

UCLA

Presentations

Title

Data, data use, and scientific inquiry: Two case studies of data practices [Presentation slides]

Permalink

<https://escholarship.org/uc/item/0dw030mk>

Authors

Wynholds, Laura A
Wallis, Jillian C.
Borgman, Christine L.
[et al.](#)

Publication Date

2012-06-01

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>



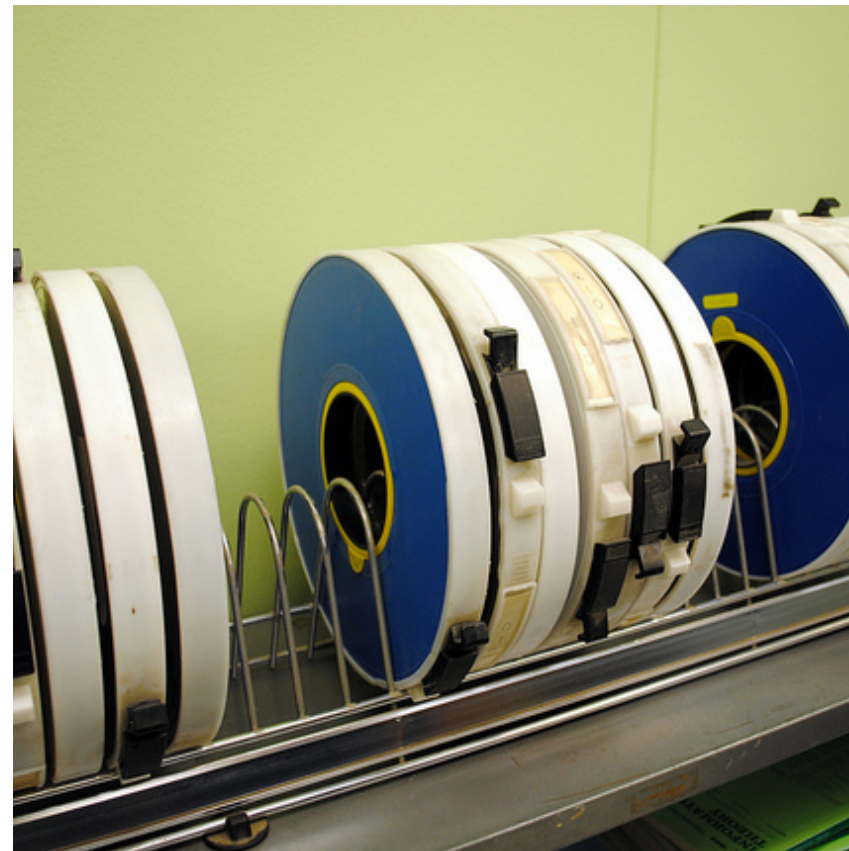
Data, Data Use, and Scientific Inquiry: Two case studies of data practices

*Jillian C. Wallis, Laura A. Wynholds,
Christine L. Borgman, Ashley Sands, & Sharon Traweek*

University of California, Los Angeles

Digital Libraries of Science Data

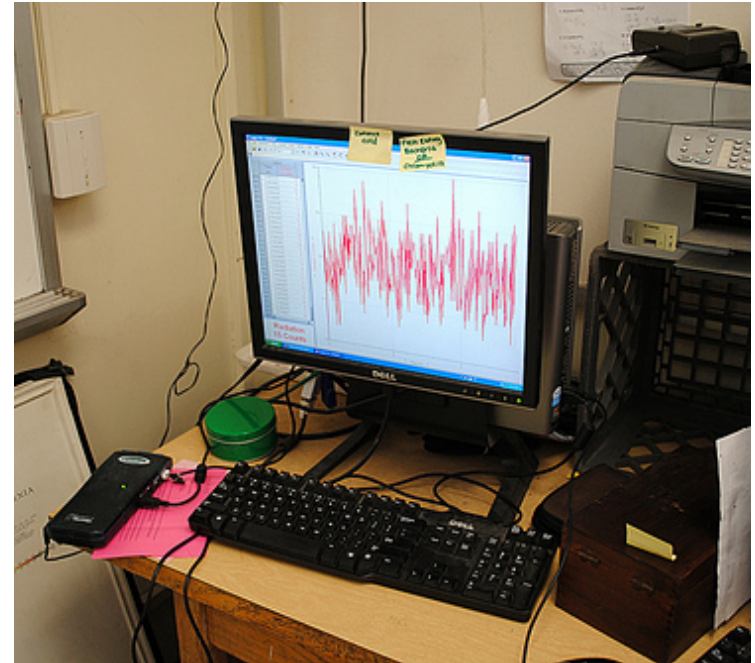
- Capture
- Curate
- Discover
- Use
- Reuse



Fastlizard4's image of tapes from the Kleinrock Internet History Center at UCLA ([flickr.com](https://www.flickr.com/photos/fastlizard4/))

Data creation and use in scientific research

- What are the characteristics of data use and reuse within *each* research community?
- How do characteristics of data use and reuse vary *within* and *between* research communities?



Fastlizard4's image of a Geiger counter setup to measure background radiation (flickr.com)

Research Sites

- Center for Embedded Networked Sensing
 - Science research
 - Environment
 - seismology
 - Technology research
 - Instrumentation
 - networks
 - Long-tail science
 - Circa 300 partners
- Sloan Digital Sky Survey
 - Science research
 - Astronomy
 - Astrophysics
 - Technology research
 - Instrumentation
 - Databases
 - Big science
 - Circa 400 partners



Sloan Digital Sky Survey

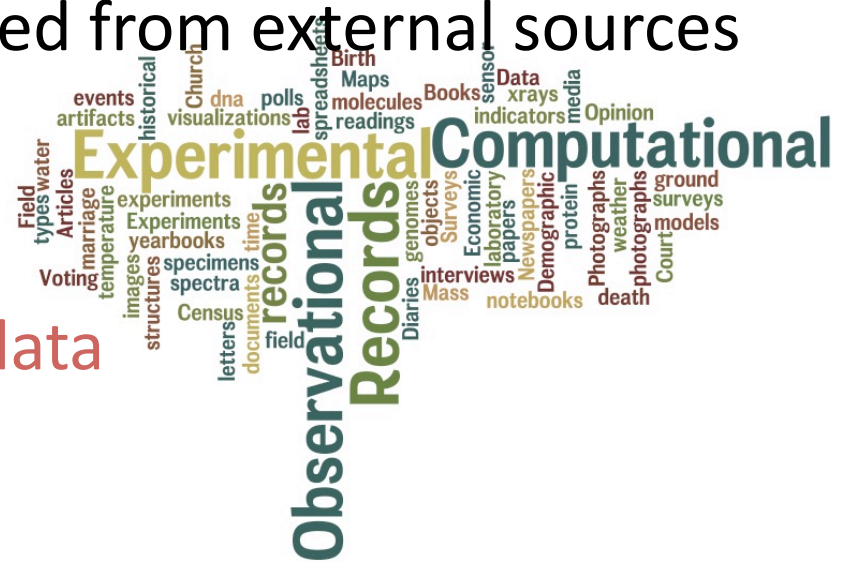
Mapping the Universe

Interview Questions

Topic	Question	CENS	SDSS
Data Types	Within your work, what is typically considered to be “data?”	X	X
	How do you distinguish between different levels or states of data?	X	
Data Sources	What are the main sources of data for your research projects?		X
	Do you routinely or have you ever used data that you did not generate yourself, or from beyond the immediate project team?	X	X
Data Use	When you look at data, what are you hoping to find in it?	X	X
	When, if ever, do you reuse your datasets?	X	X

Dimensions of Data

- Observed vs. simulated data
- Lab generated vs. field collected
- Collected by team vs. obtained from external sources
- Old vs. new data
- Raw vs. processed data
- Background vs. foreground data



Long-Lived Digital Data Collections. (2005). National Science Board. <http://www.nsf.gov/pubs/2005/nsb0540/>

Foreground vs Background

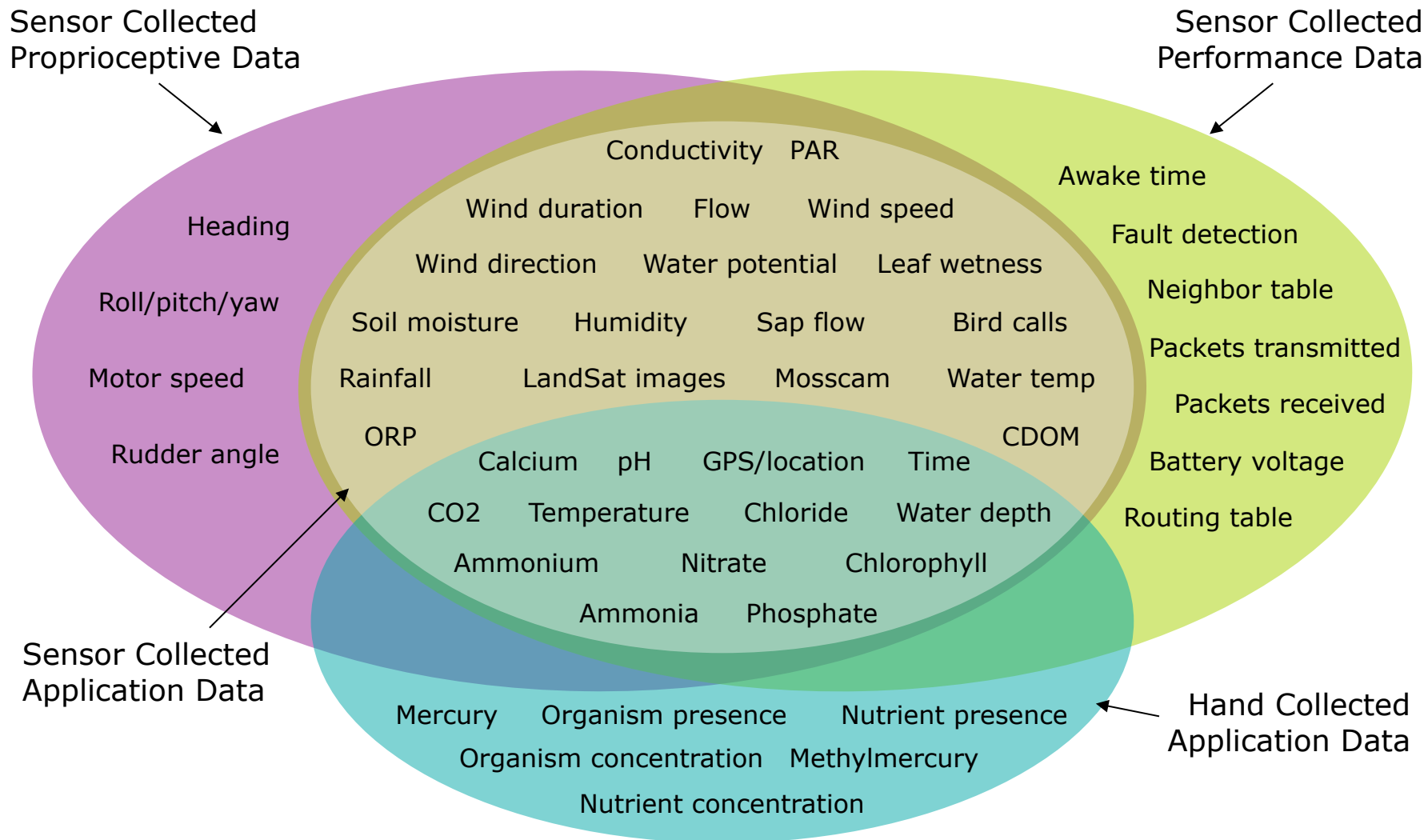
	Foreground data	Background data
Uses	Research questions	Comparison, calibration
Reuses	Internal data sources	External data sources
Disposition	Retain, curate	Discard
Value	Reference in paper	Rarely cited



CENS Data: Foreground vs background

CENTER FOR EMBEDDED NETWORKED SENSING

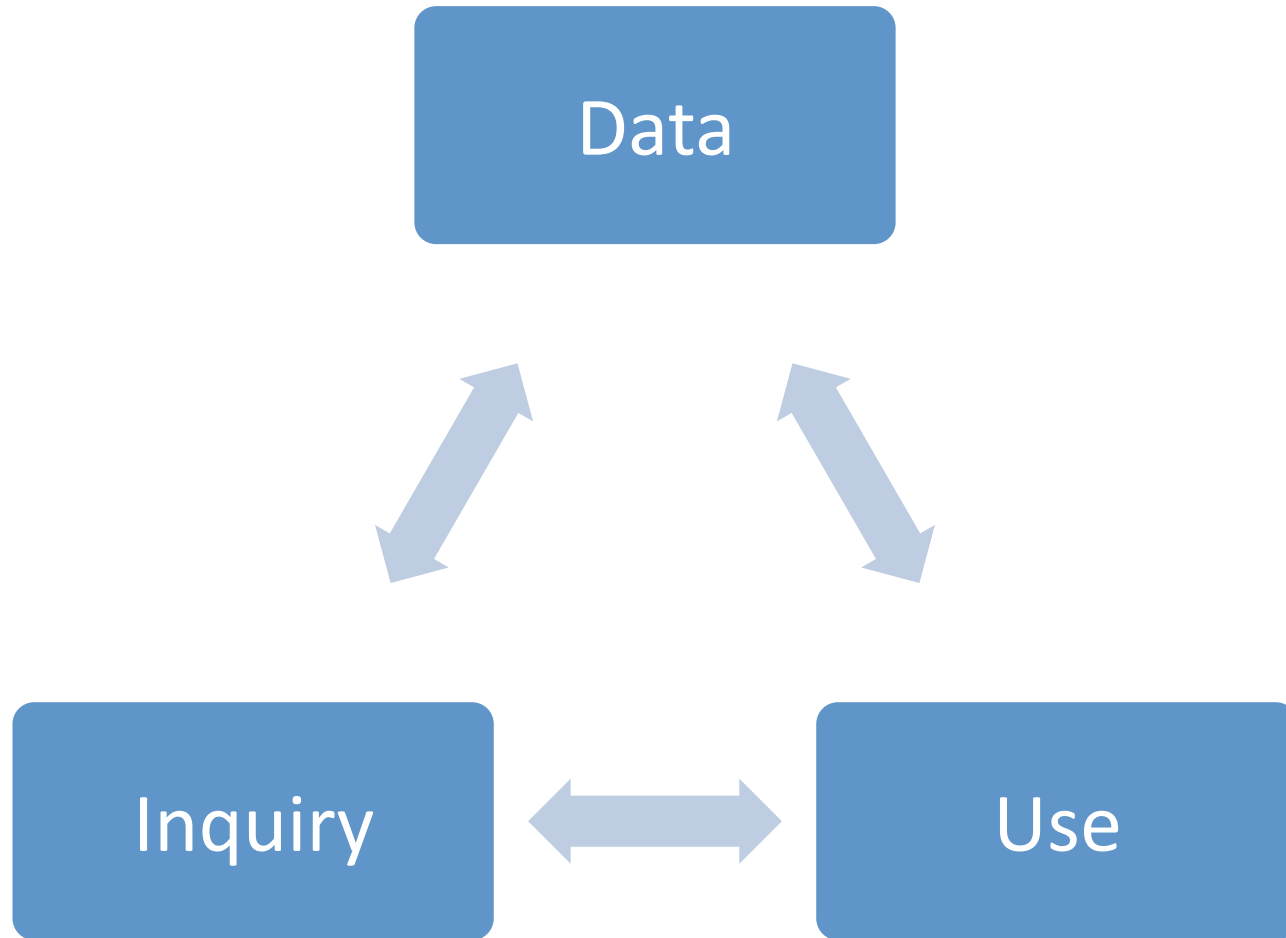
UCLA USC UCR CALTECH UCM



Astronomy data: Foreground vs. background

Type	Source Named	Genre
Catalog (Data) index	SIMBAD, VizieR	Obs
Curated Data Collection	NASA Exoplanet Database	Obs
Data Archive	Multi-mission Archive at STScI (MAST), Infrared Science Archive (IRSA)	Obs
Federated Data Query Services	Virtual Observatory Services (NVO, IVOA)	Obs
Ground Based Instruments	DEep Imaging Multi-Object Spectrograph (DEIMOS), Keck Observatories, Laser Interferometer Gravitational-Wave Observatory (LIGO)	Obs
Ground Based Sky Surveys	Deep Lens Survey, DEEP2 Galaxy Redshift Survey, Catalina Transients Survey, Palomar-Quest Survey, Sloan Digital Sky Survey (SDSS), Digitized Palomar Observatory Sky Survey (DPOSS), SDSS Value Added Catalogs	Obs
Physical Constants	NIST Atomic Spectra Database	Exp
Publications Index	SAO/NASA Astrophysics Data System	Mixed
Simulation	Millennium Simulation Database	Sim
Space Based Instruments	Chandra X-Ray Observatory, Fermi Large Area Telescope, Far Ultraviolet Spectroscopic Explorer (FUSE), Galaxy Evolution Explorer (GALEX), Hubble Space Telescope, Spitzer Space Telescope, XMM X-ray Telescope	Obs
Space Based Sky Surveys	Two Micron All Sky Survey (2MASS), Infrared Astronomical Satellite Survey (IRAS), Wide-field Infrared Survey Explorer (WISE)	Obs

Interdependencies



Conclusions

- Uses of data vary by type of inquiry
- Foreground data
 - Research questions
 - Curated
 - Cited
- Background data
 - Necessary for comparison or calibration
 - Rarely curated
 - Rarely cited
- Value of data lies in its use
- “Use” of data does not follow usual DL metrics



Acknowledgements

Research reported here is supported in part by grants from the National Science Foundation and the Alfred P. Sloan Foundation:

The Transformation of Knowledge, Culture, and Practice in Data-Driven Science: A Knowledge Infrastructures Perspective, Sloan Award # 20113194, CL Borgman, UCLA, PI; S Traweek, UCLA, Co-PI

The Data Conservancy, NSF Cooperative Agreement (DataNet) award OCI0830976, Sayeed Choudhury, Johns Hopkins University, PI

The Center for Embedded Networked Sensing (CENS) is funded by NSF Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, PI

Towards a Virtual Organization for Data Cyberinfrastructure, NSF #OCI-0750529, C.L. Borgman, UCLA, PI; G. Bowker, Santa Clara University, Co-PI; Thomas Finholt, University of Michigan, Co-PI

Monitoring, Modeling & Memory: Dynamics of Data and Knowledge in Scientific Cyberinfrastructures: NSF #0827322, P.N. Edwards, UM, PI; Co-PIs C.L. Borgman, UCLA; G. Bowker, SCU and Pittsburgh; T. Finholt, UM; S. Jackson, UM; D. Ribes, Georgetown; S.L. Star, SCU and Pittsburgh