The Durability and Fragility of Knowledge Infrastructures: Lessons Learned from Astronomy

Christine L. Borgman, University of California, Los Angeles
Ashley E. Sands, University of California, Los Angeles
Milena S. Golshan, University of California, Los Angeles
Peter T. Darch, University of Illinois at Urbana-Champaign

Available at: https://works.bepress.com/borgman/396/
The Durability and Fragility of Knowledge Infrastructures: Lessons Learned from Astronomy

Christine L. Borgman
Distinguished Professor and Presidential Chair in Information Studies
University of California, Los Angeles
http://christineborgman.info
@scitechprof

Co-authors:
UCLA: Ashley E. Sands, Milena S. Golshan
UIUC: Peter T. Darch

ASIST Paper Presentation
Copenhagen, Denmark
October 16, 2016

https://knowledgeinfrastructures.gseis.ucla.edu
Jack Meadows, 1934-2016

The founder of the University of Leicester’s astronomy department, who had a minor planet named in his honour, has died

August 18, 2016

By Jack Grove
Twitter: @jgro_the

Knowledge Infrastructures

• What new infrastructures, divisions of labor, knowledge, and expertise are required for data-intensive science?
• How are multi-disciplinary, data-intensive scientific infrastructures established – and dismantled?
• How do data management, curation, sharing, and reuse practices vary among research areas?
Research Questions

• How has astronomy developed, deployed, and managed knowledge infrastructures for their data?
• What factors contribute to the durability and fragility of knowledge infrastructures in astronomy?

https://www.polarnopyret.com/eu/function-symbols-outerwear
http://cliparts.co/fragile
Durability and Fragility

• Durability
  – Persistence over time
  – Serves intended purposes
  – Resources invested in care and maintenance

• Fragility
  – Subject to failure or degradation
  – Uncertain investments in sustainability

https://www.polarnopyret.com/eu/function-symbols-outerwear
http://cliparts.co/fragile
Research Methods

• Document analysis
  – Public and private documents and artifacts
  – Official and unofficial versions of scientific practice

• Ethnography
  – Observing activities on site
  – Embedded for days or months at a time

• Interviews
  – Questions based on our research themes
  – Compare multiple sites over time
Astronomy

- Add cool images here to talk over
- Space based / ground based
- Wave lengths
- Celestial objects
- Pubs
Seamless Astronomy

Alberto Accomazzi, Christopher Beaumont, Douglas Burke, Raffaele D’Abrusco, Rahul Davé, Christopher Erdmann, Pepi Fabbiano, Alyssa Goodman, Edwin Henneken, Jay Luker, Gus Muench, Michael Kurtz, Max Lu, Victoria Mittelbach, Alberto Pepe, Arnold Rots, Patricia Udomprasert (Harvard-Smithsonian CfA); Mercé Crosas (Harvard Institute for Quantitative Social Science); Christine Borgman (UCLA); Jonathan Fay & Curtis Wong (Microsoft Research); Alberto Conti (Space Telescope Science Institute)
Astronomy Knowledge Infrastructures

• Observations: continuity over millennia
• Astronomy became digital: 1970s-
• Array of stakeholders: international
• Private and public funding: renewal
• Consensus mechanisms: Decadal survey
# Infrastructure: General chronology

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NASA funded infrastructure</strong></td>
<td>GSFC</td>
<td>Hubble (HST)</td>
<td>ADS</td>
<td>Chandra</td>
<td>SOFIA</td>
<td>Fermi (FGST)</td>
<td>NuSTAR</td>
</tr>
<tr>
<td>observatories</td>
<td>IPAC</td>
<td>NED</td>
<td>IRSA</td>
<td>HEASARC</td>
<td>Spitzer</td>
<td>Herschel</td>
<td></td>
</tr>
<tr>
<td>missions</td>
<td>IRAS</td>
<td></td>
<td>SATC</td>
<td>2MASS</td>
<td>Suzuki</td>
<td>Kepler</td>
<td></td>
</tr>
<tr>
<td>centers</td>
<td>STScI</td>
<td></td>
<td></td>
<td>MAST</td>
<td>Swift</td>
<td>Palomar</td>
<td></td>
</tr>
<tr>
<td>archives &amp; data tools</td>
<td></td>
<td></td>
<td></td>
<td>NExScI</td>
<td></td>
<td>(PTF)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spitzer</td>
<td></td>
<td>Planck</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WIRE</td>
<td></td>
<td>WISE</td>
<td></td>
</tr>
<tr>
<td><strong>Other infrastructure</strong></td>
<td>SAO</td>
<td>arXiv</td>
<td>Aladin</td>
<td>Keck</td>
<td>Dataverse</td>
<td>ADSASS</td>
<td></td>
</tr>
<tr>
<td>observatories</td>
<td>CfA</td>
<td>VizieR</td>
<td>NVO</td>
<td></td>
<td>GitHub</td>
<td>DDT</td>
<td></td>
</tr>
<tr>
<td>centers</td>
<td>CDS/ SIMBAD</td>
<td></td>
<td>IVOA</td>
<td></td>
<td>LSST</td>
<td>Jupyter</td>
<td></td>
</tr>
<tr>
<td>surveys</td>
<td>FITS</td>
<td></td>
<td>SDSS</td>
<td></td>
<td>Pan-STARRS</td>
<td>Notebook</td>
<td></td>
</tr>
<tr>
<td>data tools</td>
<td></td>
<td></td>
<td>Gemini</td>
<td></td>
<td>VAO</td>
<td>Zenodo</td>
<td></td>
</tr>
<tr>
<td>standards</td>
<td></td>
<td></td>
<td>Subaru</td>
<td></td>
<td>WWT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Note: This table lists key infrastructure developments over several time periods. Each period highlights significant projects and collaborations in astronomy, from observatories and missions funded by NASA to other infrastructure projects funded by various organizations. The table includes observatories, missions, centers, archives, and data tools, providing a comprehensive overview of the evolution of astronomical infrastructure.
Figure 1. Relationships between Publications, Objects, Observations and the corresponding major actors in the curating process and their activities (in red).

Data Standards and Infrastructure Fabric

VO Standards
- Standard Format: VOTable
- Access Protocols: ADQL, TAP, SIAP, SSAP

and others
ADS Collaborators

- **GSky**
- **GScholar**
- **WTW**
- **IVOA**

**Libraries**
- **ComPADRE**
- **JSTAGE**
- **arXiv**
- **SPIRES**

**Services**

**Repositories**

**Archives**
- **NED**
- **MAST**
- **MAST**
- **HEASARC**
- **Vizier**
- **Chandra**
- **SIMBAD**

**Publishers**
- **Springer**
- **IOP**
- **T&F**
- **Elsevier**
- **Wiley**
- **CrossRef**
Seamless Astronomy: ADS All Sky Survey

1. Article-object matching
2. Astrometric measurement
3. Astro-referenced images

Object Data

Literature

Optical images

Non optical images

Astrotagged literature

Data viewers

NASA archives

All-sky literature heatmap

Historical data layer

Additional database

Primary database

Image extraction

NASA

Zooniverse

Slide: Alberto Pepe
Some Durability Features

• **Data Standards**
  - Flexible Image Transport System (FITS)
  - Coordinate systems

• **Metadata and Discovery Systems**
  - Centre de Données Astronomiques de Strasbourg (CDS)
  - NASA Extragalactic Database (NED)
  - Astrophysics Data System (ADS)

• **Infrastructure Fabric**
  - Virtual Astronomy Observatory
  - International Virtual Observatory of Astronomy

---

![NED Logo](https://ned.ipac.caltech.edu/level5/Golombek/Golombek2_2.html)
Some Fragility Features

• Ground vs. Space-Based Missions
  – Ground: SDSS, LSST, Keck
  – Space: Hubble, Chandra

• Empirical vs. Theoretical Inquiry
  – Empirical: acquire and analyze observations
  – Theoretical: models, simulated data

• Sky Surveys vs. Investigator-Led Inquiry
  – Surveys: Systematic documentation of night sky
  – Investigators: Question-driven studies

https://ned.ipac.caltech.edu/level5/Golombek/Golombek2_2.html
Conclusions

• Infrastructures are fragile
• Durability is an accomplishment
• Visible infrastructure
  – Instruments
  – Institutions
• Invisible infrastructure
  – Data, metadata, provenance...
  – Information work

Telescope for the Sloan Digital Sky Survey, Apache Point, New Mexico

LSST All Hands Meeting, August 2014, Arizona State University. Arrow to Peter Darch
"THE LIBRARY CLOSED HOURS AGO AND MY PAPER... IS DUE... TOMORROW!"

"GOOD HEAVENS, BETTY! HOW COULD YOU FORGET? YOU CAN ALWAYS USE ADS!"
Acknowledgements

Christine Borgman

Peter Darch

Ashley Sands

Irene Pasquetto

Bernie Randles

Milena Golshan
Big Data, Little Data, No Data: Scholarship in the Networked World

• Part I: Data and Scholarship
  – Ch 1: Provocations
  – Ch 2: What Are Data?
  – Ch 3: Data Scholarship
  – Ch 4: Data Diversity

• Part II: Case Studies in Data Scholarship
  – Ch 5: Data Scholarship in the Sciences
  – Ch 6: Data Scholarship in the Social Sciences
  – Ch 7: Data Scholarship in the Humanities

• Part III: Data Policy and Practice
  – Ch 8: Releasing, Sharing, and Reusing Data
  – Ch 9: Credit, Attribution, and Discovery
  – Ch 10: What to Keep and Why