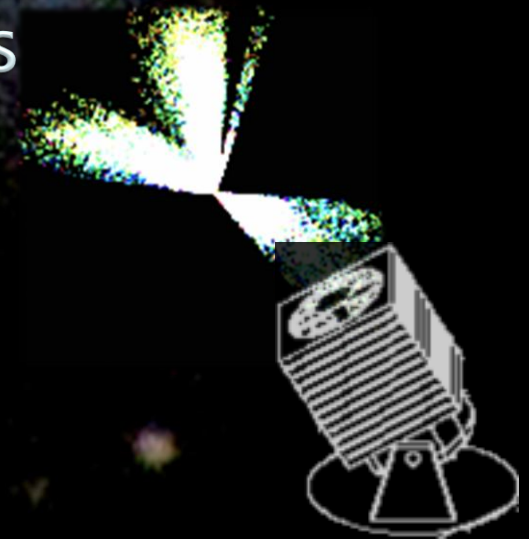


Curators to the Stars!

Berliner Bibliothekswissenschaftliches
Kolloquium

UCLA Data Conservancy Team:
Vision, Methods, Early Findings
8 June 2010

Christine Borgman,
Sharon Traweck, David
Fearon, Laura Wynolds



The Data Conservancy

Data repository

- Infrastructure for multiple sciences
- Responsive to scientists' needs

JOHNS HOPKINS
UNIVERSITY



Social science: finding contexts for data curation



User-based design

Data Requirements

CIRSS UIUC

Comparative analysis

Life
Sciences

Earth
Sciences

Social
Sciences

UCLA

Deep Case Analysis

Astronomy

Deep Case Analysis: Astronomy

Following projects and people



Sloan Digital Sky
Survey (SDSS)



Pan-STARRS



Large Synoptic
Survey Telescope

history, archives,
data practices



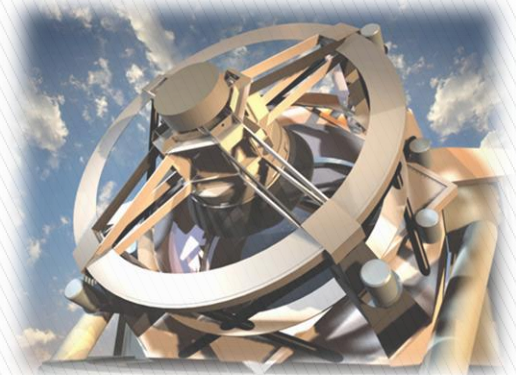
people/knowledge transfer,
development history, curation plans



Sloan Digital Sky Survey
http://www.sdss.org/photos/98_628.web1.jpg



Photo by Brett Simison
<http://pan-starrs.ifa.hawaii.edu>



Todd Mason, Mason Productions Inc. / LSST Corporation³

Following data practices & standards



Infrared Processing &
Analysis Center



“Mature” data center organization



Faculty astronomers &
small-scale projects



International Virtual
Observatory Alliance
(IVOA)



Space Telescope
Science Institute
Archives (STScI)

Curation & metadata
standards, policies

Research Questions

1. Data practices

What are the data management, curation, and sharing practices?

2. Social networks

Who uses what data when, with whom, and why?

3. Curation

What data are most important to curate, how, and for whom?

Methods

Project analysis

- History
- Documents
- People, social networks

Data practices

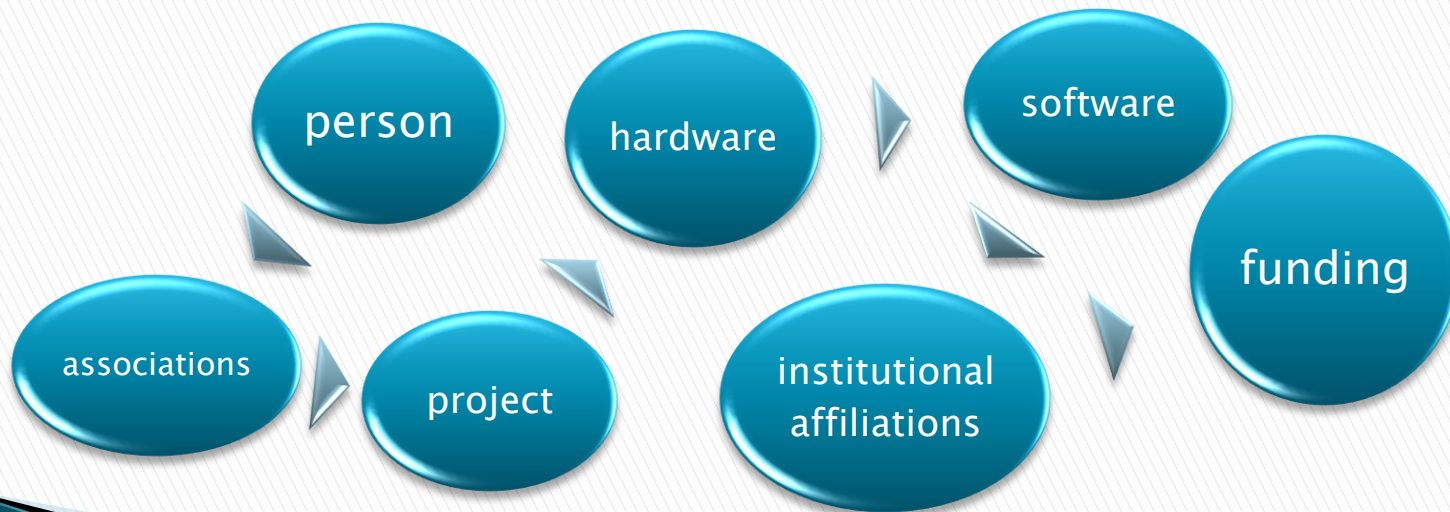
- Interviews, oral histories
- Data center observations
- Ethnography

Research repository

- Comparative analysis
- Best practices for curation
- Reflexive design and development

Data Collection & Analysis

- ▶ For data practices:
 - Open ended interview question set developed
 - Draft of code book in development
- ▶ For project and document analysis:
 - Data structure for entity mapping developed
 - Analysis of relationships over time among:

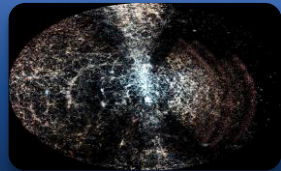


Data Collected to date

- ▶ Interviews at 3 sites
 - JHU
 - Caltech/IPAC
 - UCLA & SETI
- ▶ Documentation archive of project websites
 - SDSS
 - Pan-STARRS
 - LSST
- ▶ Archive of project publications and reports
- ▶ Bibliography of astronomy data practices

Our impressions so far...

Changes to astronomy field



Early adopters of

- Digital Data
- Large-scale data sets

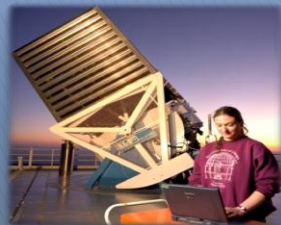


Established Data Centers



Data sharing practices

- Collaborative use of research instruments
- Data standards exist; some need for non-std
- Shared and local tools for data analysis



Evolving data management professions

- Data management roles
- Disparate relationships to data
- Data managers move with the data

Early observations

Scenarios

Organization type

Data center archives

Individual and team investigators

Outliers (SETI)

...or is each site we visit a unique scenario?

Personas

Profession

Data Wrangler

Data Gatherer

Modeler

Theorists

Career Stage

Post-doc

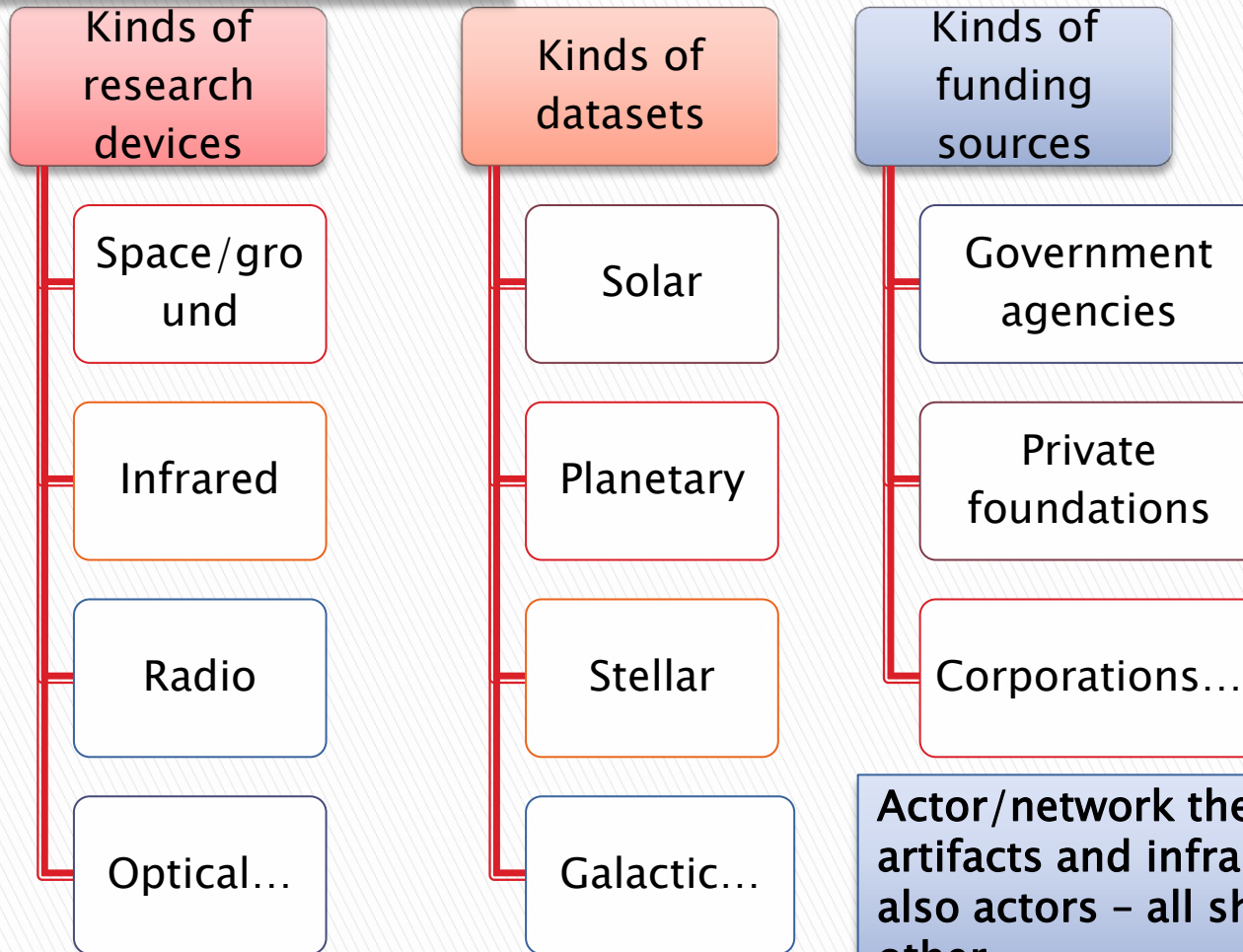
Mid-career

Team leader

...or one persona per interview?

Early observations

Other contexts



Actor/network theorists:
artifacts and infrastructures are also actors – all shape each other

Curators to the Stars!

Acknowledgments

NSF Data Conservancy Funding

Microsoft Research Partners:

Catherine van Ingen

Catherine Marshall

CENS Data Practices Team:

Sharon Traweek

David Fearon

Laura Wynolds

Alberto Pepe

Jillian Wallis

Matthew Mayernik

Katie Shilton

