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

Wynholds A Laura, University of California, Los Angeles
Jillian C Wallis, University of California, Los Angeles
Christine L Borgman, University of California, Los Angeles
Ashley Sands, University of California, Los Angeles
Sharon Traweek, University of California, Los Angeles

Available at: https://works.bepress.com/borgman/263/
Data, data use, and scientific inquiry: Two case studies of data practices
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

Science, 331(6018), February, 2011
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)
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
## Interview Questions

<table>
<thead>
<tr>
<th>Topic</th>
<th>Question</th>
<th>CENS</th>
<th>SDSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Types</td>
<td>Within your work, what is typically considered to be “data?”</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>How do you distinguish between different levels or states of data?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Sources</td>
<td>What are the main sources of data for your research projects?</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Do you routinely or have you ever used data that you did not generate yourself, or from beyond the immediate project team?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Use</td>
<td>When you look at data, what are you hoping to find in it?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>When, if ever, do you reuse your datasets?</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
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
<table>
<thead>
<tr>
<th></th>
<th>Foreground data</th>
<th>Background data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uses</strong></td>
<td>Research questions</td>
<td>Comparison, calibration</td>
</tr>
<tr>
<td><strong>Reuses</strong></td>
<td>Internal data sources</td>
<td>External data sources</td>
</tr>
<tr>
<td><strong>Disposition</strong></td>
<td>Retain, curate</td>
<td>Discard</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>Reference in paper</td>
<td>Rarely cited</td>
</tr>
<tr>
<td>Type</td>
<td>Source Named</td>
<td>Genre</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Catalog (Data) index</td>
<td>SIMBAD, VizieR</td>
<td>Obs</td>
</tr>
<tr>
<td>Curated Data Collection</td>
<td>NASA Exoplanet Database</td>
<td>Obs</td>
</tr>
<tr>
<td>Data Archive</td>
<td>Multi-mission Archive at STScI (MAST), Infrared Science Archive (IRSA)</td>
<td>Obs</td>
</tr>
<tr>
<td>Federated Data Query Services</td>
<td>Virtual Observatory Services (NVO, IVOA)</td>
<td>Obs</td>
</tr>
<tr>
<td>Ground Based Instruments</td>
<td>DEep Imaging Multi-Object Spectrograph (DEIMOS), Keck Observatories, Laser Obs Interferometer Gravitational-Wave Observatory (LIGO)</td>
<td></td>
</tr>
<tr>
<td>Physical Constants</td>
<td>NIST Atomic Spectra Database</td>
<td>Exp</td>
</tr>
<tr>
<td>Publications Index</td>
<td>SAO/NASA Astrophysics Data System</td>
<td>Mixed</td>
</tr>
<tr>
<td>Simulation</td>
<td>Millennium Simulation Database</td>
<td>Sim</td>
</tr>
<tr>
<td>Space Based Instruments</td>
<td>Chandra X-Ray Observatory, Fermi Large Area Telescope, Far Ultraviolet Obs Spectroscopic Explorer (FUSE), Galaxy Evolution Explorer (GALEX), Hubble Space Telescope, Spitzer Space Telescope, XMM X-ray Telescope</td>
<td></td>
</tr>
<tr>
<td>Space Based Sky Surveys</td>
<td>Two Micron All Sky Survey (2MASS), Infrared Astronomical Satellite Survey (IRAS), Obs Wide-field Infrared Survey Explorer (WISE)</td>
<td></td>
</tr>
</tbody>
</table>
Interdependencies

Data

Inquiry  Use
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; STraweek, 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