Title

The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management (slides)

Authors
Borgman, Christine L.
Darch, Peter T.
Sands, Ashley E.
et al.

Publication Date
2014-09-01

The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management

Christine L. Borgman, Peter T. Darch, Ashley E. Sands, Jillian C. Wallis, Sharon Traweek

Knowledge Infrastructures Project, UCLA

Digital Libraries 2014, JCDL/TPDL, London, September 2014

http://www.astro.princeton.edu/~rhl/PrettyPictures/NGC2419-core.jpg

THE WHITE HOUSE
WASHINGTON

Open data 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

NSF National Science Foundation
WHERE DISCOVERIES BEGIN

Australian Government
National Health and Medical Research Council

Supported by
wellcome trust

Policy RECommendations for Open Access to Research Data in Europe
RECODE

Knowledge Infrastructures

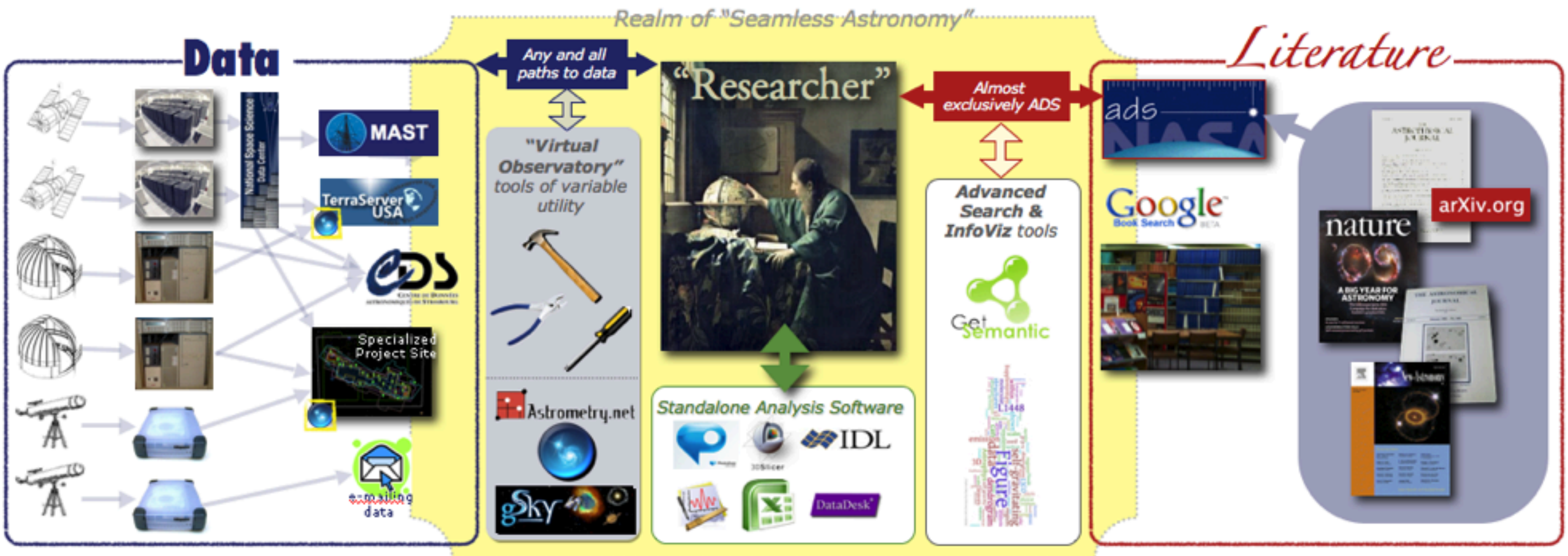


Image: Alyssa Goodman, Astronomy, Harvard

http://knowledgeinfrastructures.org

Knowledge Infrastructures Project
Research Questions

1. What new infrastructures, divisions of labor, knowledge, and expertise are required for data-intensive science?
2. How are the infrastructures of multi-disciplinary, data-intensive scientific endeavors established and how are they dismantled?
3. How do data management, curation, sharing, and reuse practices vary among research areas?
4. What data are most important to curate, from whose perspective, and who decides?

http://knowledgeinfrastructures.gseis.ucla.edu/

Knowledge Infrastructures Project
Research Design

	Big Data	Small Data
Ramping up data collection	Large Synoptic Survey Telescope (LSST)	Center for Dark Energy Biosphere Investigations (C-DEBI)
Ramping down data collection	Sloan Digital Sky Survey, Parts I & II (SDSS)	Center for Embedded Network Sensing (CENS)

Research Methods

Sites	Interviews	People	Institutions	KI Project
CENS	77	72	4	2002-2013
SDSS	118	103	21	2009-
C-DEBI	49	49	16	2012-
LSST	16	10	4	2014-
Total*	260	232	40	

Interviews conducted to date by Knowledge Infrastructures Team

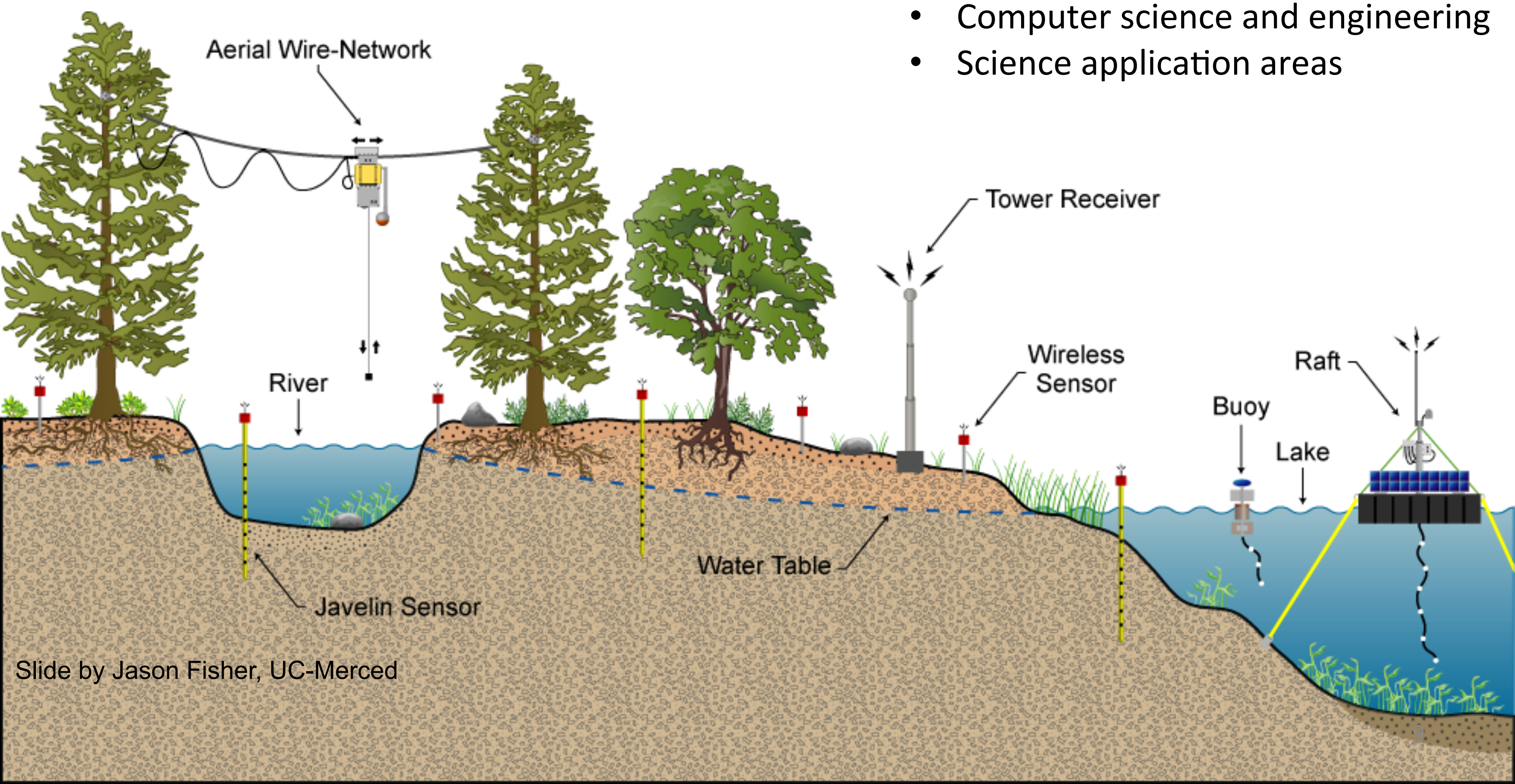
* The cells do not total because of overlapping participation in institutions and projects.

Research Sites and Data

	Big Data	Small Data
Ramping up data collection	Large Synoptic Survey Telescope (LSST) [4]	Center for Dark Energy Biosphere Investigations (C-DEBI) [3]
Ramping down data collection	Sloan Digital Sky Survey, Parts I & II (SDSS) [2]	Center for Embedded Network Sensing (CENS) [1]

Center for Embedded Networked Sensing (CENS)

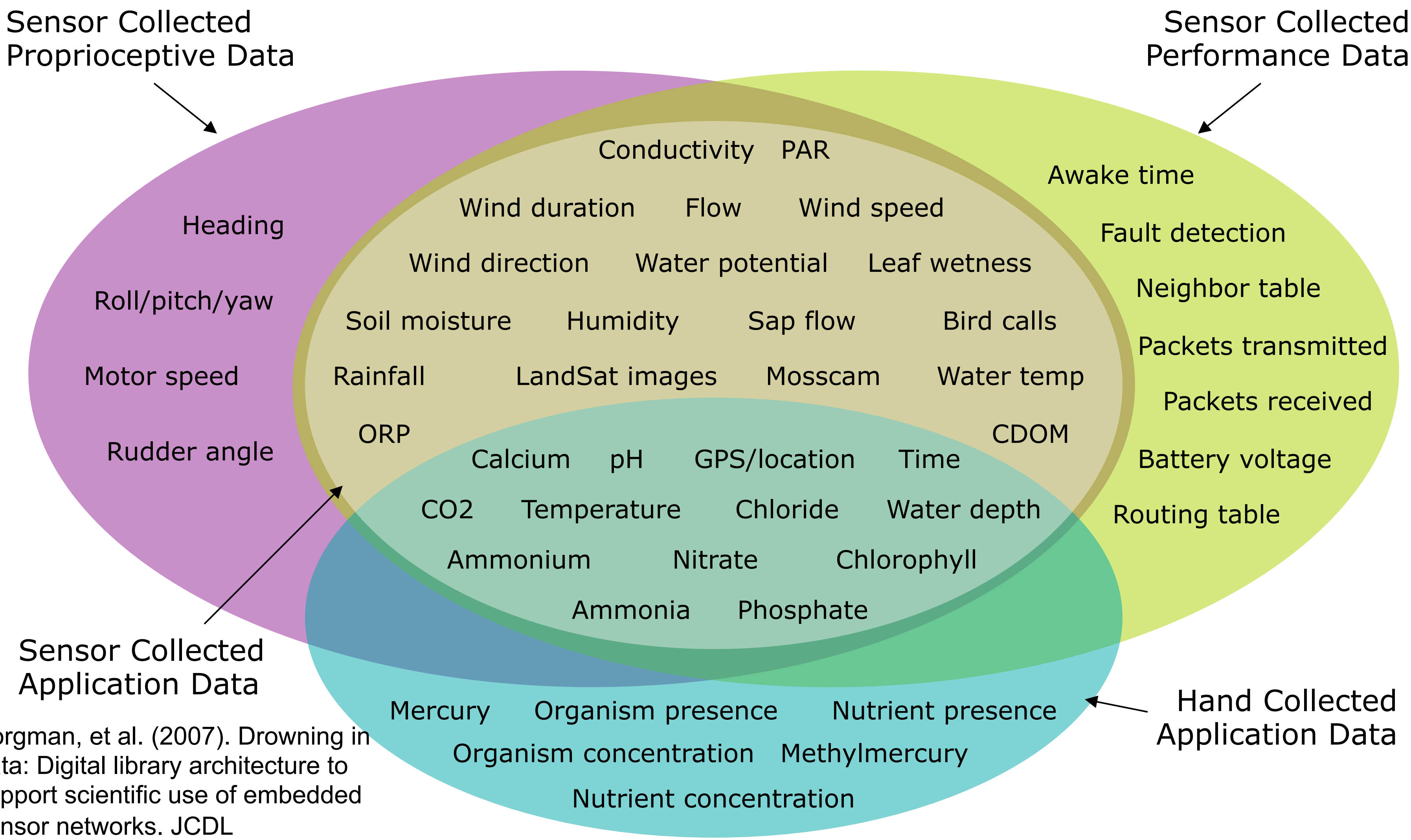
- NSF Science & Tech Ctr, 2002-2012
- 5 universities, plus partners
- 300 members
- Computer science and engineering
- Science application areas



CENS data variation

CENTER FOR EMBEDDED NETWORKED SENSING

UCLA USC UCR CALTECH UCM



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

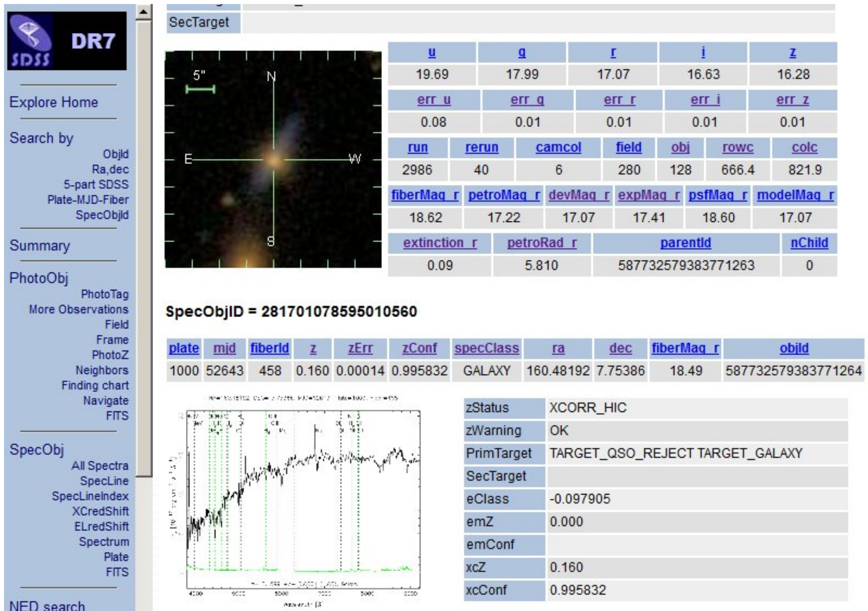
Sloan Digital Sky Survey

Mapping the Universe

- Planning: 1990s
- Data collection (I-II): 2000-2008
- 25 institutions
- 204 members
- Astronomy
- Astrophysics
- Computer science

http://skyserver.sdss3.org/dr10/en/sdss/telescope/telescope.aspx

Data



DR7 SDSS

Explore Home

Search by
- ObjId
- Ra,dec
- 5-part SDSS
- Plate-MJD-Fiber
- SpecObjId

Summary

PhotoObj
- PhotoTag
- More Observations
- Field
- Frame
- PhotoZ
- Neighbors
- Finding chart
- Navigate
- FITS

SpecObj
- All Spectra
- SpecLine
- SpecLineIndex
- XCredShift
- ELredShift
- Spectrum
- Plate
- FITS

NED search

SecTarget	

u	g	r	i	z
19.69	17.99	17.07	16.63	16.28

err_u	err_g	err_r	err_i	err_z
0.08	0.01	0.01	0.01	0.01

run	rerun	camcol	field	obj	rowc	colc
2986	40	6	280	128	666.4	821.9

fiberMag_r	petroMag_r	devMag_r	expMag_r	psfMag_r	modelMag_r
18.62	17.22	17.07	17.41	18.60	17.07

extinction_r	petroRad_r	parentId	nChild
0.09	5.810	587732579383771263	0

SpecObjID = 281701078595010560

plate	mjd	fiberId	z	zErr	zConf	specClass	ra	dec	fiberMag_r	objId
1000	52643	458	0.160	0.00014	0.995832	GALAXY	160.48192	7.75386	18.49	587732579383771264

zStatus	XCORR_HIC
zWarning	OK
PrimTarget	TARGET_QSO_REJECT TARGET_GALAXY
SecTarget	
eClass	-0.097905
emZ	0.000
emConf	
xcZ	0.160
xcConf	0.995832

SkyServer spectroscopy results

M51 Galaxy

http://www.galaxyzooforum.org/index.php?topic=280563.0

http://classic.sdss.org/includes/sideimages/m51.html

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

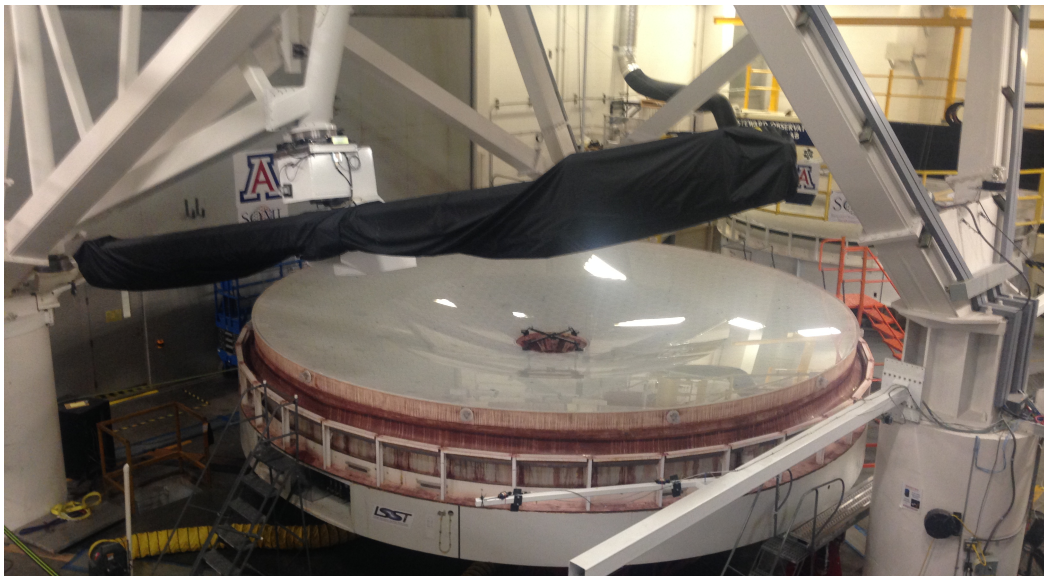
Repository for seafloor cores. Photo: Peter Darch

Data: Subseafloor microbial life

https://sites.google.com/site/adoptamicrobe/home

LSST 2014
CONSTRUCTION START

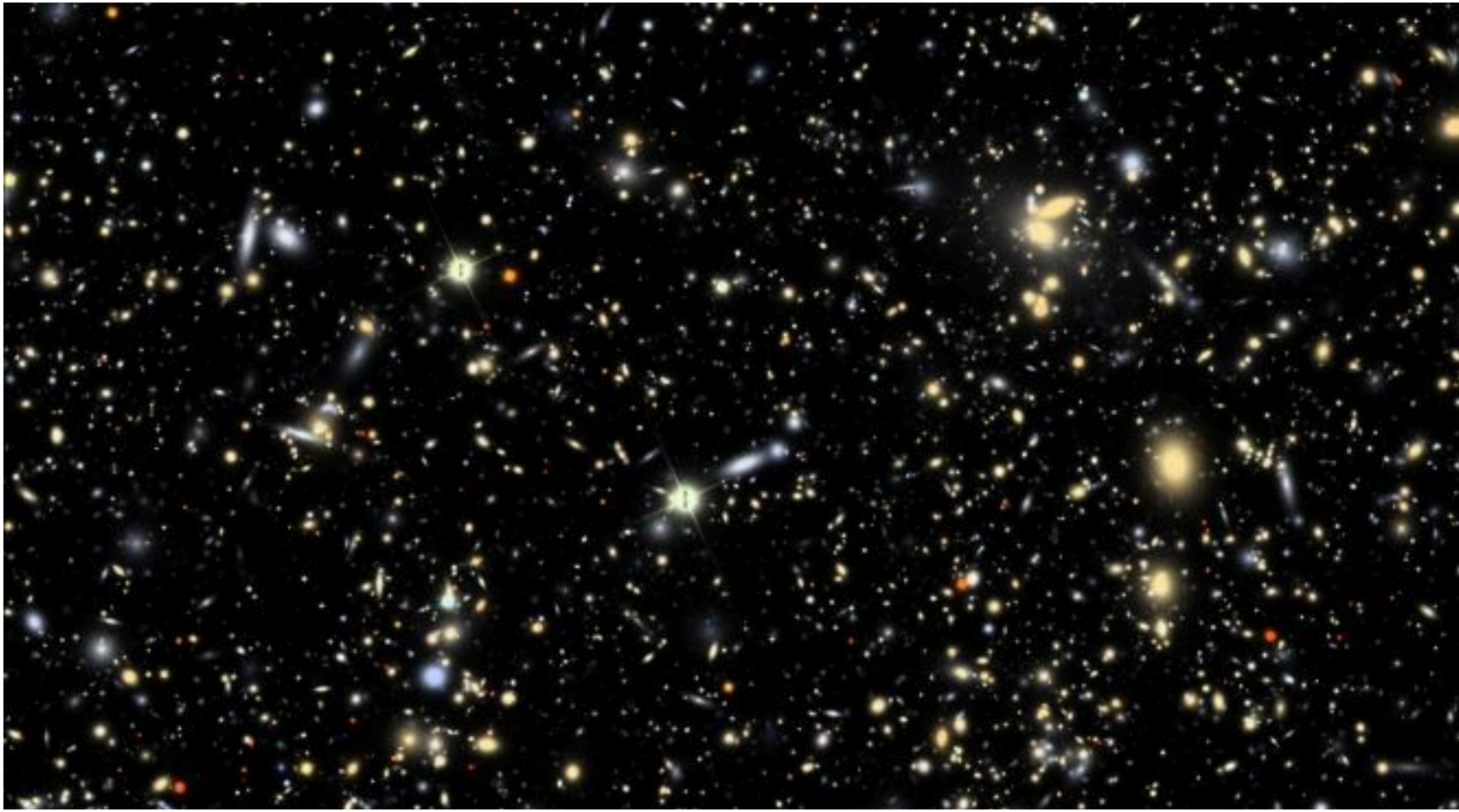
- Planning: 2000s
- Construction: 2014-??
- Data collection: 2022-2032
- Over 100 members
- Astronomy
- Astrophysics
- Computer science



LSST mirror in lab at Arizona State Unviersity.
Photo: Peter Darch

LSST All Hands Meeting, August 2014,
Arizona State University. Arrow to Peter Darch

Data



http://www.lsst.org/lsst/gallery/data/lsst-imsim-october-2010

Pairwise Comparisons of Sites

	Big Data	Small Data
Ramping up data collection	Large Synoptic Survey Telescope (LSST) [4]	Center for Dark Energy Biosphere Investigations (C-DEBI) [3]
Ramping down data collection	Sloan Digital Sky Survey, Parts I & II (SDSS) [2]	Center for Embedded Network Sensing (CENS) [1]

Ramping down: CENS and SDSS

Similarities

- Successful projects
- Research continues after funding ends
- Loose confederations of researchers
- Science-technology partners

Differences

- Scale of data
- Disposition of data
- Centrality of data to research
- Time frame of research
- Data sharing and reuse

Ramping up: C-DEBI and LSST

Similarities

- Infrastructure investments
- Mixture of big and small science
- Planned disposition of data
- Widely distributed partners

Differences

- Temporal scale
- Heterogeneity of expertise
- Heterogeneity of data practices
- Maturity of standards
- Community building

Small data: CENS and C-DEBI

Similarities

- NSF Science-Technology Centers
- Problem oriented
- Community building
- Mixture of big and small science
- Minimal data standards

Differences

- Technology vs. Science focus
- Disposition of data
- Knowledge infrastructure concerns

Big data: SDSS and LSST

Similarities

- Common personnel
- Temporal scale
- Data release

Differences

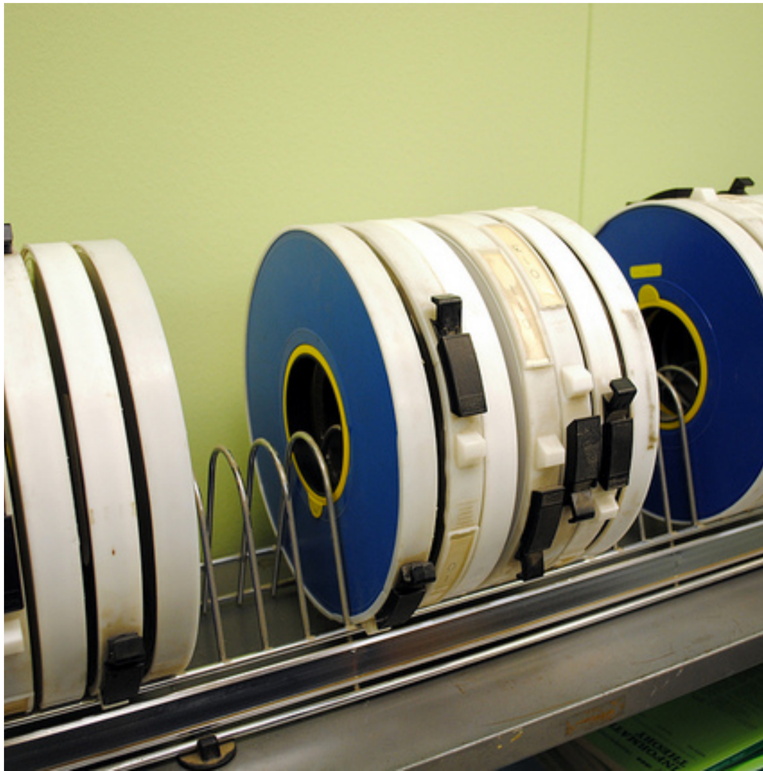
- Range of disciplines
- Scale of data collection
- Release raw vs. curated data

Discussion and Conclusions

- Knowledge infrastructure needs may vary by
  - Temporal scale of research
  - Degree of research coordination
  - Common or competing data standards
  - Release and reuse as goals
  - Communities and governance

Discussion and Conclusions

- Digital libraries for scientific data may vary by
  - Goals of research
  - Scale of data
  - Investments in data stewardship
  - Locus of digital library expertise



Fastlizard4: tapes from the Kleinrock Internet History Center at UCLA (flickr.com)

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