Big Science, Little Science, and Open Science:
Sustainability, Stewardship, and Knowledge Infrastructures

Christine L. Borgman

Distinguished Research Professor

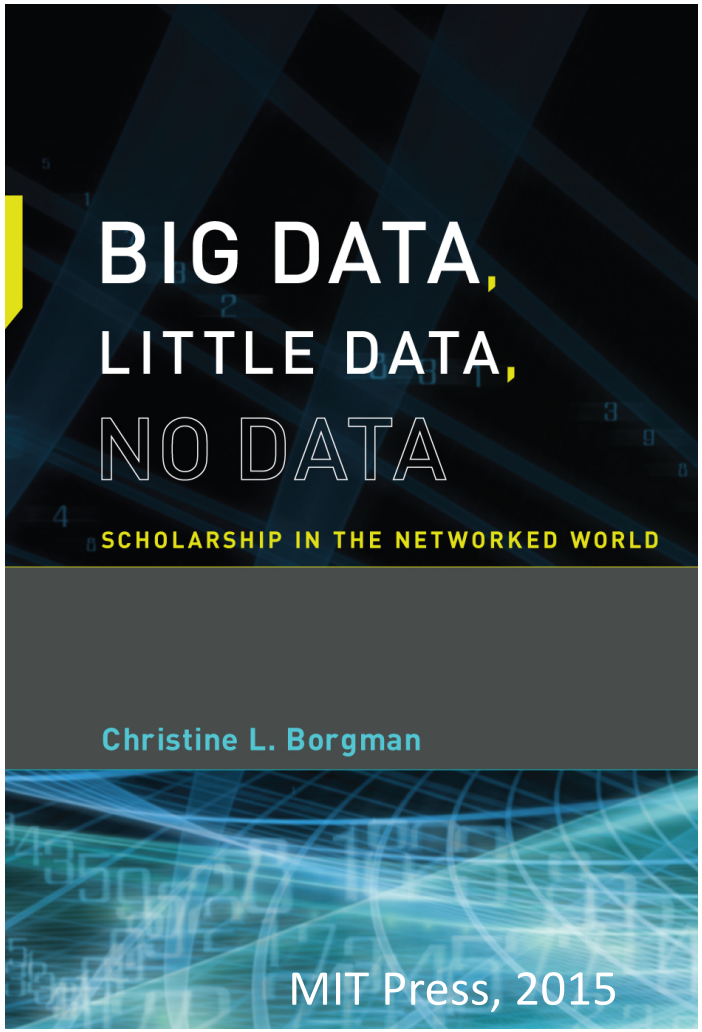
Center for Knowledge Infrastructures

University of California, Los Angeles

http://christineborgman.info

https://knowledgeinfrastructures.gseis.ucla.edu

@scitechprof



Keynote Presentation
National Open Science Plan for France:
From Strategy to Action
Paris, 4 December 2018

- Generalize open access to publications
- Structure research data and make it available through open access
- Be part of a sustainable European and international open science dynamic

Knowledge Infrastructures

"robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds"

Edwards, P. N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Knowledge Infrastructures

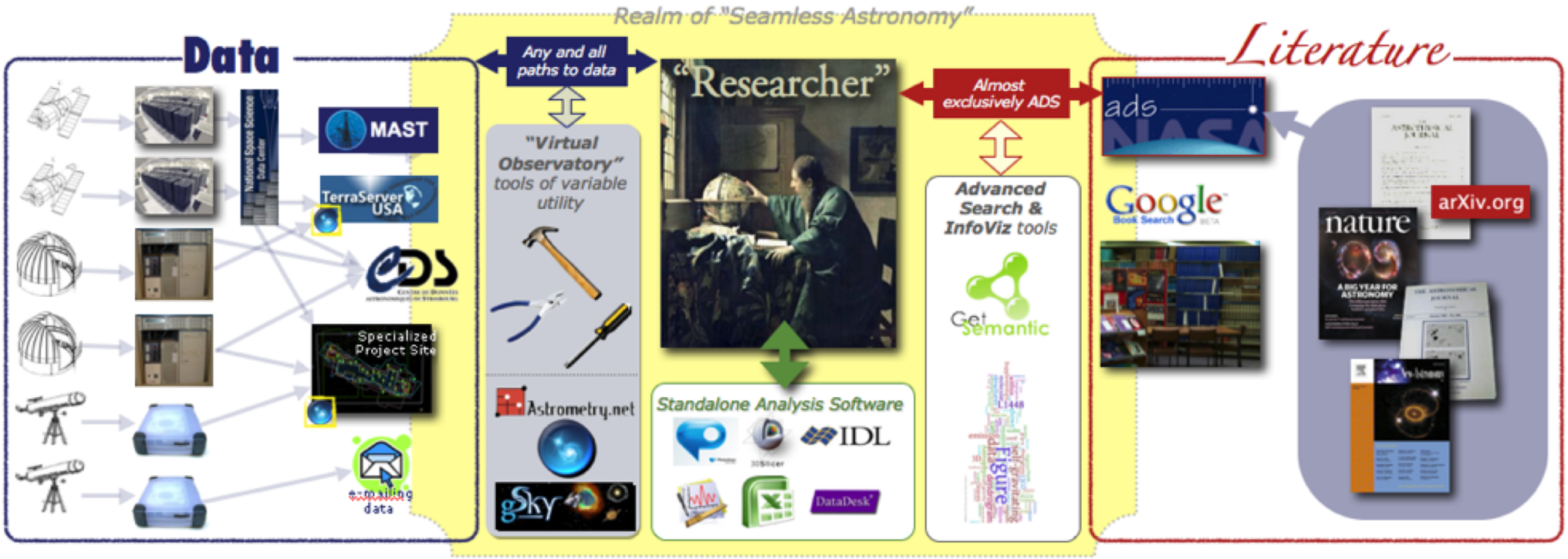
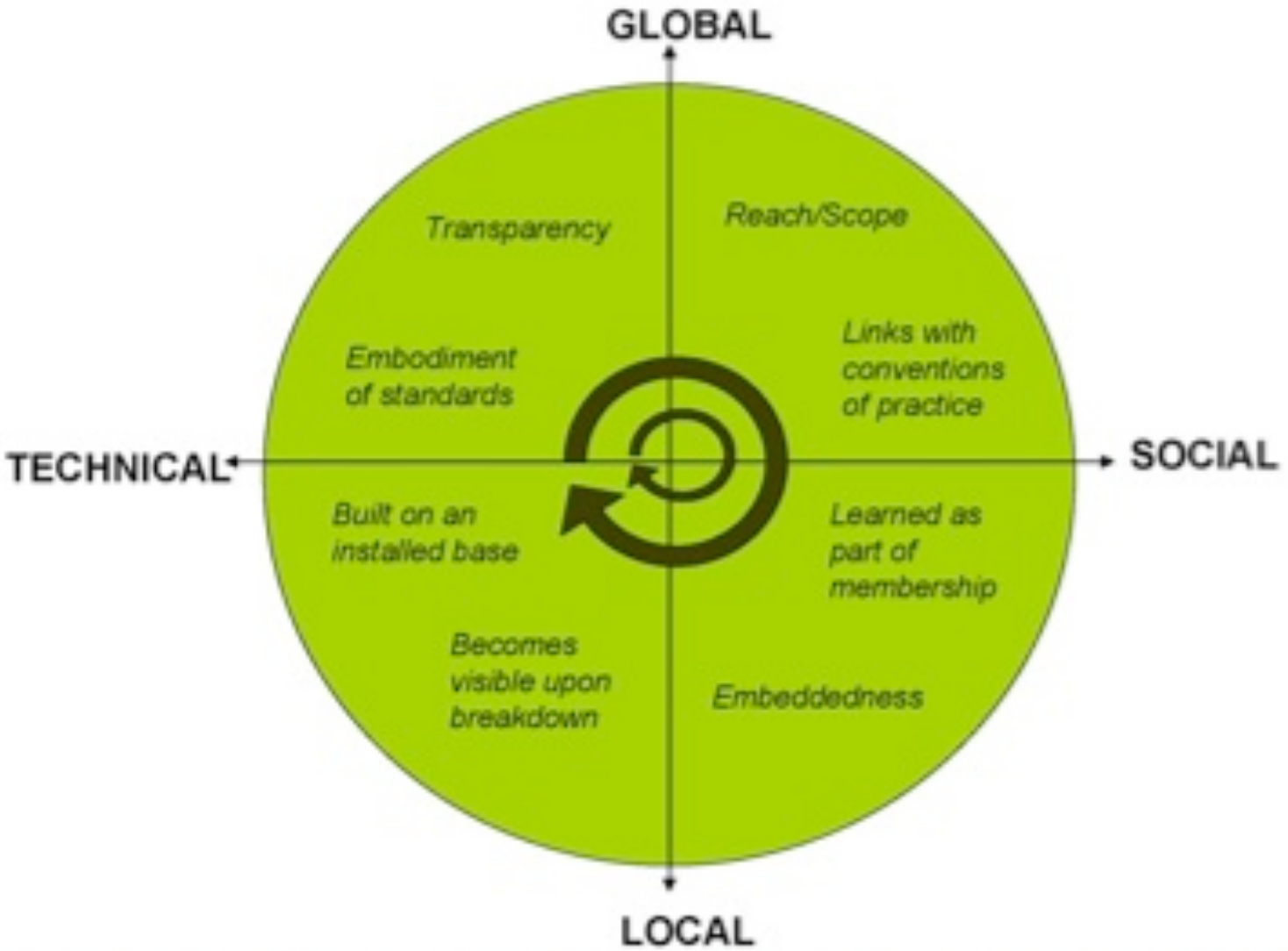


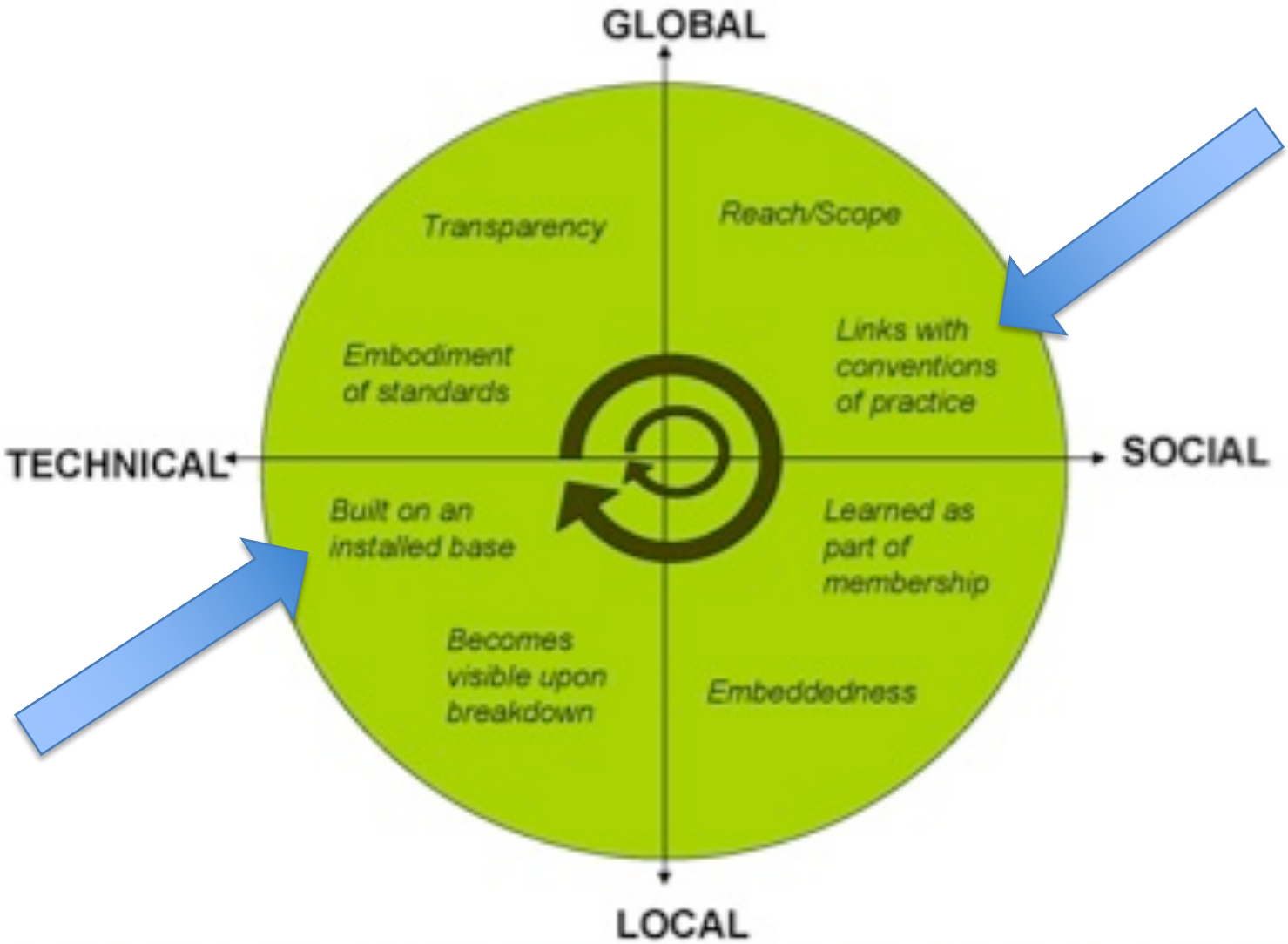
Image: Alyssa Goodman, Harvard-Smithsonian Center for Astrophysics

Infrastructures



Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111 Image by Florence Millerand.

Infrastructures



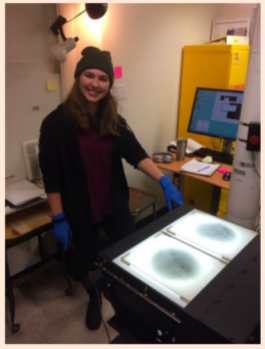
Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111 Image by Florence Millerand.

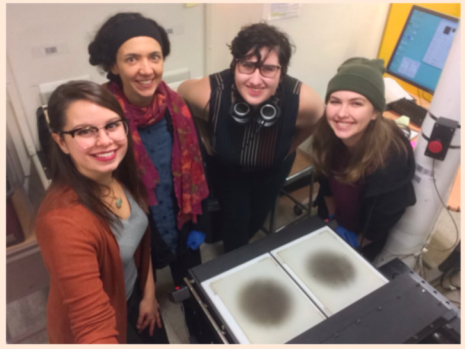
Opportunities in Open Science

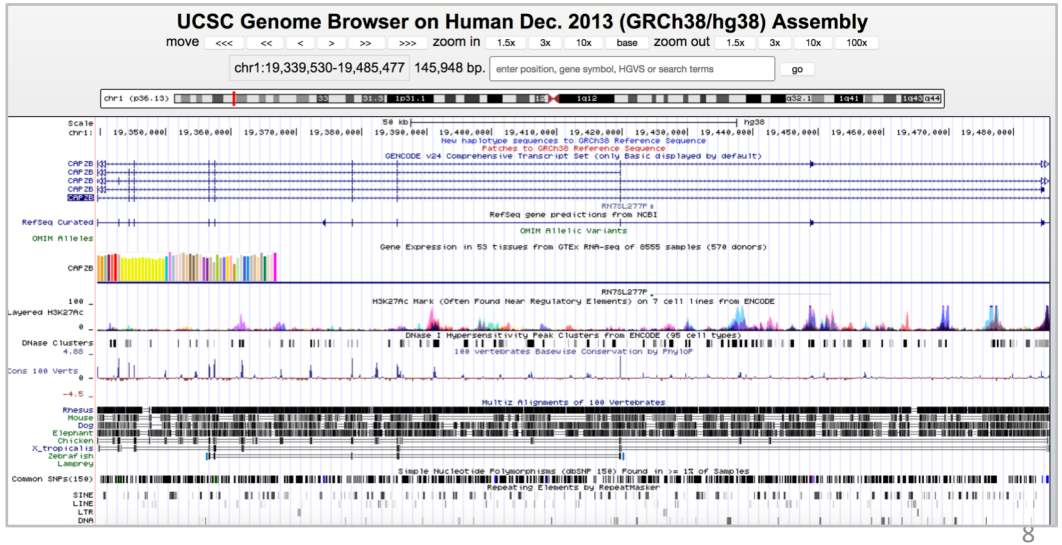
New knowledge from old data

DASCH
Digital Access to a Sky Century @ Harvard
A New Look at the Temporal Universe

Current Status: 301,991 Plates Scanned 20,401,000,000 Magnitudes.

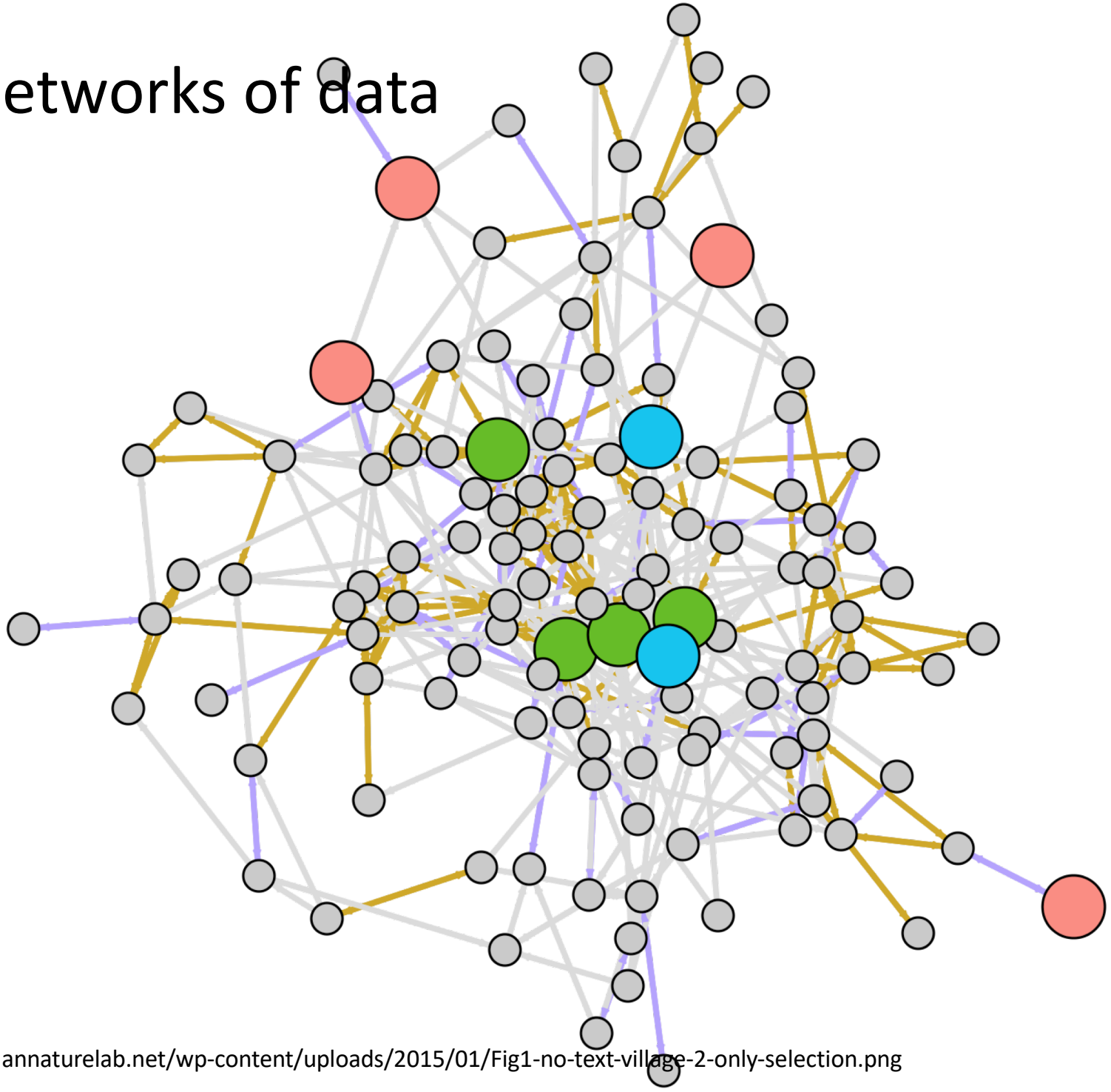




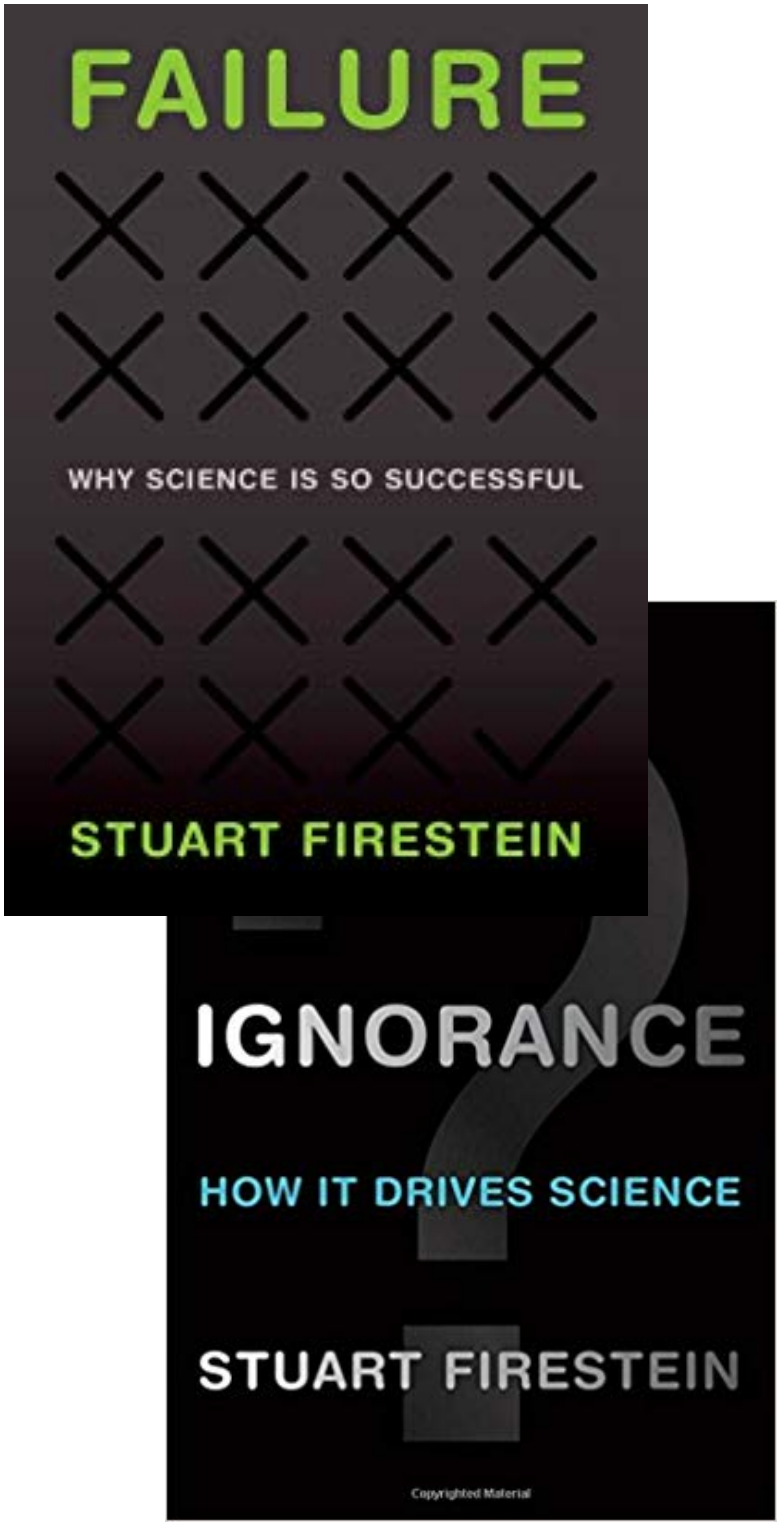


- https://library.cfa.harvard.edu/image-vocab/harvard-computers

Networks of data



http://humannaturelab.net/wp-content/uploads/2015/01/Fig1-no-text-village-2-only-selection.png



The Undiscovered: Many great discoveries in science are surprises.

https://www.radcliffe.harvard.edu/event/2018-undiscovered-symposium

Challenges in Open Science

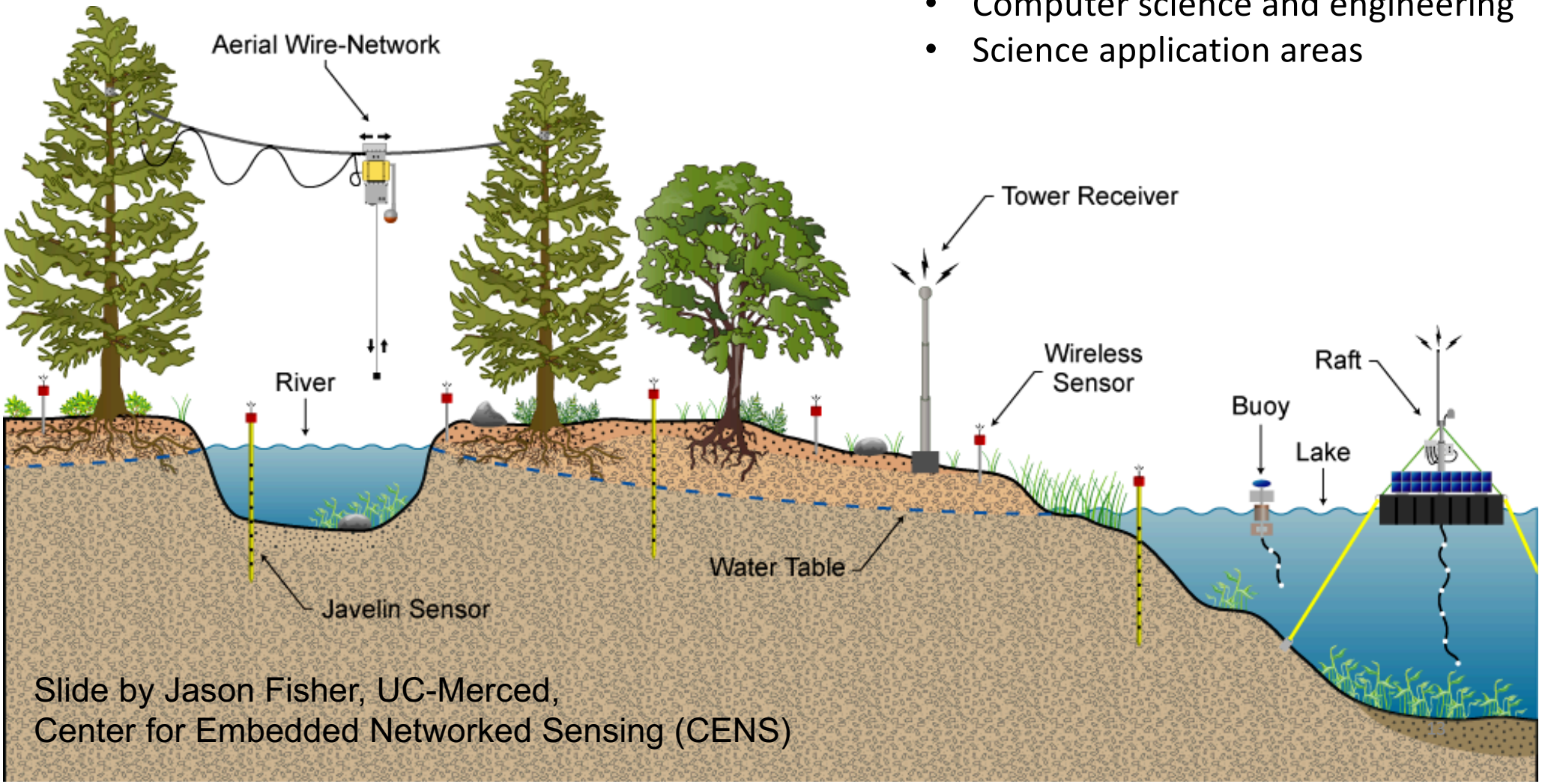
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

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

http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85327

Center for Embedded Networked Sensing

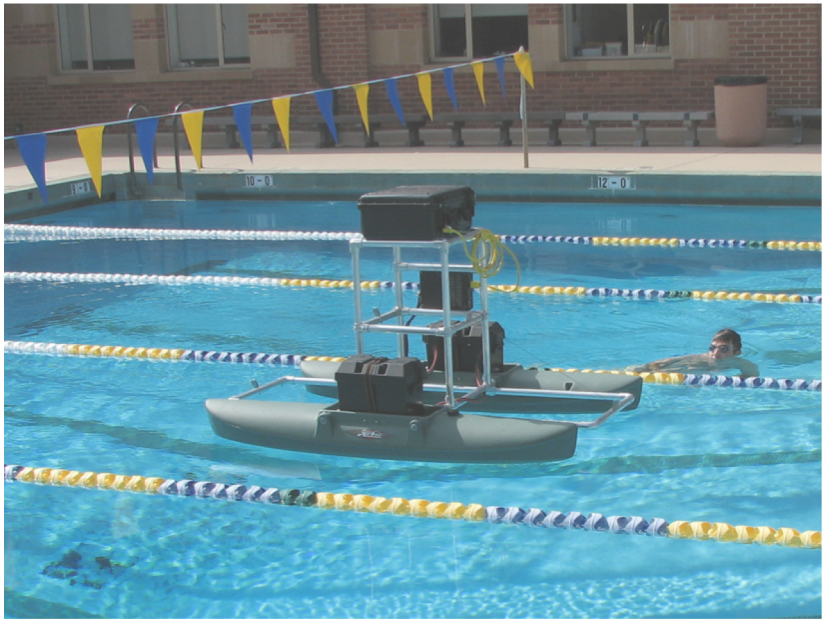
- NSF Science & Tech Ctr, 2002-2012
- 5 universities, plus partners
- 300 members
- Computer science and engineering
- Science application areas



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

Science <–> Data

Engineering researcher:
“Temperature is temperature.”



CENS Robotics team

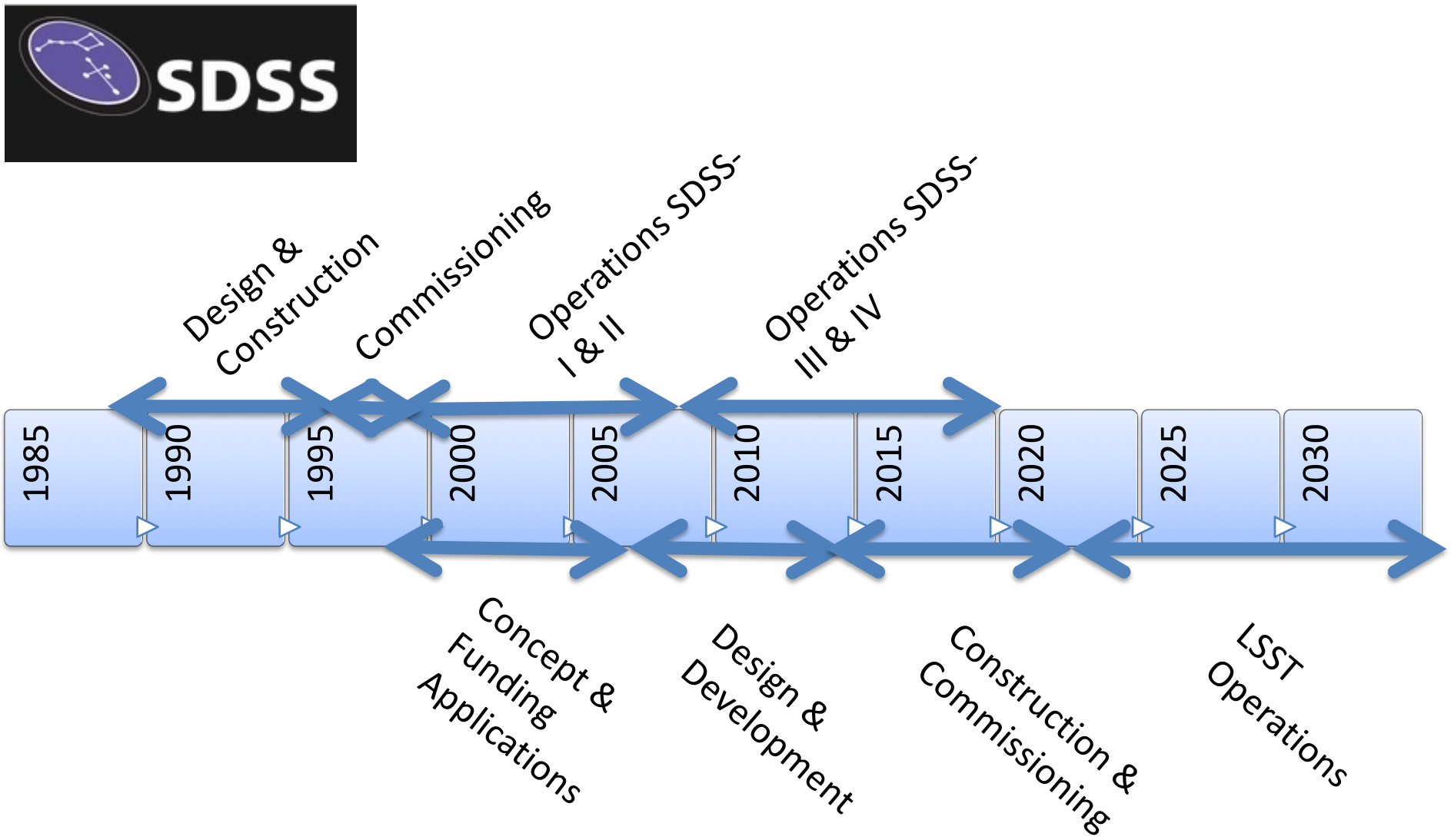
Science <–> Data

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.."*

Project Timelines

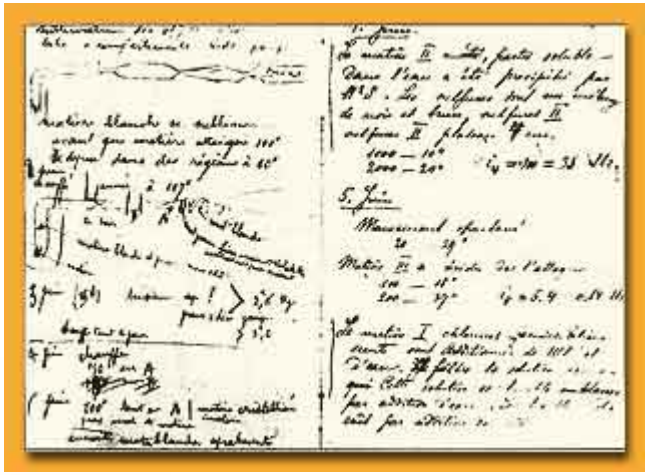


LSST Timeline: https://www.lsst.org/about/timeline

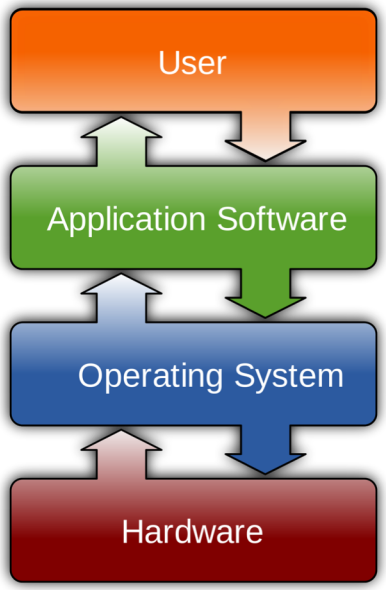
Large Synoptic Survey Telescope

Data Stewardship

Alibaba.com



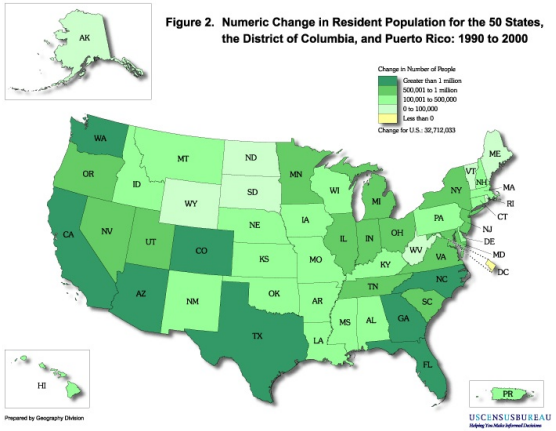
Marie Curie's notebook aip.org



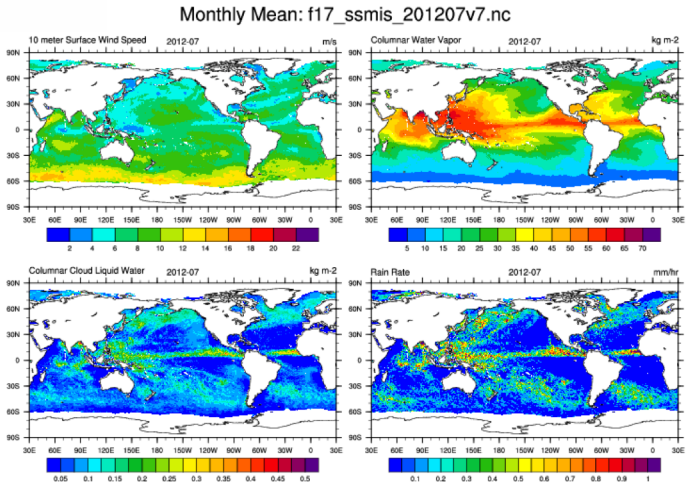
Wikipedia.org

Pisa Griffin

hudsonalpha.org



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



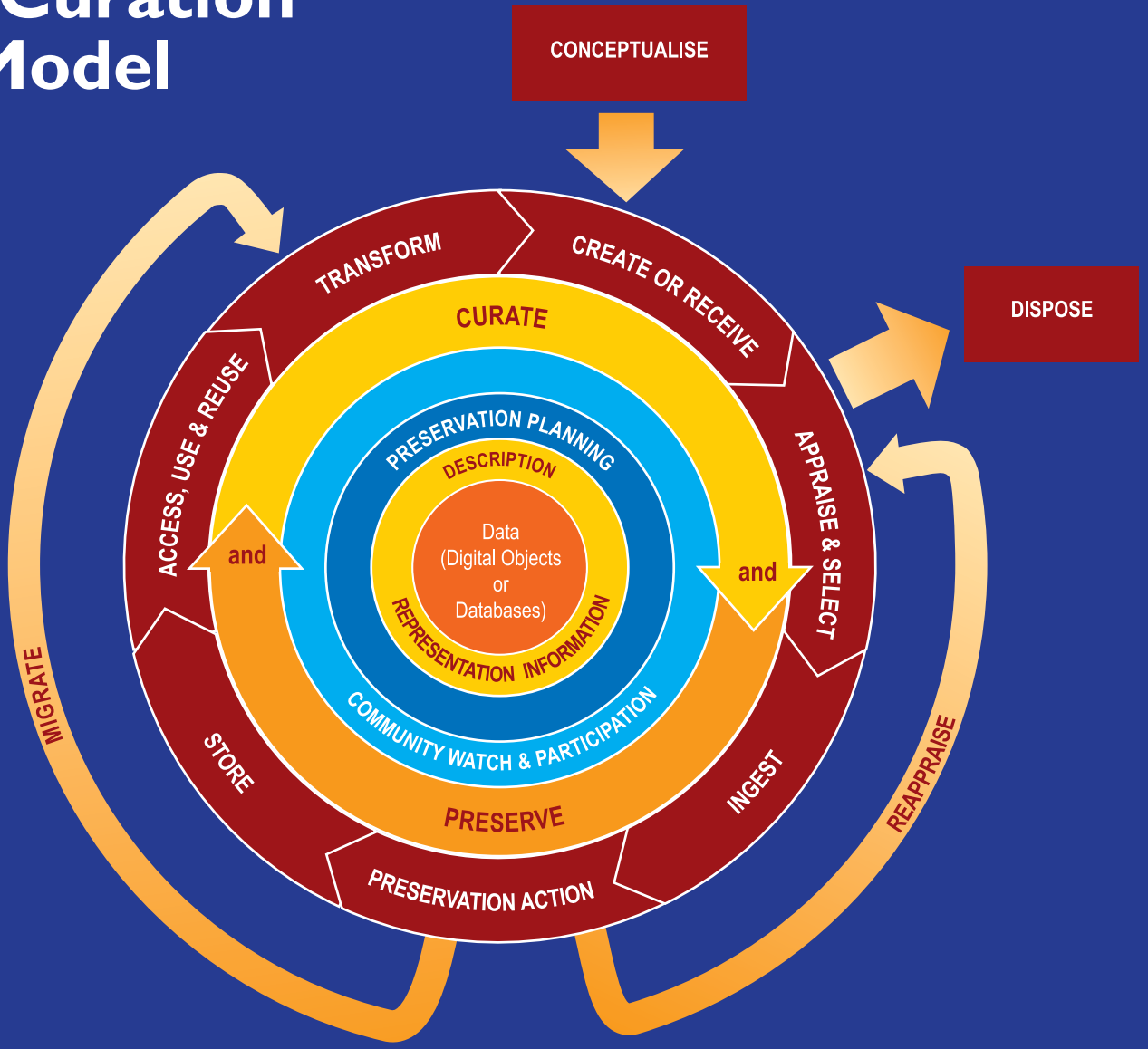
ncl.ucar.edu

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 Beledive and works all day in Kayseri.

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

The DCC Curation Lifecycle Model

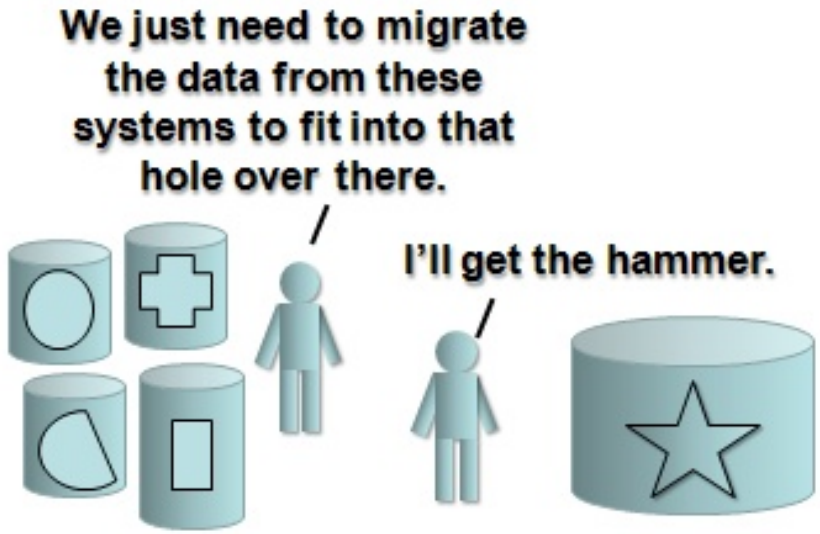


Data Stewardship: the Reality

http://www.information-age.com/cloud-computing-pharmaceutical-industry-123462676/

Getty Research Institute

Mount Wilson Solar Observatory, 2017



http://www.datamartist.com/data-migration-part-1-introduction-to-the-data-migration-delema

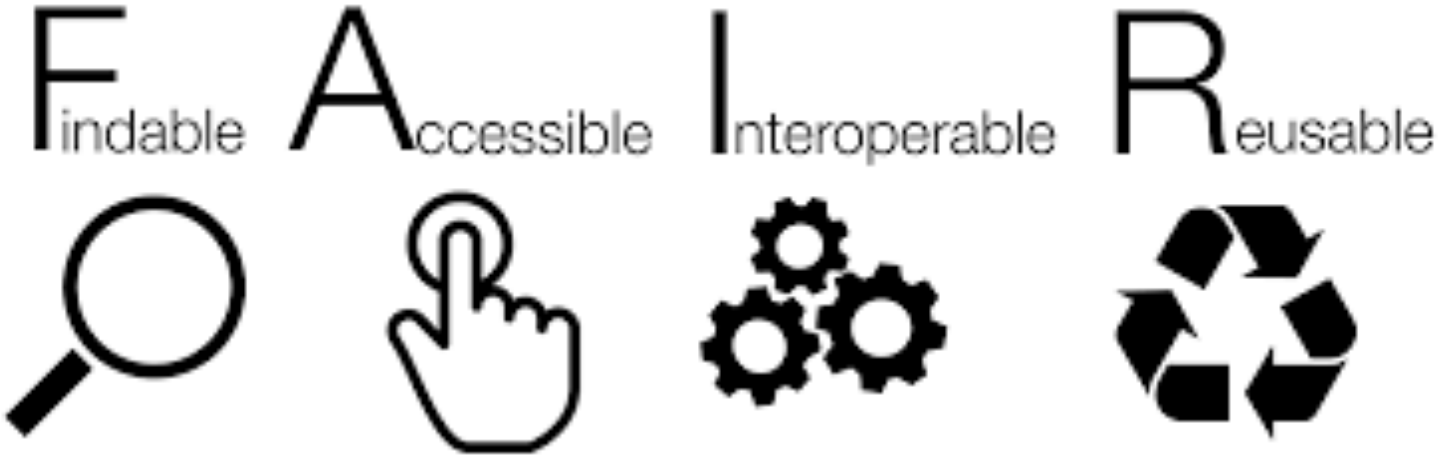
http://gsa.rice.edu/

Graduate students

https://med.nyu.edu/our-community/life-nyu-school-medicine/life-postdoc

Post-doctoral fellows

Data Stewardship: The Ideal



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, http://dx.doi.org/10.1038/sdata.2016.18

20

OPEN ACCESS Freely available online

Modeling 3D Facial Shape from DNA

Peter Claes[1], Denise K. Liberton[2], Katleen Daniels[1], Kerri Matthes Rosana[2], Ellen E. C
Laurel N. Pearson[2], Brian McEvoy[3], Marc Bauchet[2], Arslan A. Zaidi[2], Wei Yao[2], Hua T
Gregory S. Barsh[4,5], Devin M. Absher[5], David A. Puts[2], Jorge Rocha[6,7], Sandra Beleza
Rinaldo W. Pereira[9], Gareth Baynam[10,11,12], Paul Suetens[1], Dirk Vandermeulen[1], Jen
James S. Boster[14], Mark D. Shriver[2]*

1 Medical Image Computing, ESAT/PSI, Department of Electrical Engineering, KU Leuven, Medical Imaging Research Center, KU Leuven & Future Health Department, Leuven, Belgium, **2** Department of Anthropology, Penn State University, University Park, Pennsylvania, Unite Institute of Genetics, Dublin, Ireland, **4** Department of Genetics, Stanford University, Palo Alto, California, United States of America, Biotechnology, Huntsville, Alabama, United States of America, **6** CIBIO: Centro de Investigação em Biodiversidade e Recursos Genéticos, Portugal, **7** Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal, **8** IPATIMUP: Instituto de Patolog Universidade do Porto, Porto, Portugal, **9** Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Br of Paediatrics and Child Health, University of Western Australia, Perth, Australia, **11** Institute for Immunology and Infectious Diseases, Murdo **12** Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Australia, **13** Center for the Integration of Genetic Healthca Pennsylvania, Philadelphia, Pennsylvania, United States of America, **14** Department of Anthropology, University of Connecticut, Storrs, America

SHORT COURSE
HIGHLIGHT

Genomic Data and Models for Political Science

apsa AMERICAN POLITICAL SCIENCE ASSOCIATION

2018 APSA Annual Meeting & Exhibition

THE UNIVERSITY OF CHICAGO PRESS JOURNALS

The Journal of Politics / Vol. 73, No. 1, Jan. 14, 2011 / A Genome-Wide Analys...

JOURNAL ARTICLE

A Genome-Wide Analysis of Liberal and Conservative Political Attitudes

Peter K. Hatemi, Nathan A. Gillespie, Lindon J. Eaves, Brion S. Maher, Bradley T. Webb, Andrew C. Heath, Sarah E. Medland, David C. Smyth, Harry N. Beeby, Scott D. Gordon, Grant W. Montgomery, Ghu Zhu, Enda M. Byrne and Nicholas G. Martin

The Journal of Politics
Vol. 73, No. 1 (Jan. 14, 2011), pp. 271-285

Published by: The University of Chicago Press on behalf of the Southern Political Science Association
DOI: 10.1017/s0022381610001015
https://www.jstor.org/stable/10.1017/s0022381610001015
Page Count: 15

AI that can determine a person's sexuality from photos shows the dark side of the data age

Devin Coldewey @techcrunch / Sep 7, 2017 Comment

Pasquetto, I.V. (2018). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses*.Phd Dissertation. https://escholarship.org/uc/item/1sx7v77r

'This is just the beginning': Using DNA and genealogy to crack years-old cold cases

Police are harnessing consumer DNA sites to solve old murders, which could spur a massive clearing of unsolved crimes.

by Kate Snow and Jon Schuppe / Jul.18.2018 / 4:30 AM ET

SHARE

POLICY FORUM | GENETICS AND PRIVACY

Genealogy databases and the future of criminal investigation

Natalie Ram[1], Christi J. Guerrini[2], Amy L. McGuire[2]

+ See all authors and affiliations

Science

Science 08 Jun 2018:
Vol. 360, Issue 6393, pp. 1078-1079
DOI: 10.1126/science.aau1083

Pasquetto, I. V. (2018). Open Research Data: When Does Reuse Turn into Misuse? *Under Review*.

Paths Toward Open Science

Opportunities and Challenges

- Opportunities
  - Capture scientific products in digital form
  - Store, integrate, generate new knowledge
- Challenges
  - Skills and resources required to curate scientific records
  - Career paths for data science, curation work
  - Sustainability and stewardship of scientific products
  - Uses, reuses, and misuses of scientific products

Sustainable Open Science

- Create career paths
  - Data science
  - Curation and stewardship
- Commit to long-term infrastructure investments
  - Capture and sustain scholarly products
  - Stewardship of knowledge infrastructures
- Promote data reuse
  - Celebrate discovery
  - Anticipate controversy
  - Govern misuse

UCLA Center for Knowledge Infrastructures


Christine Borgman


Bernie Boscoe


Peter Darch


Milena Golshan


Irene Pasquetto


Michael Scroggins


Cheryl Thompson


Morgan Wofford