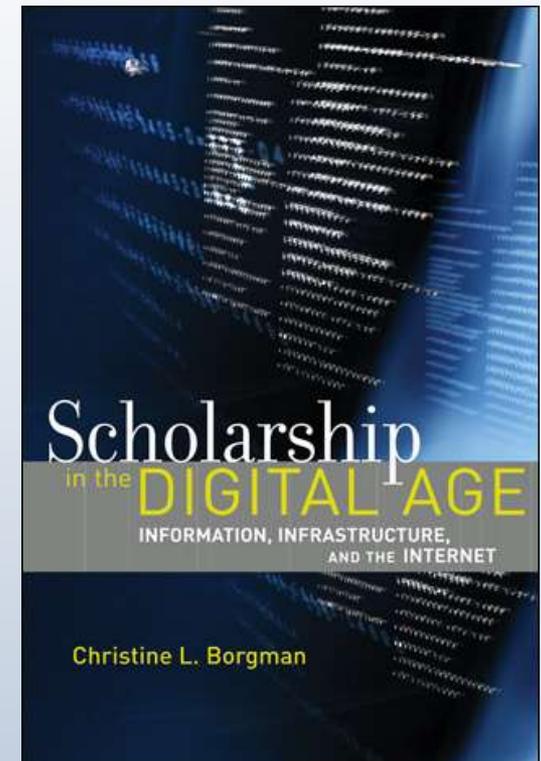


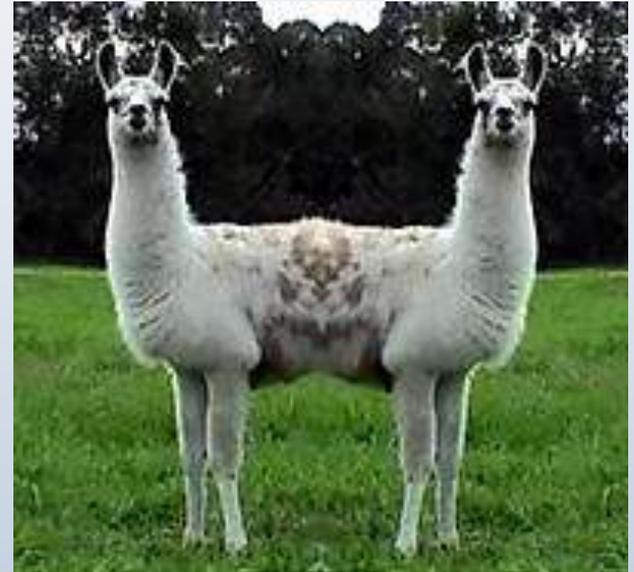
Scholarly Information Infrastructure

- Cyberinfrastructure, e-Science, e-Social Science, e-Humanities, ...e-Research
- Goal: enable new forms of scholarship that are
 - information-intensive
 - data-intensive
 - distributed
 - collaborative
 - multi-disciplinary
- Means: use information technology to
 - improve access to scholarly information
 - collaborate over distance
 - access tools, services, content at distributed sites



Driving Forces

- Technology push
 - Data deluge from highly instrumented research
 - Distributed access to content and computing resources
 - Tools and services for data collection, mining
- Collaboration pull
 - Virtual organizations
 - Share distributed resources
- Social environment
 - Culture of contribution
 - Open access publishing



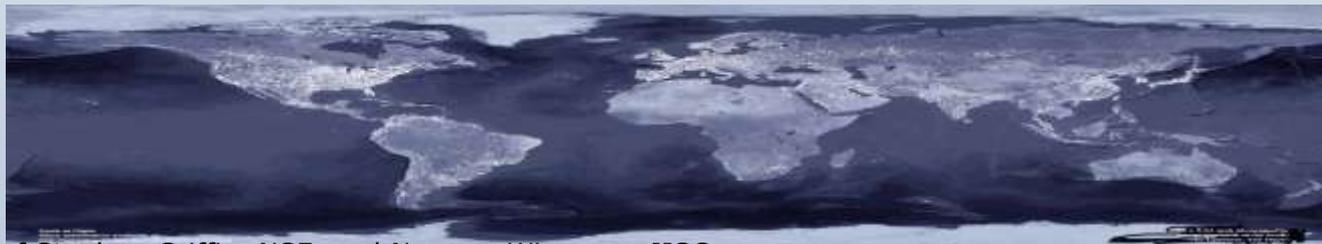
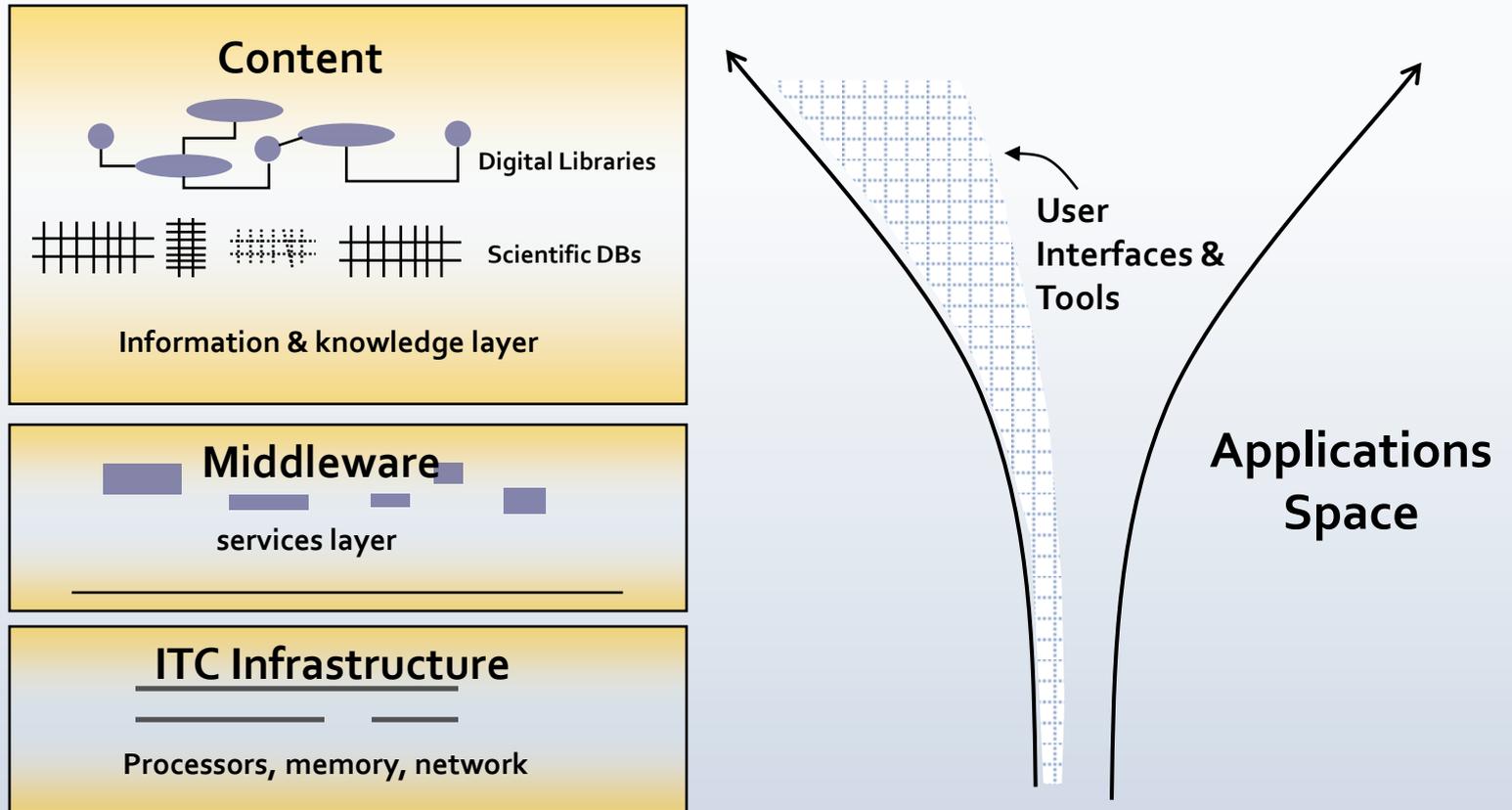
Data are the *glue* of collaborative research

- Collaborative research
 - Data creation
 - Data sharing, reuse
- Leverage research investments
 - Replicate, verify research findings
 - Ask new questions with extant data
- Scholarly capital
 - Human capital
 - Instrumentation
 - Data

<http://www.gridpp.ac.uk/pics/gridpp-group.jpg>

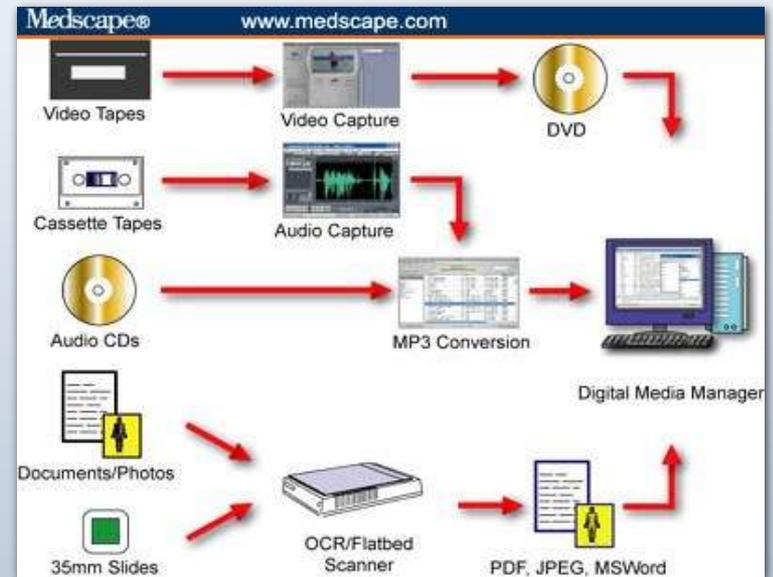


eResearch Infrastructure: Layered Model



Content layer

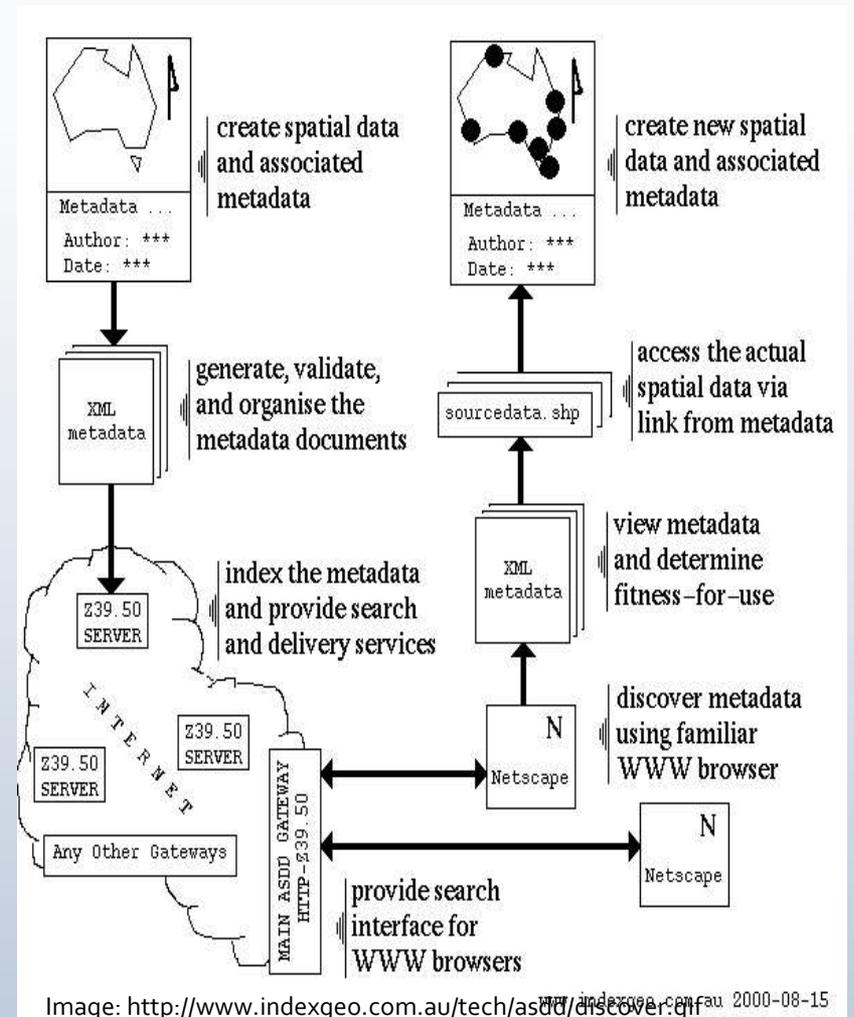
- Documents
 - Publications: books, journals, conference papers, ...
 - Semi-formal: technical reports, working papers, proposals...
 - Unpublished: websites, blogs, wikis...
- Data
 - Observational
 - Computational
 - Experimental
 - Records
- Composite objects



Value chain of information

- Links
 - Cited/citing documents
 - Publications to data sources
 - Data to publications in which reported

- Across boundaries
 - Repositories
 - Publisher databases
 - Disciplines
 - Countries

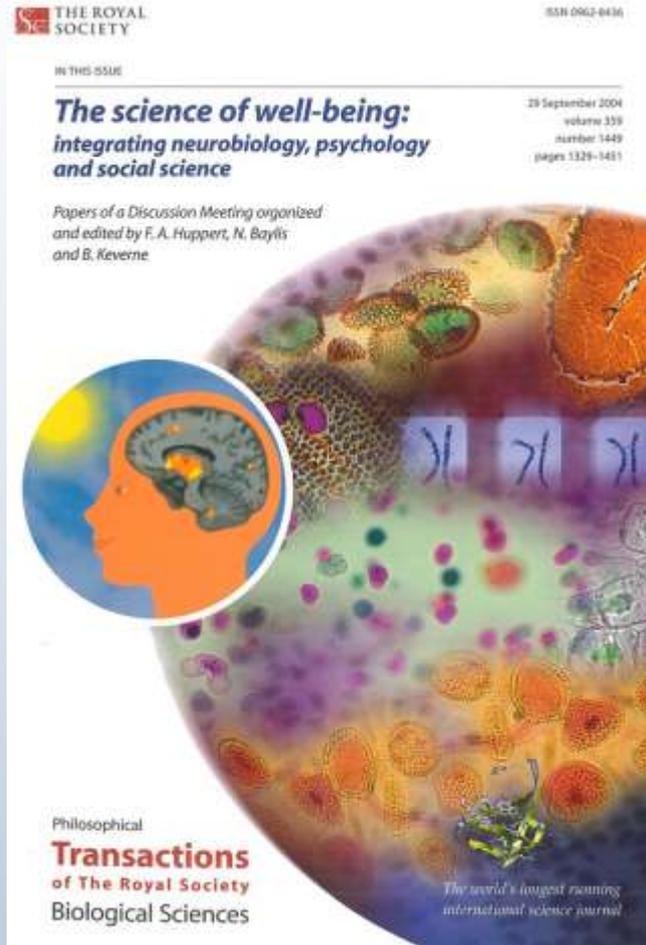


The Scholarly Journal, 1665-2008

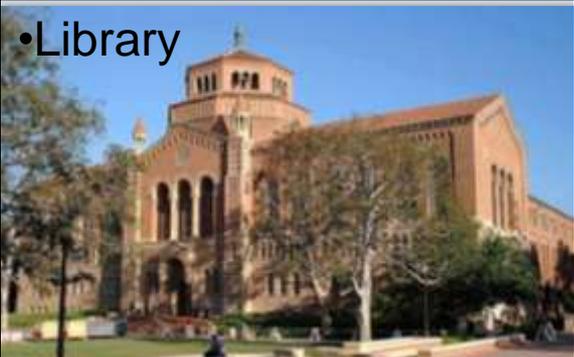
PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD

Vol. I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little with-
out Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.



How do publications enter the value chain?

Function	Print	Digital
Legitimization Authority, quality, priority, trustworthiness	<ul style="list-style-type: none"> •Peer review 	<ul style="list-style-type: none"> •Peer review
Dissemination Awareness, diffusion, publicity	<ul style="list-style-type: none"> •Publisher •Pre-print distribution <ul style="list-style-type: none"> •Copy •Mail 	<ul style="list-style-type: none"> •Publisher •Pre-print distribution <ul style="list-style-type: none"> •Post on Web •Deposit
Access, preservation, curation Availability, discovery, retrieval, persistence	<ul style="list-style-type: none"> •Library 	<ul style="list-style-type: none"> •Library •Publisher •Repository •Homepage

What are data?

Technical definition:

A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen *Reference Model for an Open Archival Information System* (2002).

Socio-technical definition:

“alleged evidence” (Buckland, 2006)

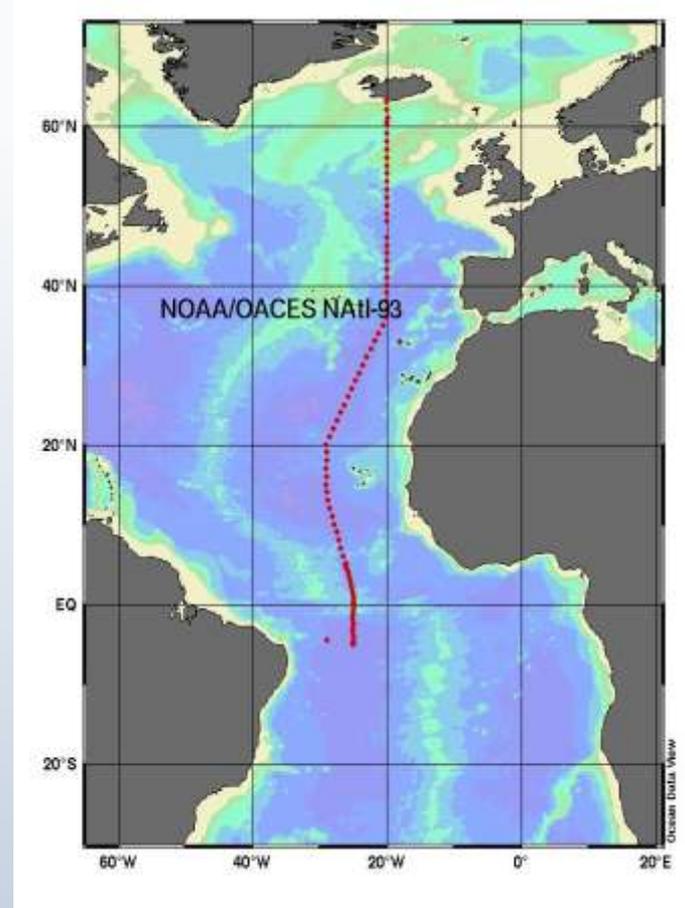


Image: http://cdiac.ornl.gov/oceans/NATl_map.jpg

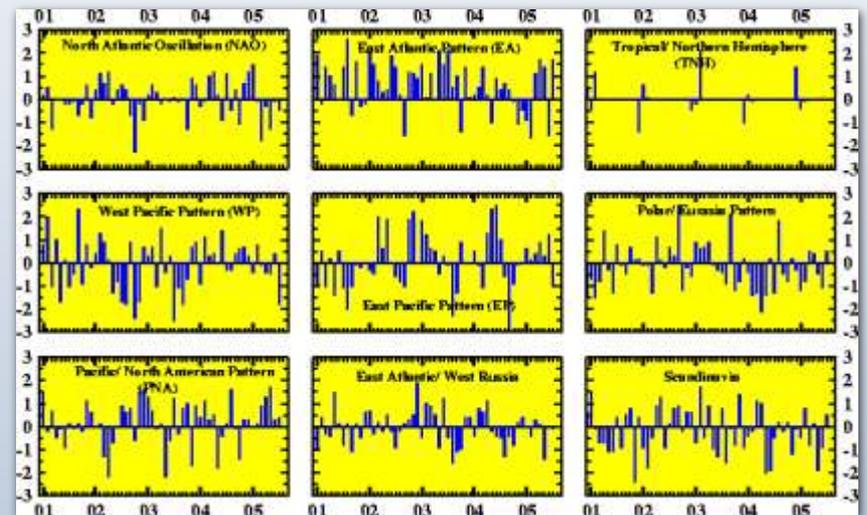
Scientific data

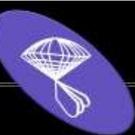
- Examples

- Ecology: weather, ground water, sensor readings, historical record
- Medicine: xrays
- Chemistry: protein structures
- Astronomy: spectral surveys
- Biology: specimens
- Physics: events, objects
- Documentation: Lab and field notebooks, spreadsheets

- Sources

- Generate own data
- Acquire from collaborators, other scientists
- Data repository





Sloan Digital Sky Survey

Mapping the Universe

Explore SDSS

About Us

KeyServer

Image Gallery

News

Education

Credits

Archive

Data Products

Status

Documents

SDSS@Work

Management

Collaboration

Survey

Operations

Publications

Other Links

Contact Us

Search



Scientific and Technical Publications

Scientific papers are based on analyses of, or presentations of, the SDSS data. **Data Release papers** describe the specific process for each data release. **Technical papers** describe the SDSS instrumentation, calibration, software, strategy, and targeting algorithms. Technical papers may include some SDSS data for illustrative purposes. This list represents the definitive list of SDSS papers submitted to peer-reviewed journals. **Other Publications Based on SDSS Data** is a list of publications in journals and astro-ph which use public SDSS data.

Scientific Publications

Title

First Author

astro-ph Journal

A MaxBCG Catalog of 12,875 Galaxy Clusters from the Sloan Digital Sky Survey	B. Koester	0701268 ApJ accepted
Luminosity dependence of the spatial and velocity distributions of galaxies: Semi-analytic models versus the Sloan Digital Sky Survey	C. Li	0701218 MNRAS submitted
Clustering Analyses of 300,000 Photometrically Classified Quasars--II. The Excess on Very Small Scales	A. Myers	0612191 ApJ accepted
Clustering Analyses of 300,000 Photometrically Classified Quasars--I. Luminosity and Redshift Evolution in Quasar Bias	A. Myers	0612190 ApJ accepted
Environment-Dependence of Properties of Galaxies in the Sloan Digital Sky Survey	C. Park	0611610 ApJ accepted
Internal and Collective Properties of Galaxies in the Sloan Digital Sky Survey	Y. Choi	0611607 ApJ accepted
The UV Properties of SDSS Selected Quasars	G. Trammell	0611549 AJ accepted
The Peculiar SN 2005hk: Do Some Type Ia Supernovae Explode as Deflagrations?	M. M. Phillips	0611295 PASP submitted
SDSS J1029+2623: A Gravitationally Lensed Quasar with an Image Separation of 22.5 Arcseconds	N. Inada	0611275 ApJL 653:97 (2006)
Topology of Structure in the Sloan Digital Sky Survey: Model Testing	J. Gott	0610762 ApJ submitted
Broad Absorption Line Variability in Repeat Quasar Observations from the SDSS	B. Lundgren	0610656 ApJ submitted
Low-Mass Dwarf Template Spectra from the SDSS	J. Bochanski	0610639 AJ 133:531 (2007)
The Clustering of Galaxy Groups: Dependence on Mass and other Properties	A. Berlind	0610524 ApJ submitted
3.6-7.9 um Photometry of L and T Dwarfs and the Prevalence of Vertical Mixing in their Atmospheres	S. Leggett	0610214 ApJ accepted
Galaxy Colour, Morphology, and Environment in the Sloan Digital Sky Survey	N. Ball	0610171 MNRAS submitted
A New Survey for Giant Arcs	J. Hennawi	0610061 AJ submitted
Using the Galactic Dynamics of M7 Dwarfs to Infer the Evolution of their Magnetic Activity	A. West	0609001 AJ 132:2507 (2006)
Model Atmosphere Analysis of the Weakly Magnetic DZ White Dwarf G165-7	P. Dufour	0608065 ApJ 651:1112 (2006)
Cluster Lensing in the SDSS I: Weak Lensing Profiles	E. Sheldon	ApJ submitted

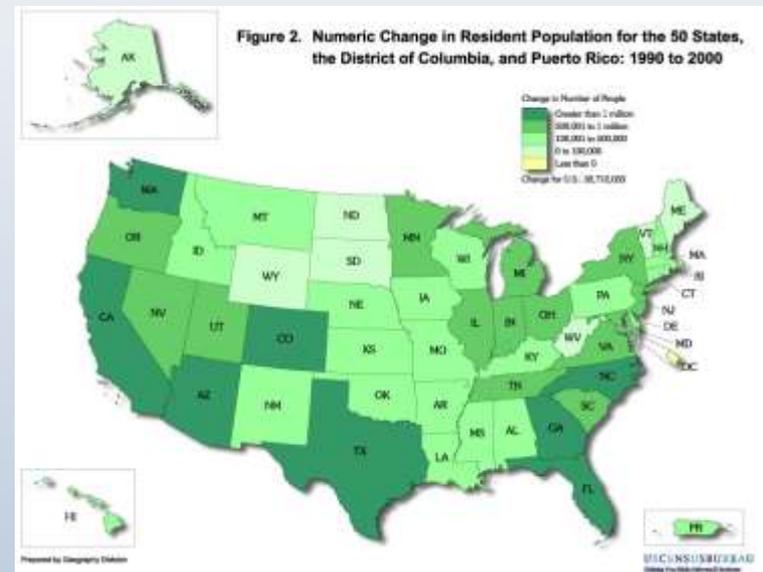
Social Scientific Data

- Examples

- Opinion polls
- Surveys, interviews
- Mass media
- Laboratory Experiments
- Field experiments
- Demographic records
- Census records
- Voting records
- Economic indicators

- Sources

- Generate own data
- Acquire from other scholars
- Data repositories: Social Surveys
- Government records
- Corporate records



About the UK Data Archive

The UK Data Archive (UKDA) is a centre of expertise in data acquisition, preservation, dissemination and promotion and is curatorial home of the largest collection of digital data in the social sciences and humanities in the UK. It is funded by the Economic and Social Research Council ([ESRC](#)), the Joint Information Systems Committee ([JISC](#)) of the Higher Education Funding Councils and the [University of Essex](#). Founded in 1967, it now houses several thousand datasets of interest to researchers in all sectors and from many different disciplines.

The UKDA is a member of the Council of European Social Science Data Archives (CESSDA), and the International Association of Social Science Information Service and Technology (IASSIST) through which it plays a lead role in international collaborative projects on issues such as data sharing, metadata and social science thesauri. UKDA is also a member institution of the national social science and historical data archive, ICPSR in Michigan and the International Federation of Data Archives (IFDO).

The UKDA provides resource discovery and support for secondary use of quantitative and qualitative data in research, teaching and learning. As a lead partner of the [Economic and Social Data Service](#) (ESDS), the UKDA is responsible for:

- overall integration and management of the ESDS
- access and preservation, focusing on the central activities of data acquisition, processing, preservation and dissemination
- [ESDS Qualidata](#), a specialist service for a range of qualitative datasets
- [ESDS Longitudinal](#), undertaken jointly with the Institute for Social and Economic Research ([ISER](#))

and supports:

- [ESDS International](#), working with Manchester Information and Associated Services ([MIMAS](#)), providing access to international micro data
- [ESDS Government](#), working with the Cathie Marsh Centre for Census and Survey Research ([CCSR](#)), facilitating access to large-scale government datasets

The UKDA also provides preservation services for other data organisations, supports the National Centre for e-Social Science ([NCeSS](#)) and facilitates international data exchange through agreements with other national archives. The UKDA hosts [AH Digital History](#), one of the five Centres of the Arts and Humanities Data Service, and the [Census Registration Service](#), facilitating access to the census data resources for UK higher and further education.

[About UKDA](#)[FAQ](#)[Contacts](#)[Associate resources](#)[About this web site](#)

Humanities data

- Examples
 - Newspapers
 - Photographs
 - Letters
 - Diaries
 - Books
 - Articles
 - Birth, death, marriage records
 - Church records
 - Court records
 - School and college yearbooks
 - Maps...

- Sources
 - Search libraries, archives, public records
 - Acquire from other scholars
 - Data repositories: Beazley, Arts & Humanities Data Service (UK)
 - Corporate records, mass media



ECAI Silk Road Atlas

Land Empires Routes Cultures

Cultures

The peoples of Eurasia encompass many language families and practice many of the world's religions. Their practices, foods, material cultures and built environments reflect the diversity of the climatic and topographic conditions that they live in. Political events tend to highlight the conflicts among cultures, classes, and religions that exist, even in a very small area. Yet practices and artifacts also move and become shared across great distances. The Mongolian stirrup changed the face of medieval warfare as much as the airplane did for the twentieth century. The concept of monotheism moved east and west from its earliest home in West Asia.

This site showcases musical cultures as an example of the dialectic between local and global experience that typified the Silk Road. Particular instruments and sounds were often very local, but bowed, drummed and other instruments, and particular sounds and genres, crossed great distances, centuries before they were aided by the modern technologies that we take for granted.

[Instruments - Interactive Java TimeMap](#)
Map showing regional influence of musical culture - Silk Road.

[Religion - Interactive Java TimeMap](#)
A worldwide mapping of religious adherence.

[Musical Instruments of the Silk Road](#)
Showcases the musical instruments of the Silk Road.

[Images from the Huntington Archive of Buddhist and Related Art](#)



<http://ecai.org/silkroad/cultures/index.html>



[Rome Reborn 2.0](#)

[Google Earth Rome Reborn](#)

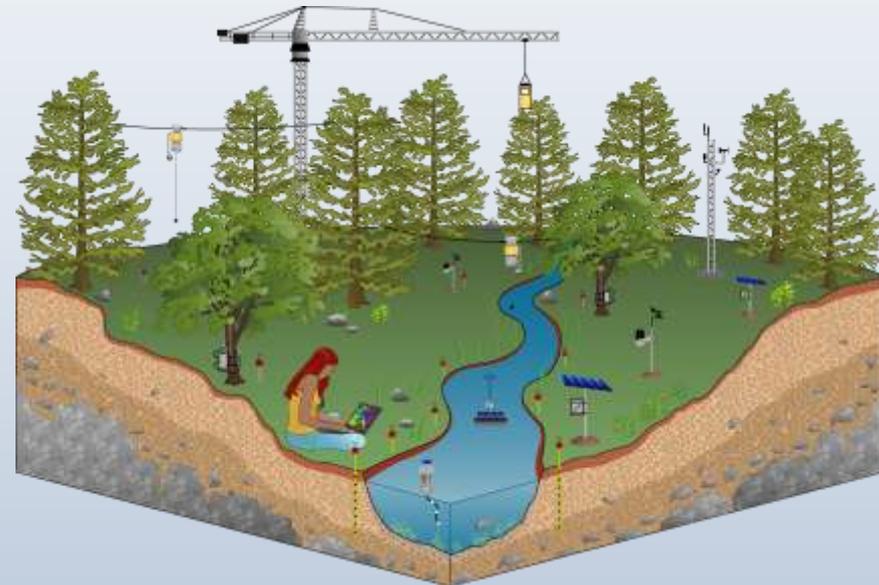
Roman Forum, Western End, ca. 400AD, copyright Regents of the University of California

How do data enter the value chain?

Function	Reported in a publication	Contributed to a data repository
<p>Legitimization Authority, quality, priority, trustworthiness</p>	<ul style="list-style-type: none"> •Peer review in context <ul style="list-style-type: none"> •Quality of method •Evidence for conclusions •Verify, reanalyze? •Author reputation 	<ul style="list-style-type: none"> •Peer review <ul style="list-style-type: none"> •Quality of metadata, documentation •“test drive” the data
<p>Dissemination Awareness, diffusion, publicity</p>	<ul style="list-style-type: none"> •Description in a publication 	<ul style="list-style-type: none"> •Repository publisher
<p>Access, preservation, curation Availability, discovery, retrieval, persistence</p>	<ul style="list-style-type: none"> •Request to author •Author maintains own data •Author directs requestor to data source 	<ul style="list-style-type: none"> •Repository <ul style="list-style-type: none"> •Immediate access •Embargo period •Curation responsibility

Center for Embedded Networked Sensing

- NSF Science & Technology Center 2002-
- Goal of CENS
 - "We envision a world where researchers, students, industry and government routinely use distributed sensor and actuator networks to understand and control both natural and artificial systems."
- CENS community
 - Some 280+ researchers across 5 research institutions
 - Technological research areas
 - Systems
 - Multi-scaled Actuated Sensing
 - Sensors
 - Statistics and Data Practice



CENS Objective & Application Areas

Seismic

- Create **programmable, distributed, multi-modal, multi-scale, multi-use observatories** to address compelling science and engineering issues

Terrestrial

- ...and reveal the previously unobservable.

- From the natural to the built environment...

- From ecosystems to human systems...

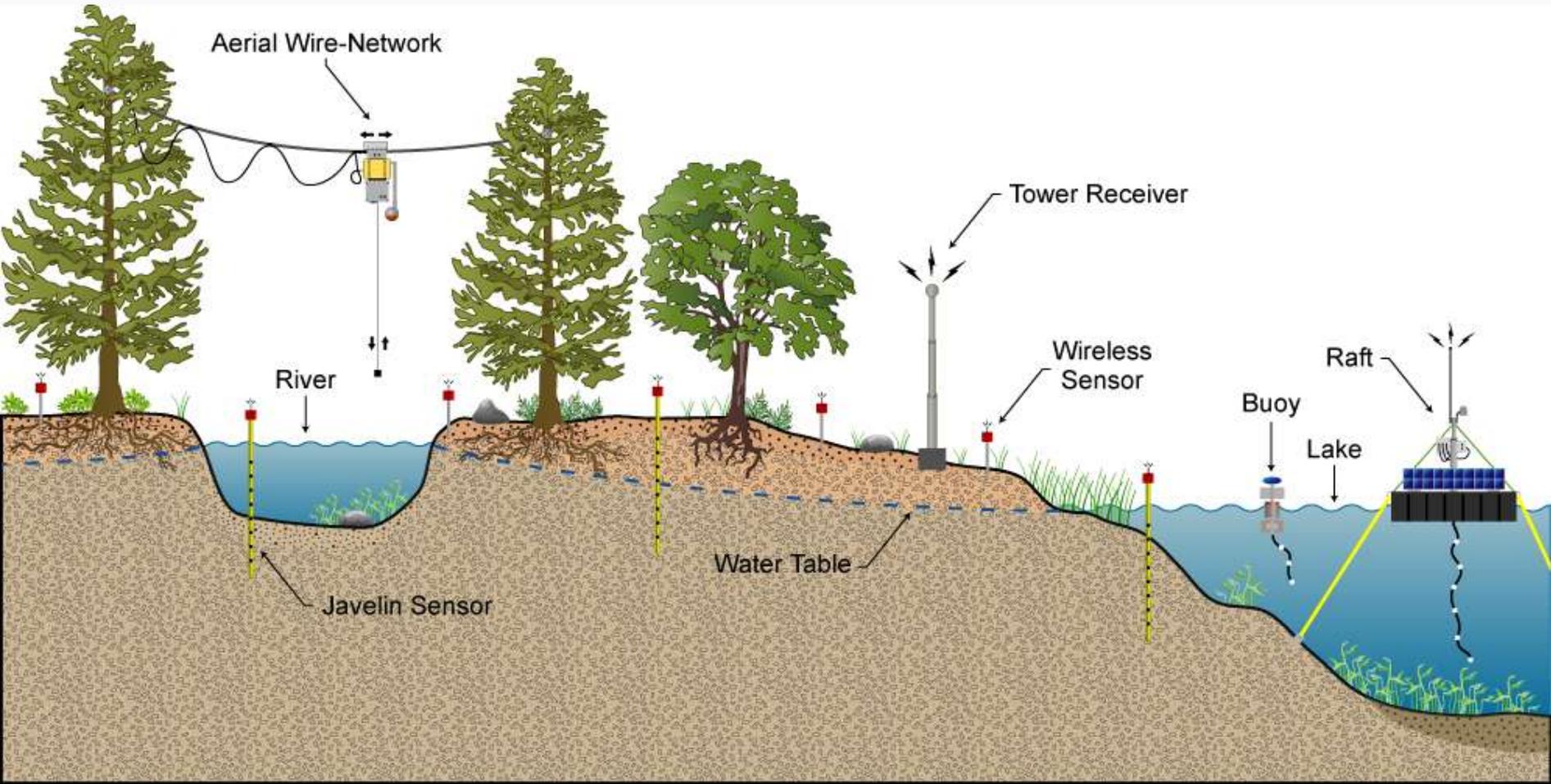
Urban

Contaminant transport

Aquatic



Field Deployment of Embedded Sensor Networks



Interview Study: Research Questions

- **Research problem:** CENS is committed to sharing data from our research
- **Research questions (selected):**
 - What are CENS data?
 - When, how, and with whom will they share data?
 - What contextual information is necessary for primary and for secondary users to interpret the data?
 - What resources exist to provide metadata?
- **Implications:**
 - What is an appropriate architecture for capturing, storing, and providing access to CENS data?
 - How can we leverage these metadata resources?

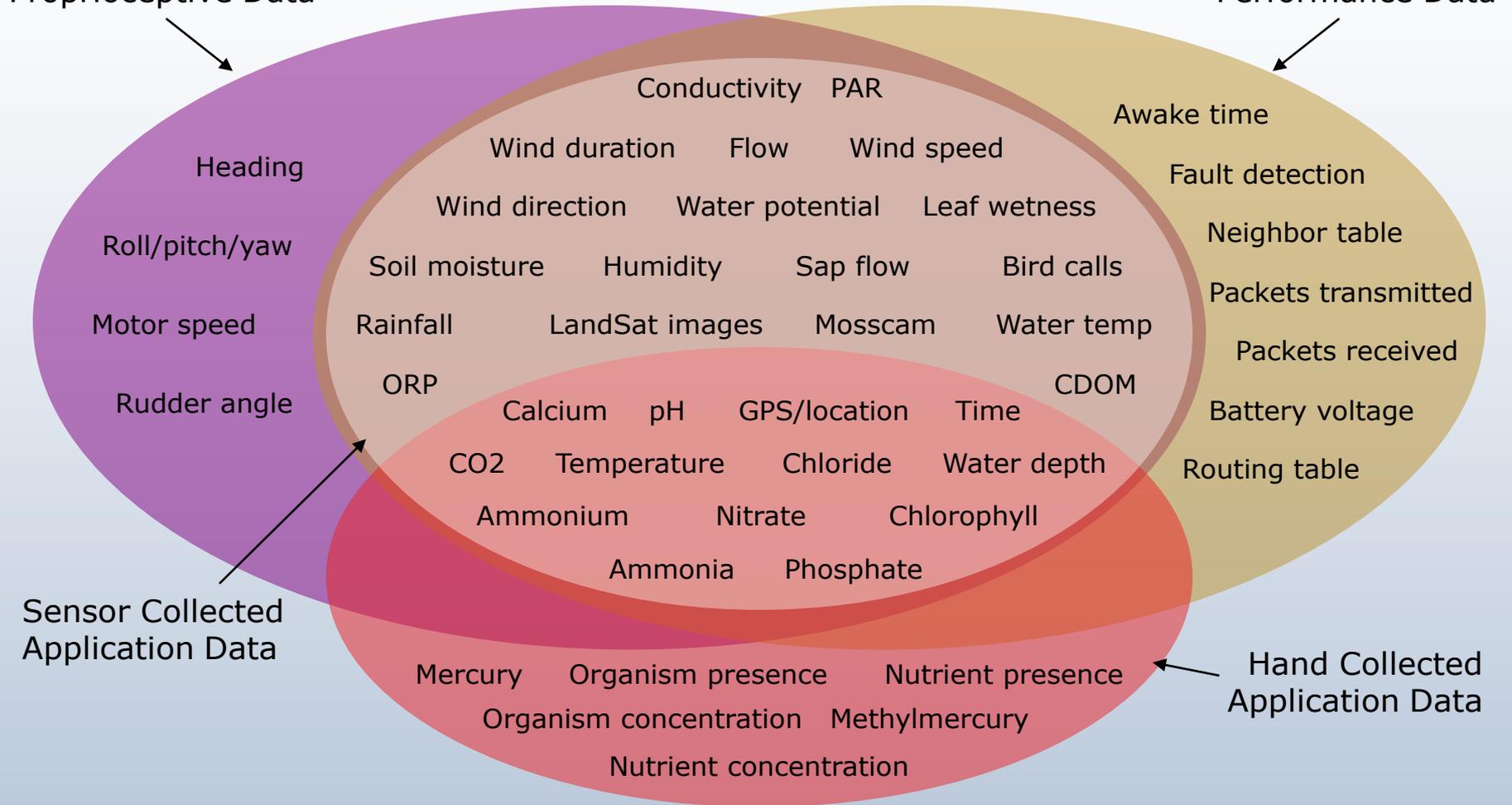
Documenting Data for Interpretation

- *"Temperature is temperature."*
- *"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.."*

What are CENS Data?

Sensor Collected
Proprioceptive Data

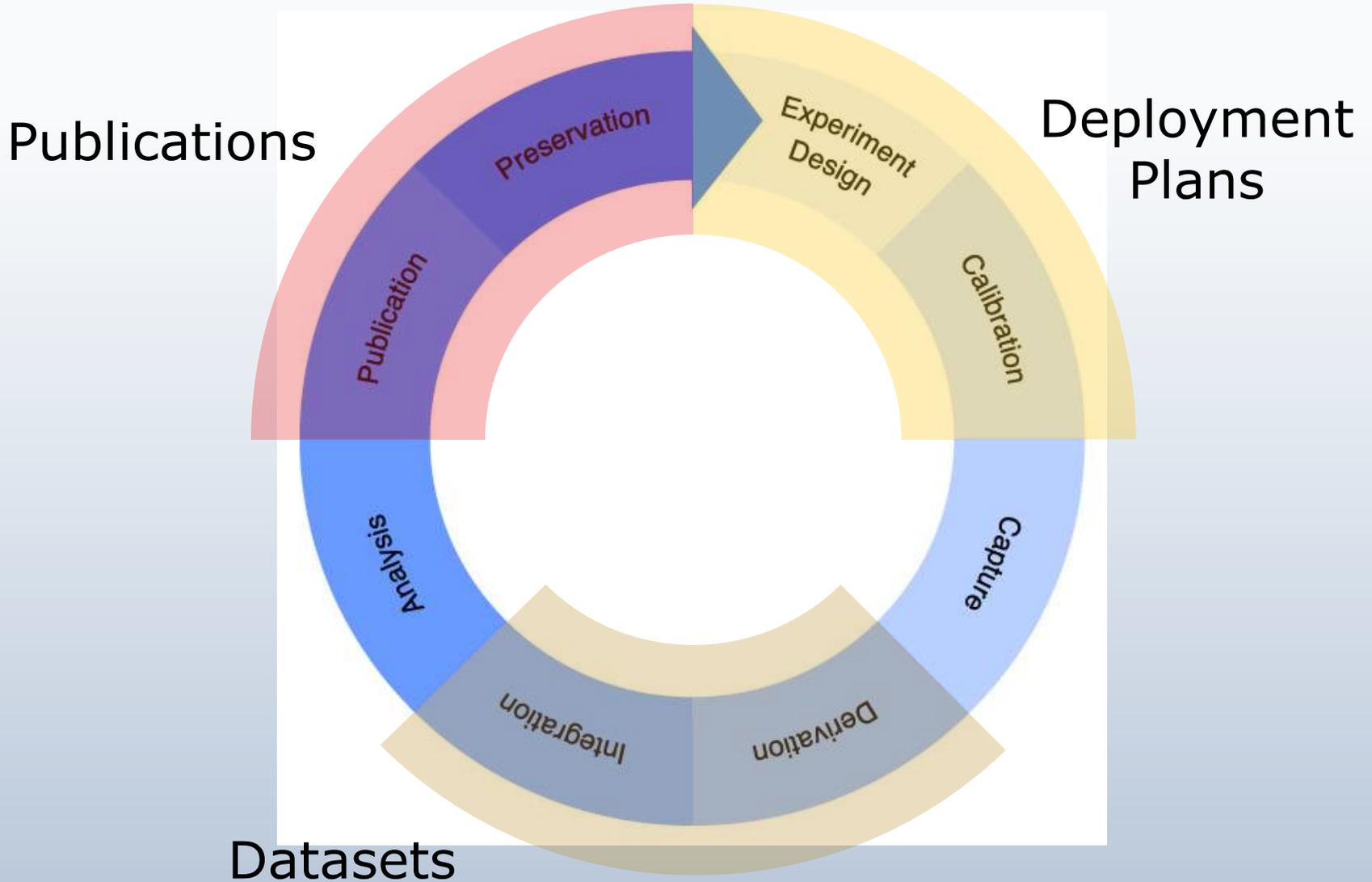
Sensor Collected
Performance Data



Sensor Collected
Application Data

Hand Collected
Application Data

Fulfilling the Data Life Cycle



Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge



*A 21st Century Agenda for the
National Science Foundation*

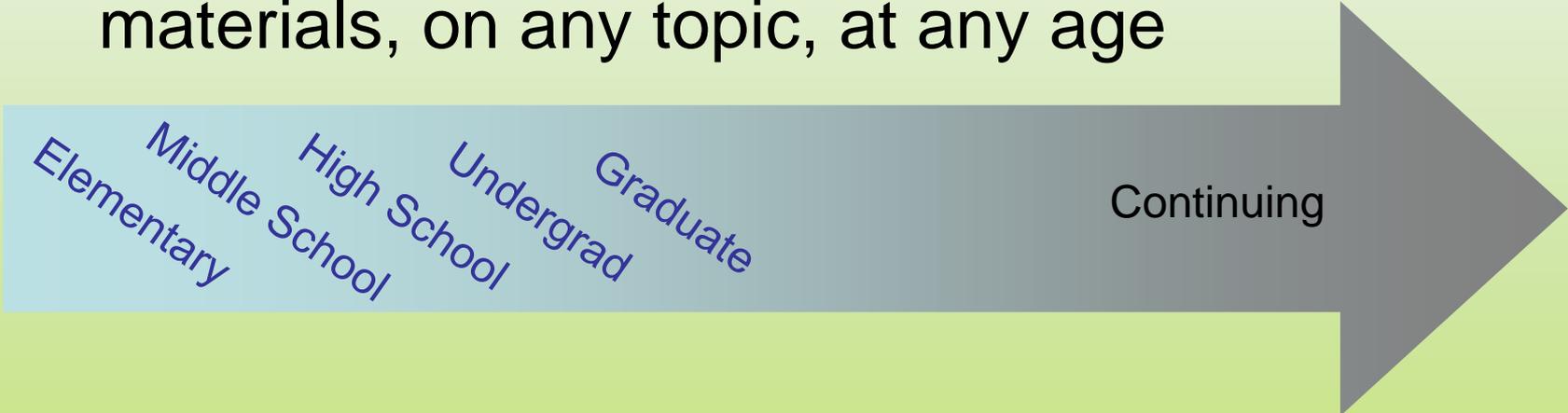
Report of the NSF Task Force on Cyberlearning
July 2008

What Is Cyberlearning?

- The use of *networked* computing and communications technologies to support learning
- Interactions among communities of learners across space and time
- Customized interaction with diverse materials, on any topic, at any age

Elementary
Middle School
High School
Undergrad
Graduate

Continuing



Why Is Cyberlearning Important?

- Leverages learning through
 - Communication technologies
 - Students' technology skills
- Extends capacity of educational institutions into life-long learning opportunities
 - Increases public understanding of science
 - Prepares citizens for complex, evolving, global challenges

War

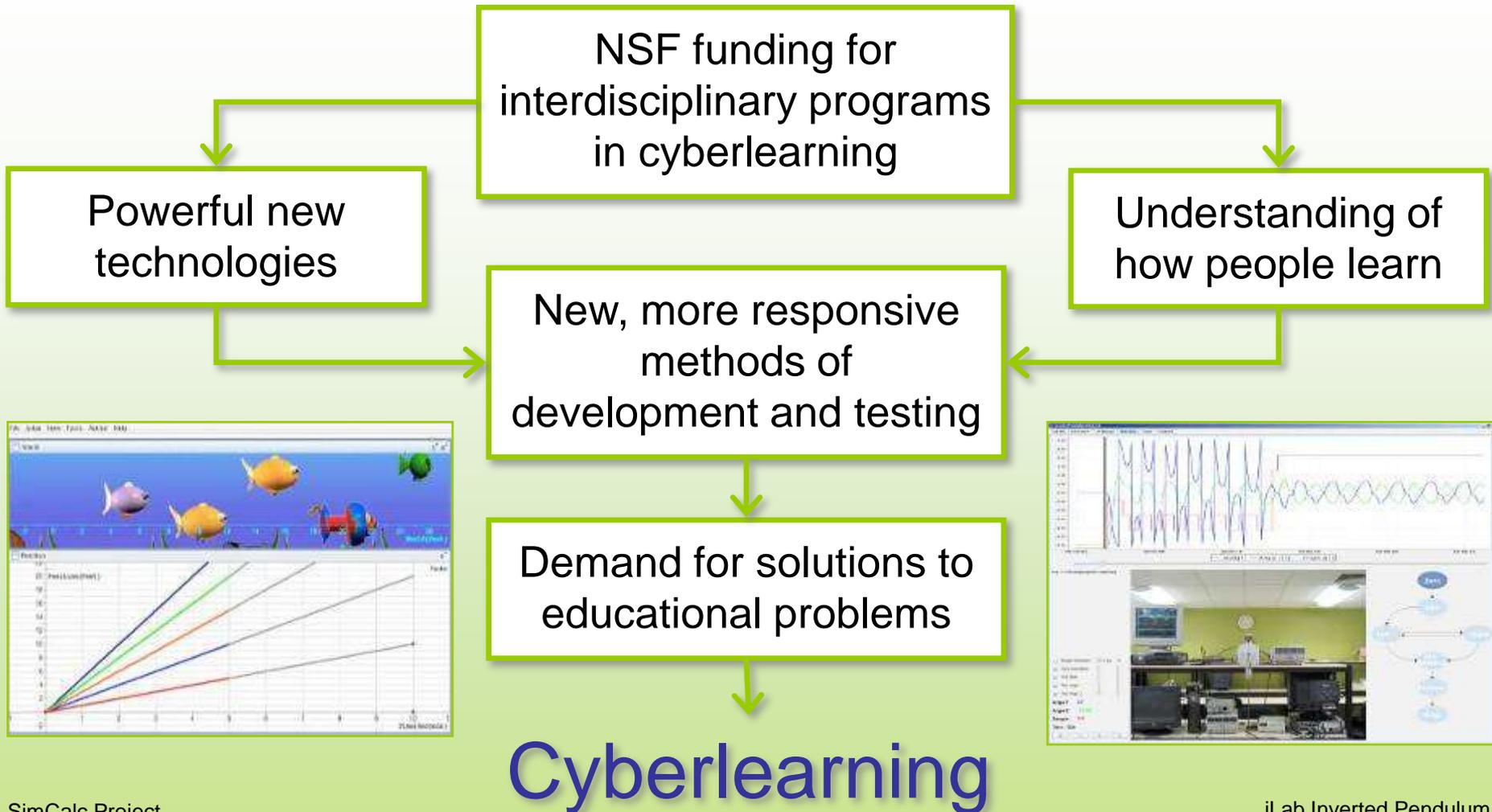
Recession

Global Warming

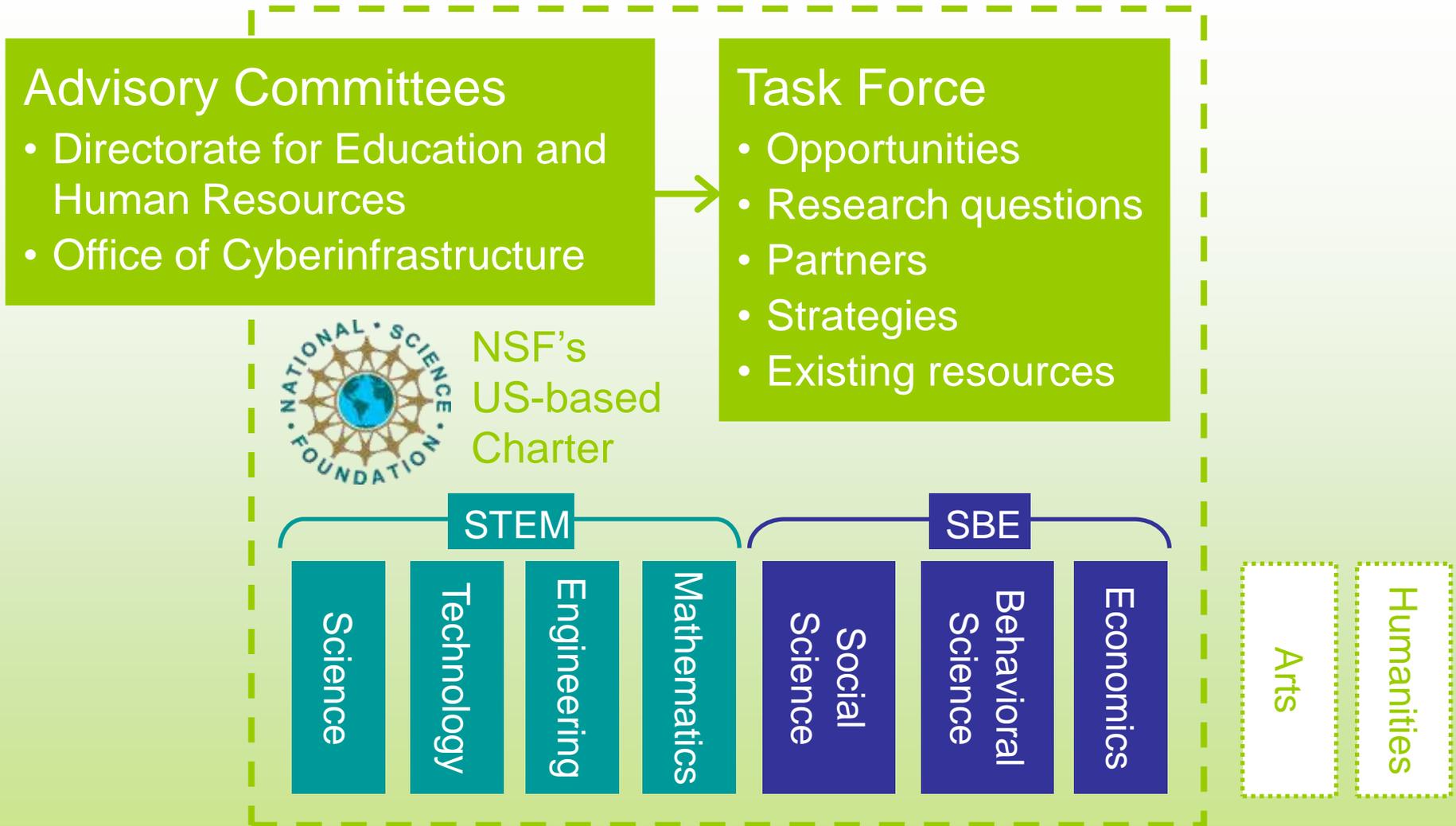
Epidemics

Poverty

Why Cyberlearning Now?



Task Force Charge



Enable Students to Use Data

- **Strategy:** Transforming STEM disciplines and K–12 education
 - New ways of looking at and understanding content
 - Preparing students for “computational thinking”
- **Opportunity:** Teaching students and teachers how to harness large amounts of data
 - Scientific research
 - Responsible use of data



Task Force Recommendations

1. Build a vibrant cyberlearning field
2. Instill a “platform perspective”
3. Emphasize the transformative power of technology
4. Promote open educational resources
5. Sustain NSF-sponsored projects

Scholarly Infrastructure for Data

- Scholars' concerns
- Public's concerns
- Librarians' concerns



Scholars' incentives to share data

- Open science
- Collaboration
- Reciprocity
- Recognition
- Coercion



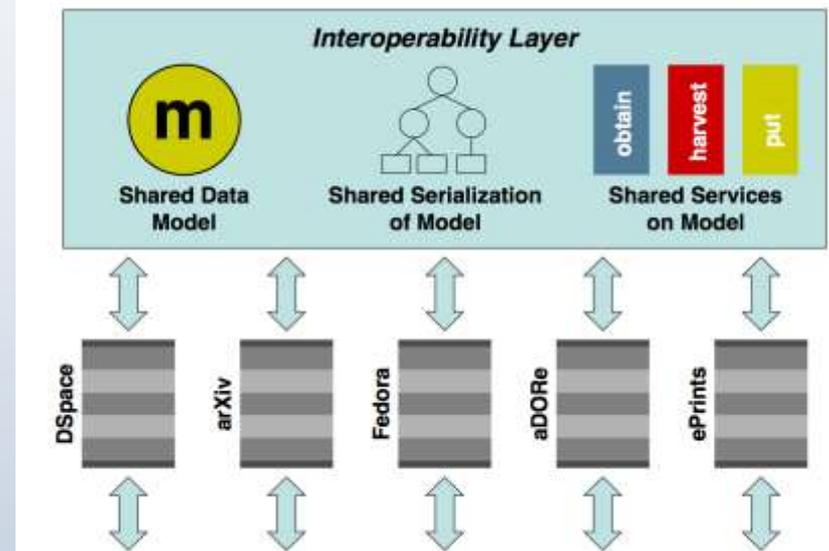
Incentives *not* to share



- Rewards for publication, not for data management
- Effort to document data
- Competition, priority of claims
- Intellectual property
 - control of own resources
 - access to resources controlled by others

Infrastructure for data: Scholars' concerns

- Agreements among research partners
 - Ownership, access, use, reuse of data
 - Release of data to others
- Agreements within disciplines
 - Syntax and semantics of data
 - Embargoes, ownership, release
- Agreements between disciplines
 - Syntax and semantics of data
 - Embargoes, ownership, release
- Technology and policy to facilitate
 - Use and reuse of data
 - Discovery and reuse of data



Infrastructure for data: Public's concerns

- How to obtain data from publicly funded research?
 - Preserved, curated
 - Made available for reuse
- How to set policy for research funding agencies?
 - Data management plans in proposals
 - Deposit of
 - Research reports
 - Datasets
- How open is open access?
 - Deposit rules?
 - Licensing?
 - Embargo rules?
 - Fees for labor to release?



Infrastructure for data: Librarians' concerns

- Scalable and sustainable infrastructure
 - Unfunded mandates
 - Short term vs. long term solutions
- Campus responsibilities
 - Expertise for data management planning
 - Responsibility for orphaned data
- Open access
 - Barriers to deposit
 - Publications vs. data
 - Open data
- Data from course management systems
 - Pedagogical content
 - Learner data
 - Open textbooks
- Policy/technology conflicts
 - Computation vs. curation
 - Library support for virtual organizations



Image: Christine L. Borgman, 1995

Acknowledgements & Thanks

- Funding for CENS: National Science Foundation Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, Principal Investigator
- Funding for CENS Data Practices Research:
 - National Science Foundation: CENS award; CENSEI grant (Education and Human Resources Directorate); SGER award (Office of Cyberinfrastructure); Human and Social Dynamics award (Social, Behavioral and Economic Sciences Directorate)
 - Microsoft Technical Computing Initiative gifts
- Graduate Student Research Assistants
 - Jillian Wallis
 - Alberto Pepe
 - Matthew Mayernik
 - Andrew Lau
 - David Fearon



Microsoft®

Christine Borgman is a member of the Scientific Advisory Board to the Microsoft Technical Computing Initiative



National Science Foundation
WHERE DISCOVERIES BEGIN