Borgman, et al. (2007). Drowning in data: Digital library architecture to support scientific use of embedded sensor networks. JCDL

Documenting Data for Interpretation

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

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

Repository for seafloor cores. Photo: Peter Darch

Social science data

56. Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

RESPONSE	PUNCH	YEAR									COL: 240
		1972-82	1982B	1983-87	1987B	1988-91	1993	1994	1996	1998	ALL
Strong Democrat	0	2197	143	1271	151	864	227	423	400	370	6,046
Not very strong Democrat	1	3482	109	1655	89	1,282	321	644	577	597	8,756
Independent, close to Democrat	2	1768	44	904	51	578	190	341	356	349	4,581
Independent (Neither, No response)	3	1736	30	855	32	721	205	369	457	477	4,882
Independent, close to Republican	4	1106	8	743	9	571	158	282	258	244	3,379
Not very strong Republican	5	2011	8	1259	15	1,170	299	519	500	484	6,265
Strong Republican	6	1009	8	751	2	662	180	321	307	239	3,479
Other party, refused to say	7	243	0	75	1	44	17	44	43	63	530
Don't know	8	10	0	0	0	0	0	0	0	0	10
No answer	9	64	4	29	3	15	9	49	6	9	188

See Appendix D: Recodes, for original question format and method of recoding. See Appendix N for changes across surveys. If planning to perform trend analysis with this variable, please consult GSS Methodological Report No. 56.

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

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

CLAROS *data*

The world of art on the semantic web

Home Objects People Places SPARQL Search

CLAROS Data

The CLAROS Data service provides a RESTful interface for the data of the CLAROS Project, and complements the CLAROS Explorer.

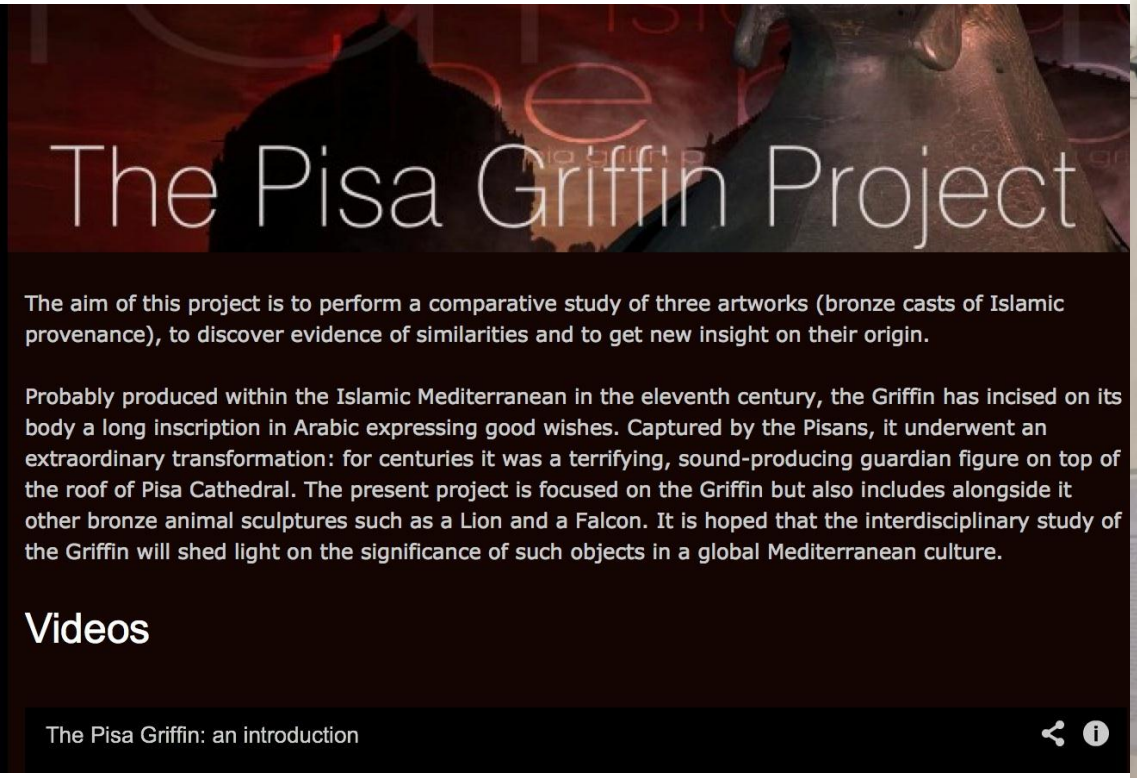
This service provides metadata about archaeology and art in machine-readable formats such as RDF, JSON and KML.

The data for the CLAROS project are modelled using the Erlangen OWL-DL 1.0 implementation of the CIDOC Conceptual Reference Model using the `http://purl.org/NET/crm-owl#` namespace.

You can use the Objects and People views to start exploring, or try your hand at a SPARQL query.

Hosted by the University of Oxford's e-research centre, OeRC
This interface was built using Fuseki, humfrey, and a tweaked frontend, all open-source software.

http://data.clarosnet.org/



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

Arte islamica, ippogrifo, XI sec 03, own work

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

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] mātāpitṛbhrātṛbhaginīsamacittāḥ 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.

Zacchetti, S. (2005). In Praise of the Light: A Critical Synoptic Edition with an Annotated Translation of Chapters 1-3 of Dharmaraksa's Guang zan jing, Being the Earliest Chinese Translation of the Larger Prajnaparamita. Tokyo, Japan: The International Research Institute for Advanced Buddhology, Soka University. Retrieved from http://iriab.soka.ac.jp/orc/Publications/BPPB/index_BPPB.html

Humanities data

CBETA 中華電子佛典協會

登入 | Register

Google™ 自訂搜尋 大藏經搜尋

▼ CBETA 首頁 ▼ 電子佛典集成光碟 ▼ 下載 ▼ 檔案櫃 ▼ 服務 ▼ CBETA相關網站 版權宣告

CBETA 首頁

熱門連結

- jCBReader 跨平台版本 NEW
- ePub 電子書 NEW
- 漢文大藏經
- 漢文大藏經 Android App
- 漢文大藏經（線上手機版）

我們也有 g+ 專頁 追蹤

CBETA 電子佛典集成光碟 2011

下載光碟

索取光碟

光碟說明

光碟更新 NEW

新聞與公告

標題	張貼日期
CBETA 2013 新年公告	2012/12/30 - 13:56
新增「國圖善本佛典」數位內容	2012/09/20 - 18:18
CBETA 開放測試 jCBReader Linux 版本	2012/08/15 - 15:45

新的討論版主題

- CBETA電子佛典部落格 編輯報告 2013/05/20
- 《緒高僧傳》牛頭智巖的標點
- 「磻溪王氏子。壯歲謁冥庵。勤剪髮。聞別傳之旨。」的標點
- CBETA電子佛典部落格 編輯報告 2013/04/20
- 佛說摩利支天經

閱讀全文

最新回應

- RE: 《廣弘明集》卷29錯字
  1 週 4 天 前
- RE: 《緒高僧傳》牛頭智巖的標點
  1 週 4 天 前
- 非常準確，楳，就是梅字。英也花，楳也是花。 不過，未見
  1 週 5 天 前
- 樓主斷句顯為更勝，呵呵。 同理，最後一句也可改為

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

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, January). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

Data sharing

- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

Incentives

- Publications that report the research

Vs.

- Data that are reusable by others

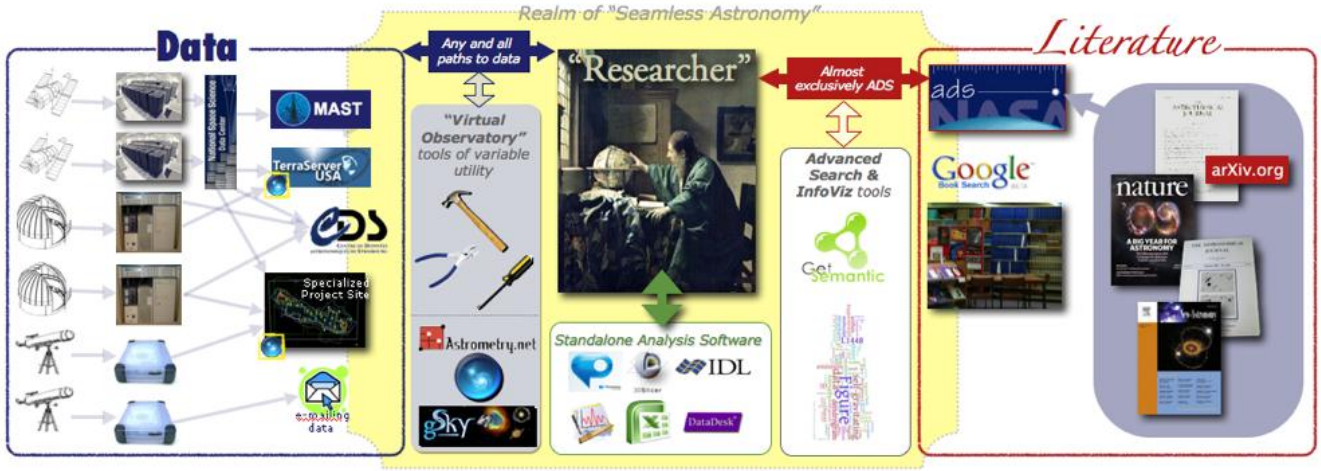
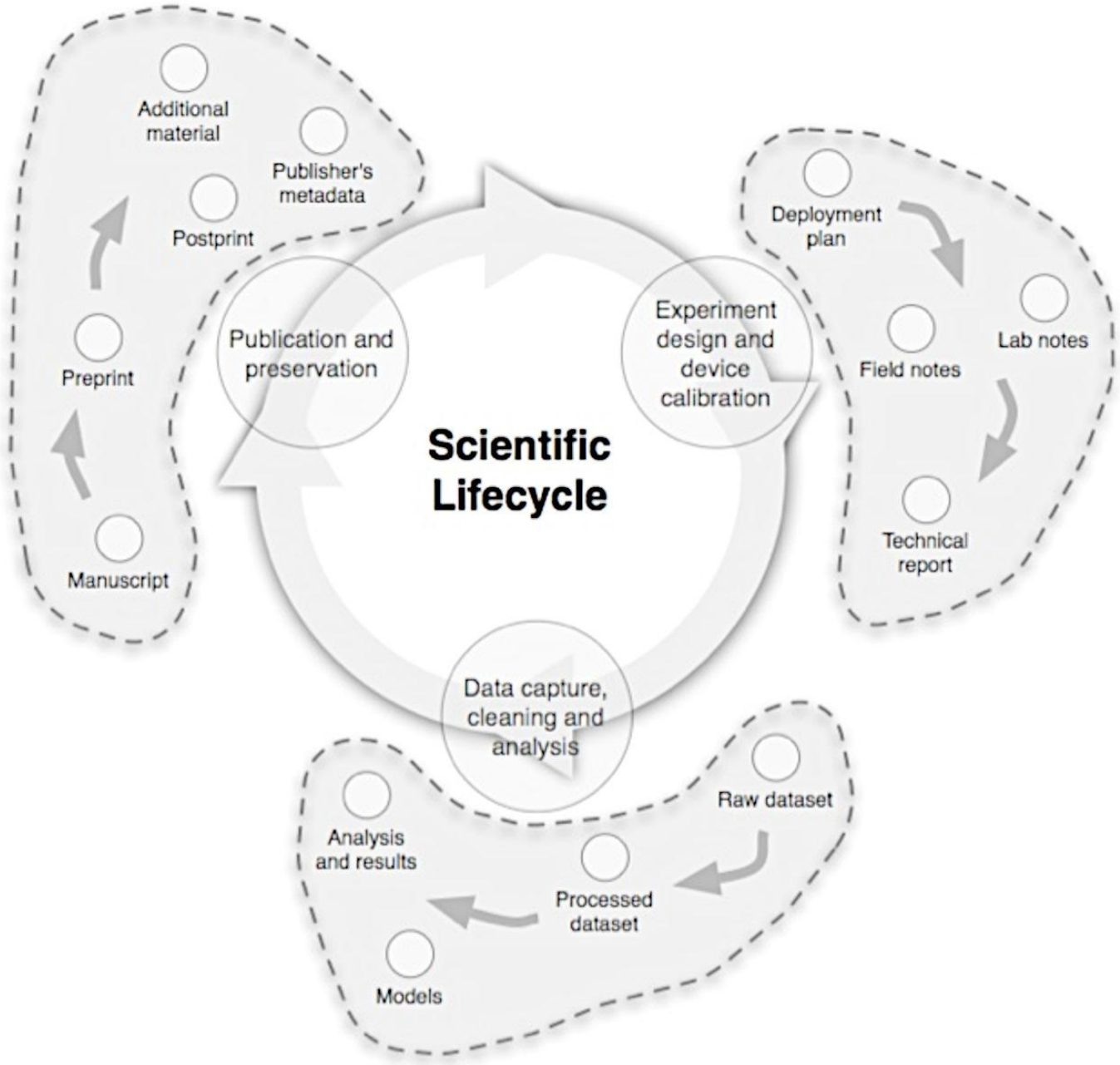


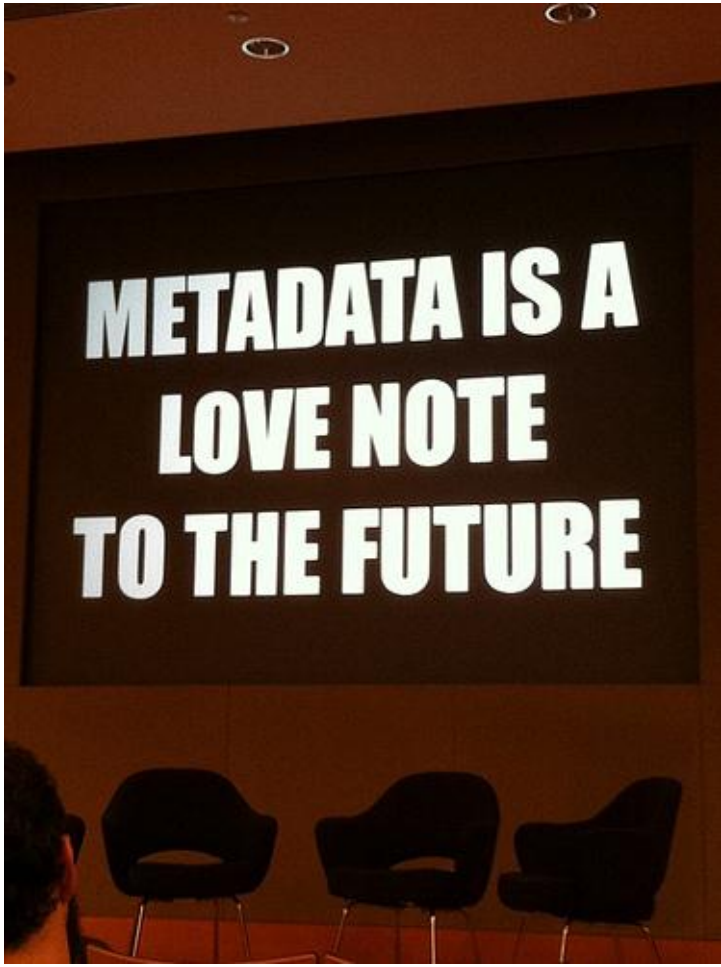
Image: Alyssa Goodman, Harvard Astronomy



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.

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

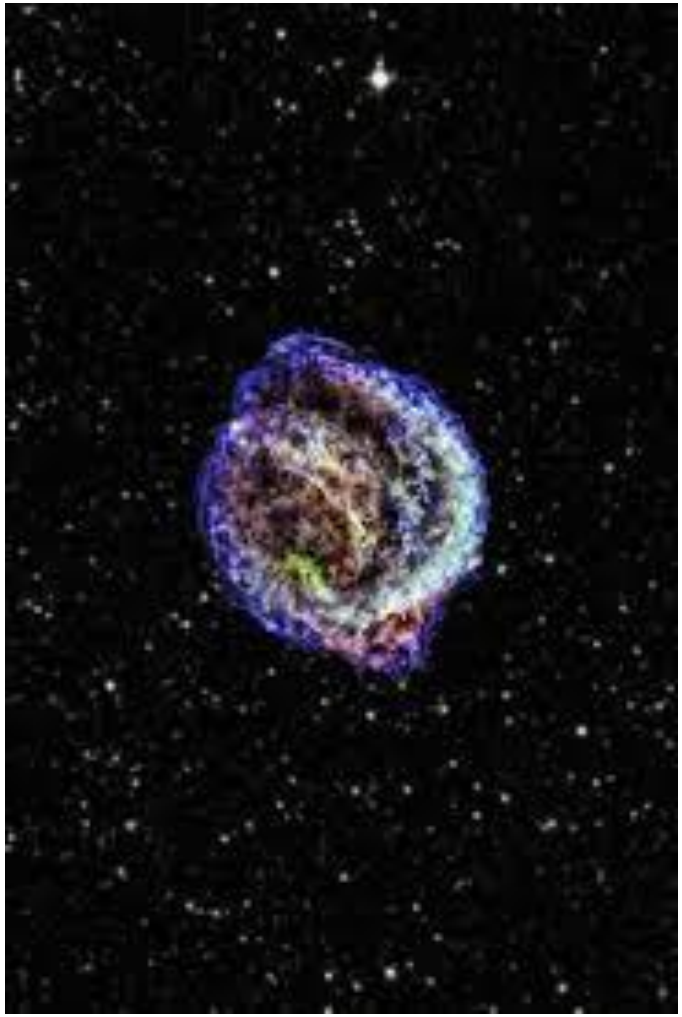
- Libraries: Origin or source
- Museums: Chain of custody
- Internet: Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.*

*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

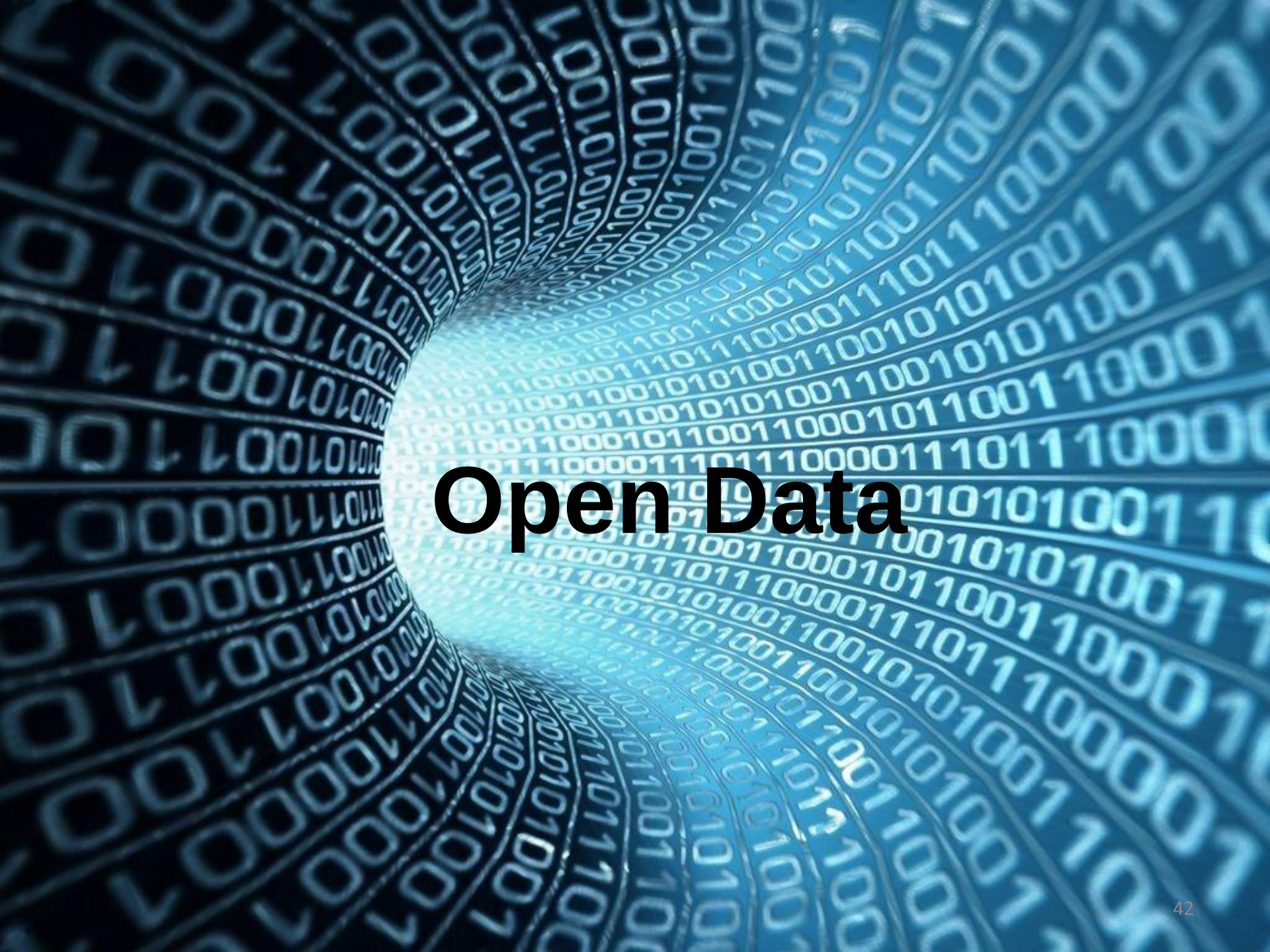


http://chandra.harvard.edu/photo/2013/kepler/kepler_525.jpg

Data sharing

- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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


Open Data

Open Data - 1

- 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). Home page. http://opendatacommons.org/

```
/FontMatrix matrix def
/FontBBox[2048 -1164 1 index div -628 2 index div 4096 3 index div 2062 5 -1 rol
/sfnts [<
747275650009000000000000637674200000000000000009C000007DA6670676D0000000000000878
019901AC01C101C501C901E101F601F601F60222022202280236023F0243024602670285028502940
01160125011800EA00EA00AE0000003E05BB008A04D70053003FFF8CFFD500150028002200990062
B200402F2B59B002602D2C21B0C051580C6423648BB81555621BB200802F2B59B002602D2C0C6423
B8FF80B30809341CB8FF8040E80809343609352F4A0959045809A7090626062B082C0B280C281328
D4401E091202550C0C047F180118401112025518401D1D02551810251E073C2C04002FED3FEDC42B
1700030209131415002159020071B25090B1E050526250018OC0D0D065518021010065518OCB8FFF8
1F1E4A62A1A4FB432E7900020044FFE40405058E001F002D024BB1020243545840262F4010100255
2F2B2BDD2B2BC01112392F2BCDD0CD003FED5DC45D5D2B323FED12392F5D5D5DCD31301B4019125F
11242517012B50100110302A2912110608070A101B011BB80152B38F2D012DB802F6B2012A2BB801
025D5972103C103C111217391112393911123901111239391239393130437940 7A4B573A4524351C
2D602D702D04802D902D02B02D01002D102DC02DD02D042D6037A67F182B10F65D5D71723CFD3C10
FFEAB40C0C02550CB8FFE2B40D0D02550CB8FFD6B4101002550CB8FFDEB50F0F02550C05BD03E200
2F2F2F3FDDCD3F3F10ED10ED313001B00D4B5458BF0046FFE80045FFE8002FFFE80030FFE8B51A18
1010065529B8FFF2B70F0F0655292935341112392F2B2B2BDD2B2B2BC01112392F2B2B2BCD2B2B2B
3236353427260200D07E6B76CF7FCF7A677DCC53356B429F82617E694703AF9E87AF7BFC80A58BAD
961A9C1E9621982A9E2BA816A61AAB1CAD2BB916BE2BCD2BDA2BEC2BFB2B2A202D732573288F1397
201A01601A701A021A120B003FC45D5DED5D5D2F3FC45DED5D5D5D1217393F012F2B2BCD2F2BCDD4
363702902126775C4656201F5F92CBBD75546C2115170D21211C9E62455761FEDE2D2D9B7B364D33
393FDD5DCD31301B40350127600D5D362027602770 27B02704340B371F3A20481F4820051A08134F
FFE0400A1339082013391B201339012B2B2B2B002B012B2B5959591321152322061514171313 3635
C0B215391AB8FFF040131539360815392830143929301439110816390 9B8FFE0401B163929401139
0000000900010000000000000000100000721FE4500571000FB74FADF1000000100000000000000
00340363004403630004036 3B2242F1FBA034E006D0800400E1F7F027F037F047F050430440112BF
014A001F000D0126400B1F0DC61F0D571F0D371F0D410D019E0141000D00420141000D001E014100
021E0024455258B90024021E4459594BB8020153205C58B9010F00224544B1222245445958B90C00
2B2B2B2B2B2B2B2B2B2B2B2B2B2B2B2B0001737500737300456944007373017374 2B2B2B2B2B732B00732B
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
```

Research Area	Status	Round 1	Round 2	Totals
Application Scientists	Faculty	7	6	13
	Staff	5	2	7
	Student	3	2	5
Technology Researchers	Faculty	4	2	6
	Staff	1	2	3
	Student	2	7	9
Totals		22	21	43

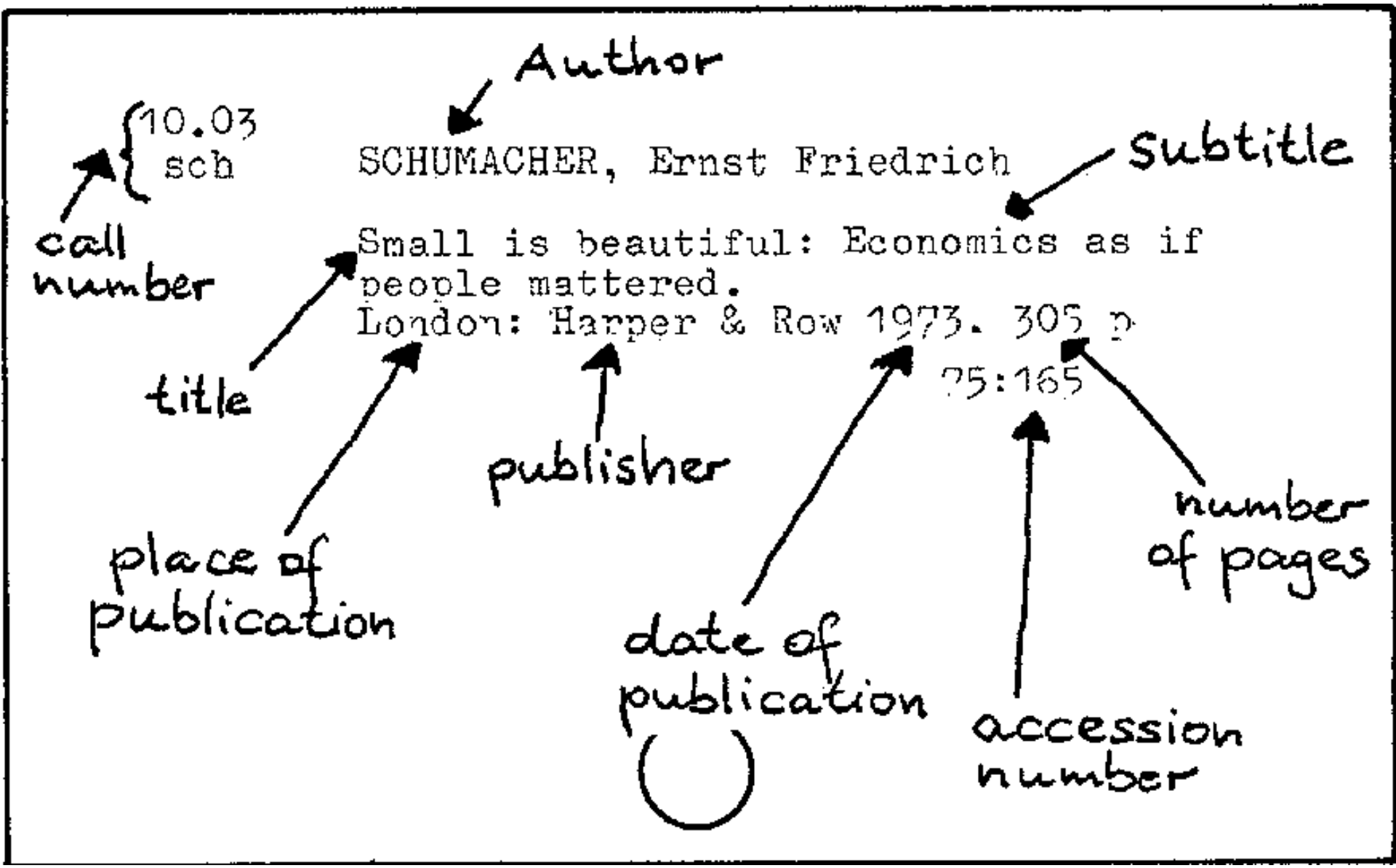
doi:10.1371/journal.pone.0067332.t001

Open Data - 2

- Data that meets the criteria of intelligent openness. Data must be accessible, useable, assessable and intelligible.

Boulton, G., et al. (2012). Science as an open enterprise. *The Royal Society*. http://royalsociety.org/policy/projects/science-public-enterprise/report/

Indian Girl Pitching up Haystack
Uploaded by The U.S. National Archives on December 9, 2009



http://www.nzdl.org/gsdl/collect/fnl2.2/archives/HASH01a1/51ac3343.dir/p027.gif

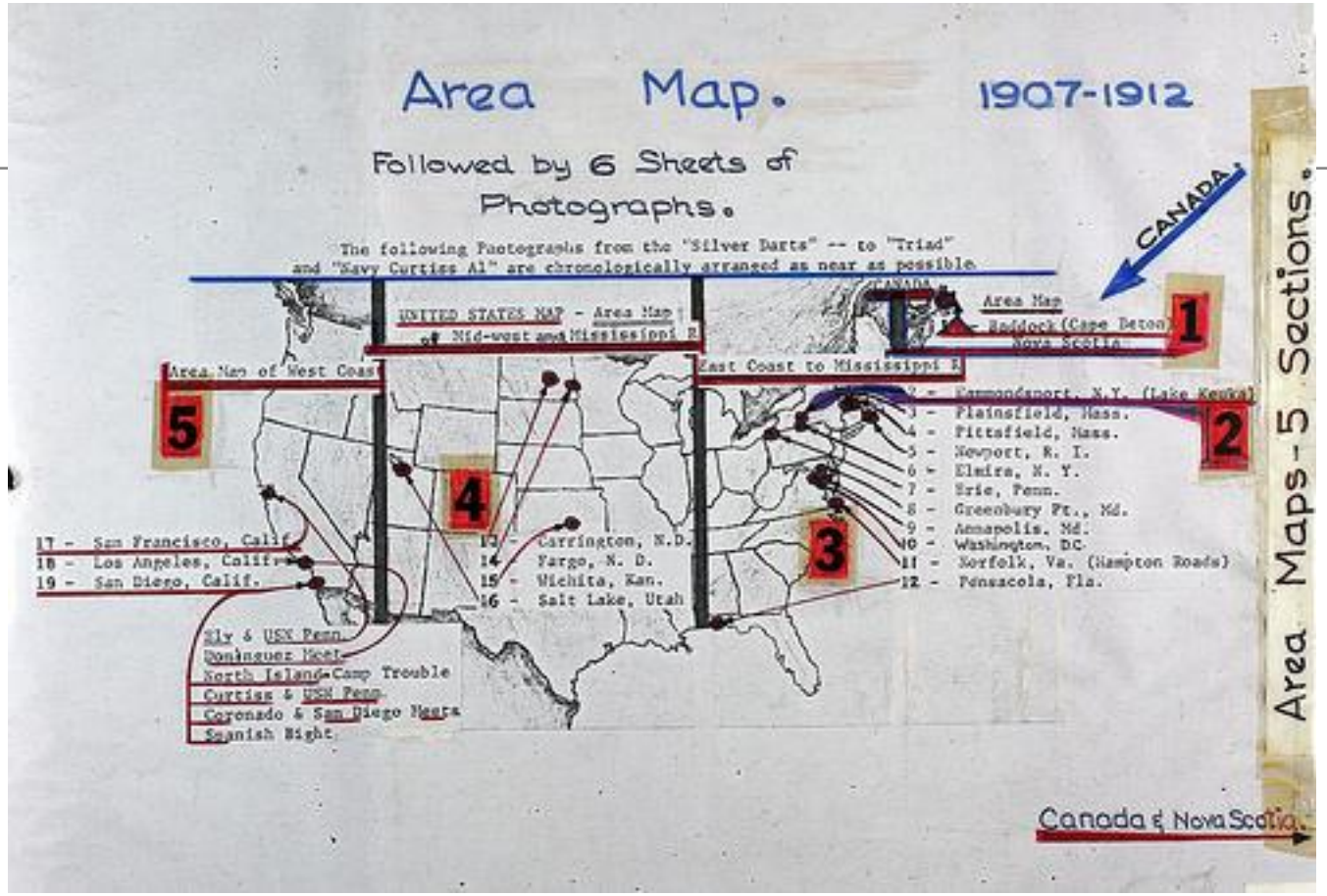
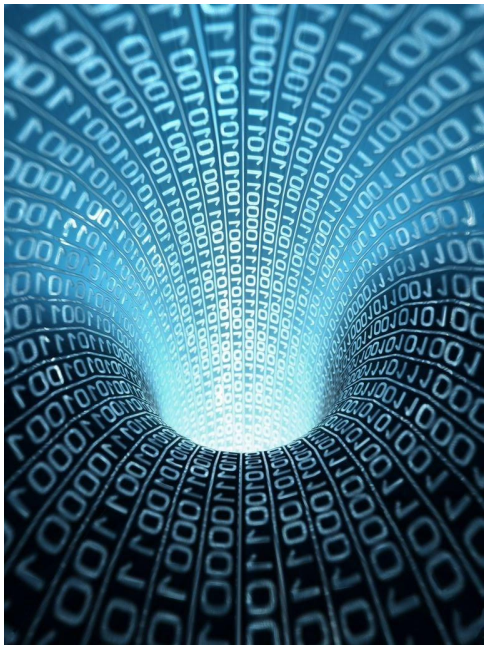


Image from a photo album (AL-27) which belonged to Ray Fife, a pioneer aviation mechanic who built a complete plane at age 16. In 1963 he built a reproduction Curtiss Pusher Biplane. This album documents the building of the pusher and also contains photos used to research the project.

Open Data - 3

- 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

Discovery and Interpretation

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

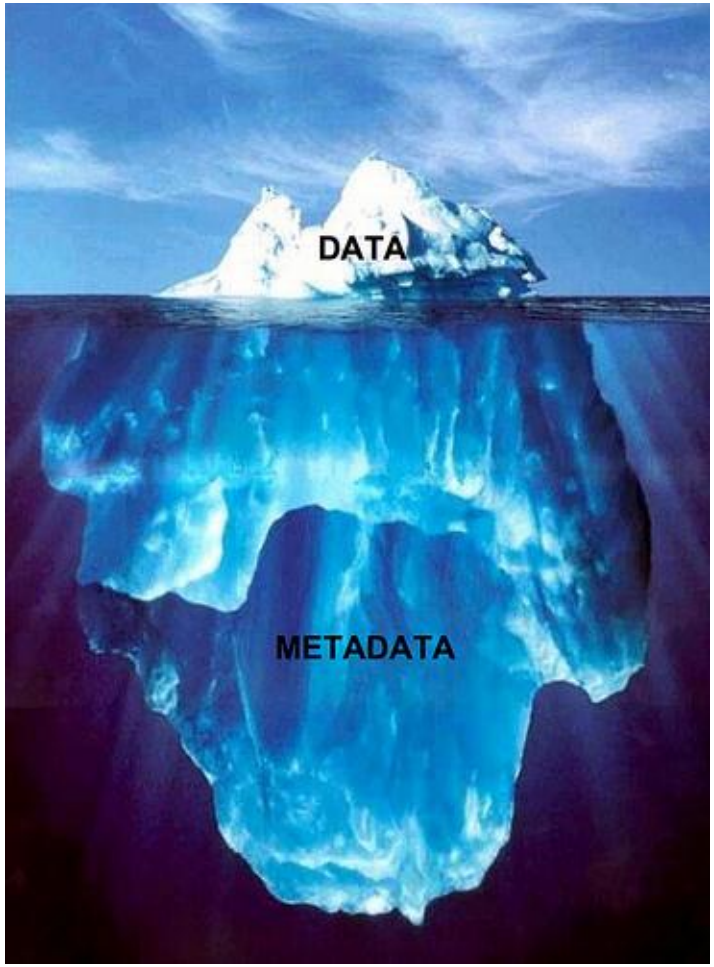


Image from Soumitri Varadarajan blog. Iceberg image © Ralph A. Clevenger. Flickr photo

Data sharing

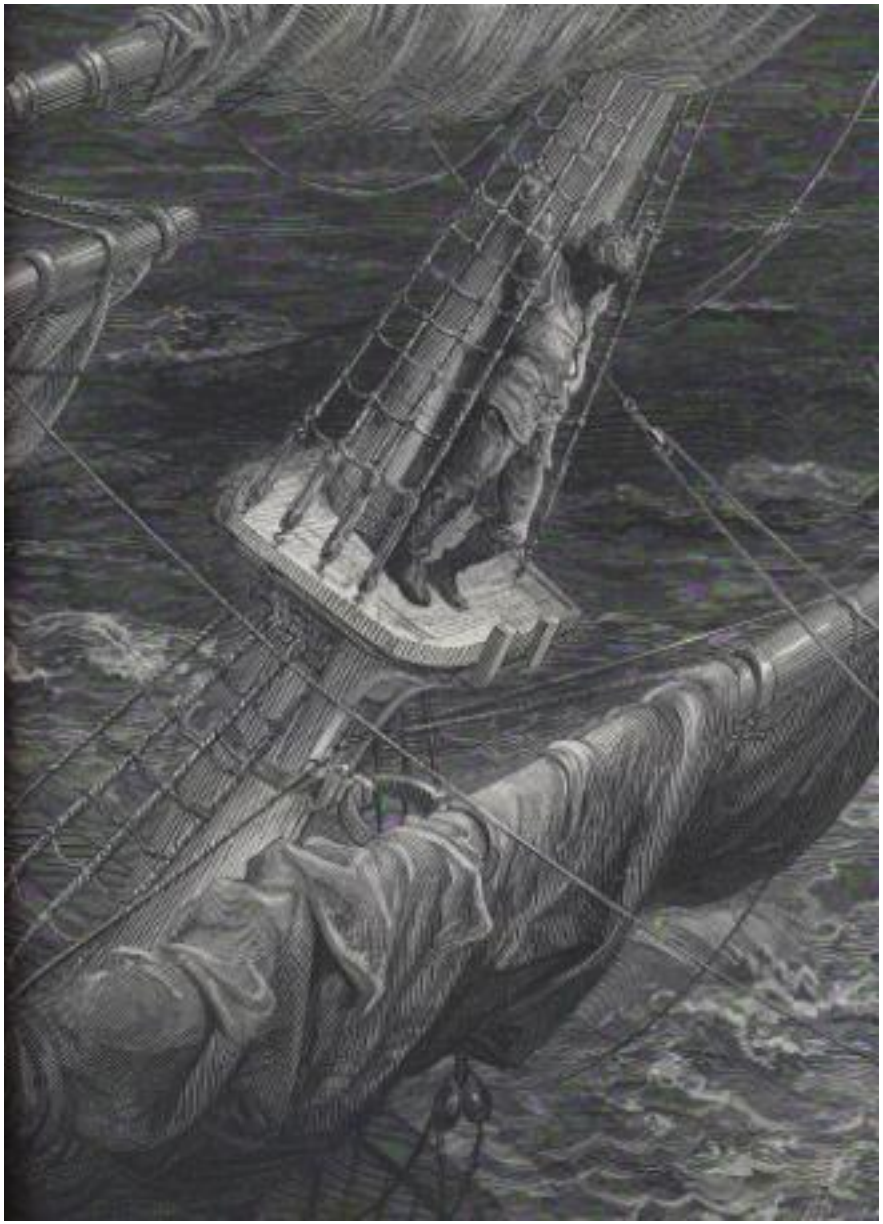
- Scholars' perspectives
  - Rewards
  - Responsibility
  - Data
  - Incentives
- Open data
- Knowledge infrastructure

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

Day after day, day after day,
We stuck, nor breath nor motion;
As idle as a painted ship
Upon a painted ocean.

Water, water, every where,
And all the boards did shrink;
Water, water, every where,
Nor any drop to drink.

Stanzas from
The Rime of the Ancient Mariner
Samuel Taylor Coleridge, 1798



Gustave Dore, Ancient Mariner Illustration, 1798

Emerging themes in data practices

- Scarcity or abundance of data
- Centrality of data to research
- Time frame of research
- Heterogeneity of expertise
- Maturity of standards
- Community building

Borgman, C. L., et al. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. IEEE/ACM *Digital Libraries Conference,* London

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.

To share data, scholars need

- Fresh water
  - Tools
  - Services
  - Skills
  - Resources
  - Incentives

http://environment.nationalgeographic.com/environment/habitats/water-pressure/

To share data, scholars need

- Life boats
  - Repositories
  - Governance models
  - Provenance models
  - Data stewardship workforce

Patent Model, Life Boat, 1841; Smithsonian American History Museum

Knowledge Infrastructures

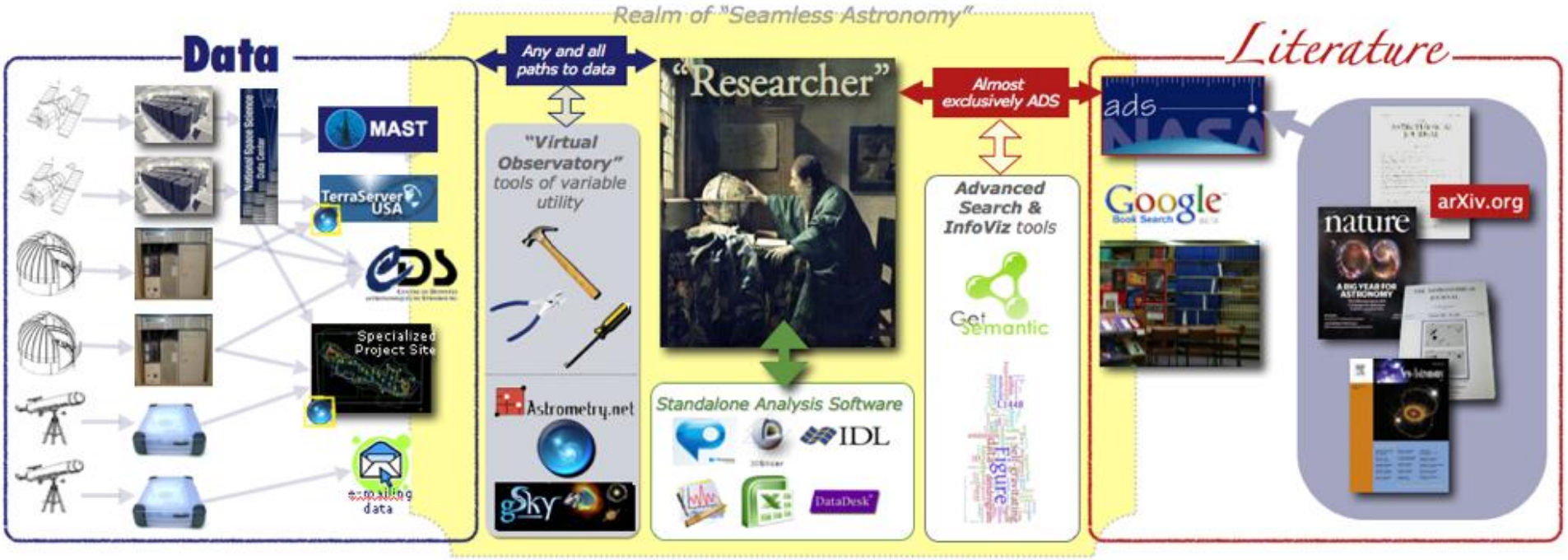
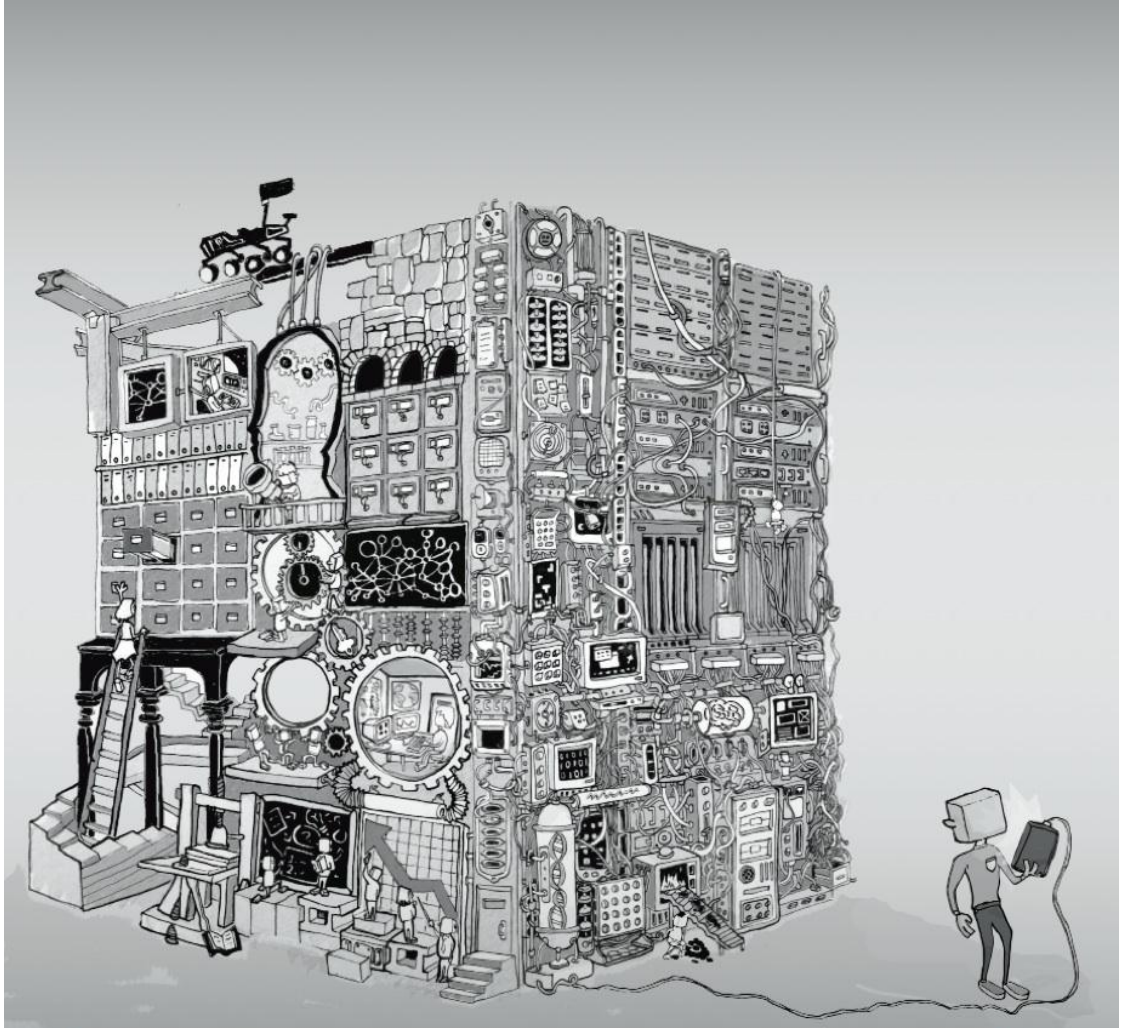


Image: Alyssa Goodman, Harvard Astronomy



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

Acknowledgements

UCLA Data Practices team

- Peter Darch, Milena Golshan, Irene Pasquetto, Ashley Sands, Sharon Traweek
- Former members: Rebekah Cummings, David Fearon, Ariel Hernandez, Elaine Levia, Jaklyn Nunga, Matthew Mayernik, Alberto Pepe, Kalpana Shankar, Katie Shilton, Jillian Wallis, Laura Wynholds, Kan Zhang

- Research funding: National Science Foundation, Alfred P. Sloan Foundation, Microsoft Research, DANS-Netherlands
- University of Oxford: Balliol College, Oliver Smithies Fellowship, Oxford Internet Institute, Oxford eResearch Center, Bodleian Library