国際シンポジウム「大学における研究データ管理の意義と支援人材育成」

Darch, Peter
イリノイ大学アーバナ・シャンペーン校情報学部 ： 助教

Downie, J. Stephen
イリノイ大学アーバナ・シャンペーン校情報学部 ： 教授

冨浦, 洋一
九州大学データ駆動イノベーション推進本部研究データ管理支援部門 ： 部門長

石田, 栄美
九州大学データ駆動イノベーション推進本部研究データ管理支援部門 ： 教授

Building Research Data Infrastructure in Libraries: the Impact of Culture and Values

Professor Peter Darch
University of Illinois at Urbana-Champaign
ptdarch@illinois.edu

Kyushu University, June 19th 2023

School of Information Sciences

The Power of Information

Data sharing policies

Slide: Christine Borgman

- Japan Science and Technology Agency
- U.S. Federal research policy
- European Union
- Research Councils of the UK
- Australian Research Council
- And others

School of Information Sciences

The Power of Information

Reasons to share datasets

- Enhance research integrity
    - Reproducibility
    - Transparency
- Facilitate new science
    - Answer new questions
    - Integrate with other datasets

School of Information Sciences

The Power of Information

Who might reuse datasets?

Slide: Christine Borgman

- The researcher who produced the data
- Other members of the research team
- Other researchers
    - In the same discipline
    - In other disciplines
- Others outside the academy

School of Information Sciences

The Power of Information

Aspiration (FAIR) vs. reality

- **F**indable, **A**ccessible, **I**nteroperable, **R**eusable
- Reality:
  - Lack of planning
  - Lack documentation (records, standards)
  - Lack of understanding of data reusers' needs
  - Lack of support
  - Lack of infrastructure
- Challenges start early on in research cycle

School of Information Sciences

The Power of Information

Knowledge Infrastructures (KI) project

Slide: Christine Borgman

- What new infrastructures, divisions of labor, knowledge, and expertise are required for data-intensive science?

- How are the infrastructures of multi-disciplinary, data-intensive scientific endeavors established and how are they dismantled?

- How do data management, curation, sharing, and reuse practices vary among research areas?

ALFRED P. SLOAN FOUNDATION

School of Information Sciences
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

UCLA DEPARTMENT OF INFORMATION STUDIES

School of Information Sciences

The Power of Information

Knowledge Infrastructures (KI) project

Slide: Christine Borgman

Christine Borgman

Peter Darch

Ashley Sands

Morgan Wofford

Irene Pasquetto

Bernie Randles

Milena Golshan

ALFRED P. SLOAN FOUNDATION

School of Information Sciences
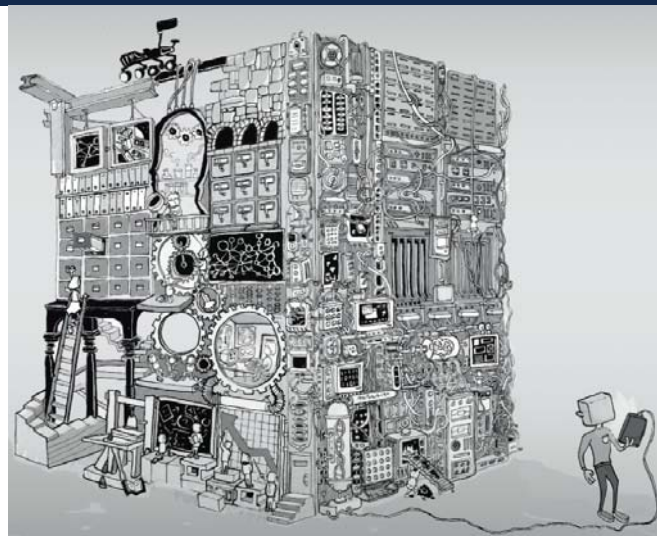UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

UCLA DEPARTMENT OF INFORMATION STUDIES

School of Information Sciences

The Power of Information

Infrastructures are complex

https://knowledgeinfrastructures.gseis.ucla.edu/

School of Information Sciences

The Power of Information

Research methods

Slide: Christine Borgman

- Ethnography
  - Observing on site (offices, laboratories, meetings) and online
  - Embedded for weeks or months at a time
- Interviews
  - Questions based on research themes
  - Compare multiple sites over time
- Document analysis
  - Public and private documents and artifacts
  - Official and unofficial versions of scientific practice

School of Information Sciences

The Power of Information

KI research sites

Slide: Christine Borgman

Domain	Focus	Topic
Astronomy sky surveys	Place: sky and universe	Survey of night sky
Deep subseafloor biosphere	Place: under ocean floor	Microbial life and environment
Biomedical collaboration	Problem: data sharing and reuse in an interdisciplinary context	Genomics of four model organisms
Computational science	Problem: Data analysis at scale	Computing in physical and life sciences
Astronomy phenomena	Place: sky and universe	Orbits, black holes, gravity

School of Information Sciences

The Power of Information

Sloan Digital Sky Survey (SDSS-I/II)

- Ground-based telescope

- Decades-long project
  - Planning 1990s
  - Data collection phases I and II from 2000-2008
  - Subsequent projects continue today

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

School of Information Sciences

The Power of Information

Sloan Digital Sky Survey (SDSS-I/II)

- 160+ terabytes data total

- Open data

- As 2008 approached:
  - Potential loss of funding
  - Potential loss of expertise

- What to do to ensure data remained available?

School of Information Sciences

The Power of Information

Library infrastructure and data reuse

- What constitutes the infrastructure upon which digital services are built for research data curation?

- What decisions do stakeholders make when building these services?

- What factors shape these decisions?

Library infrastructure and data reuse

- *Darch, P. T., Sands, A. E., Borgman, C. L. (2020). Library Cultures of Data Curation: Adventures in Astronomy. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.24345.*

- *Darch, P. T., Sands, A. E., Borgman, C. L. (2021). Do the Stars Align?: Stakeholders and Strategies in Libraries' Curation of an Astronomy Dataset. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.24392.*

SDSS Data Transfer Process

- SDSS stakeholders partnered with two university libraries
- SDSS stakeholders' goal was a dataset available:
  - Over the long-term
  - To astronomers
  - To produce new science and to reproduce existing science

(Interviews and observation primarily carried out by Dr Ashley Sands)

SDSS Data Transfer Process

- The two libraries diverged in terms of:
  - The materials they curated
  - The services they implemented to support access and use
  - **The reuse possibilities afforded by their activities**
- Institutional factors account for divergence

SDSS Data Transfer Process

	Blue University Library	Red University Library
Data archive server (DAS)	X	X
Catalog archive server (CAS)	X	X
Administrative archive	X	-
Help desk	X	-
Raw data	-	X
Software	-	X

School of Information Sciences

The Power of Information

	Blue University Library	Red University Library
Data archive server (DAS)	X	X
Catalog archive server (CAS)	X	X
Administrative archive	X	-
Help desk	X	-
Raw data	-	X
Software	-	X

School of Information Sciences

The Power of Information

Blue University: Administrative Archive

- Traditional university library
  - Concerned with visibility
  - Special Collections Unit
  - Service orientation

https://www.houseandgarden.co.uk/gallery/university-libraries

- Library suggested inclusion of Administrative Archive
- Reuse potential for historians and sociologists

School of Information Sciences

The Power of Information

Blue University: Administrative Archive

- **Strategic objectives**
  - Articulate a new purpose for existing infrastructure
- **Mission, values, and internal culture**
  - Help users understand data
  - Archive university records and preserve faculty papers
- **Organizational structure**
  - Fit into existing, rigid structure
- **Infrastructure and capacity**
  - Leverage extant Special Collections Unit infrastructure
  - Leverage extant history of science expertise

School of Information Sciences

The Power of Information

Red University: Data Archive Server

- Transfer handled by autonomous unit in Red University library
  - *Institute for Innovation in Digital Curation (IIDC)*, pseudonym
  - Largely dependent on soft money
    - Research & development orientation
    - *Guardianship of Data* (*GD*) project
  - Agile culture of software development

- Data Archive Server (DAS)
  - Included at suggestion of IIDC during negotiation phase
  - Prioritized by IIDC during implementation phase
  - Enhanced reuse potential for reproducibility

School of Information Sciences **The Power of Information**

Red University: Data Archive Server

- **Strategic objectives**
  - Reinforce work of GD project
- **Mission, values, and internal culture**
  - Boost research and development mission of the IIDC
- **Organizational structure**
  - Could be prioritized at discretion of the team (Agile structure)
- **Infrastructure and capacity**
  - Leverage staff expertise and GD infrastructure for an "easy win"

School of Information Sciences **The Power of Information**

Conclusions

- Institutional factors intersected to produce outcomes
    - Strategic objectives
    - Mission, values, and internal culture
    - Organizational structure
    - Infrastructure and capacity
- Variations in factors produced variations in terms of:
    - The scholarly purposes for which the data could be used
    - The scholarly communities who found the dataset useful
- Divergent pathways led to different prospects for data reuse

Acknowledgements

And thank you for listening