Local or global? Making sense of the data sharing imperative (Keynote)

Christine L Borgman, University of California, Los Angeles
Local or global? Making sense of the data sharing imperative

Christine L. Borgman
Professor & Presidential Chair in Information Studies
University of California, Los Angeles

Keynote Presentation
Oxford Internet Institute
A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society
Deluge!!!

Data!!
Data Sharing Policy – National Science Foundation

NSF Data Sharing Policy
• Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF Data Management Plan Requirements
• Beginning January 18, 2011, proposals submitted to NSF must include a supplementary document of no more than two pages labeled “Data Management Plan”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.
Data Sharing Policy – UK agencies

Wellcome Trust

• [It is] our expectation that all our funded researchers should maximize access to their research data with as few restrictions as possible. It requires applicants whose proposed research will generate data that hold significant value as a resource for the wider research community to submit a data management and sharing plan as part of the application process. Our policy aligns with those of other organizations - such as the MRC, BBSRC, the US National Institutes of Health and the NCRI Cancer Informatics initiative.

Economic and Social Research Council

• Our Research Data Policy is built upon the principle adopted by the OECD stating that publicly-funded research data are a public good, produced in the public interest and, therefore, should be openly available to a maximum extent possible. ESRC is, therefore, committed to long-term preservation, high quality data management and strengthening the provision for secondary data analysis.
Data

Experimental

Computational

Observational

Records

Data types: historical, DNA, lab, maps, spreadsheets, polls, molecules, books, indicators, x-rays, opinion, media, ground, models, ground surveys, ground, photographs, weather, photographs, notebooks, death, notebooks, death, death.
Rationales for Sharing Research Data

1. Reproduce or verify research

http://chemistry.curtin.edu.au/research/index.cfm

http://serc.carleton.edu/cismi/broadaccess/groupwork.html
2. Public monies serve the public good
3. Others can ask new questions


4. Data curation advances innovation
Rationales for Sharing Research Data

Motivations to Share

- R1: Reproduce/verify
- R2: Public interest
- R3: Ask new questions
- R4: Advance research

Interests Served

Data producers

Data users

Infrastructure for research data

• Social practice
• Usability
• Identity
• Persistence
• Discoverability
• Provenance
• Relationships
• Intellectual property
• Policy

http://datalib.ed.ac.uk/GRAPHICS/blue_data.gif
Conclusions

- Research policy:
  - Data are intellectual property to be managed and exploited
- Funding agencies:
  - Research data are public assets
- Scholarly communication:
  - Research data are process and product
- Publishers:
  - Data are not publications
- Libraries:
  - Research data are the new special collections
- Researchers:
  - Some data will be shared, with some people, some of the time
- Internet research:
  - What are data, to whom, when, why, and to what ends?

http://plus.maths.org/content/text-bytes-and-videotape
Acknowledgements

• National Science Foundation
  – CENS: Cooperative Agreement #CCR-0120778, D.L. Estrin, UCLA, PI.
  – CENS Education Infrastructure: #ESI-0352572, W.A. Sandoval, PI; C.L. Borgman, co-PI.
  – Towards a Virtual Organization for Data Cyberinfrastructure, #OCI-0750529, C.L. Borgman, UCLA, PI; G. Bowker, Santa Clara University, Co-PI; T. Finholt, University of Michigan, Co-PI.
  – Monitoring, Modeling & Memory: Dynamics of Data and Knowledge in Scientific Cyberinfrastructures: #0827322, P.N. Edwards, UM, PI; Co-PIs C.L. Borgman, UCLA; G. Bowker, SCU; T. Finholt, UM; S. Jackson, UM; D. Ribes, Georgetown; S.L. Star, SCU)
  – Data Conservancy: OCI0830976, Sayeed Choudhury, PI, Johns Hopkins University.

• Microsoft External Research: Tony Hey, Lee Dirks, Catherine van Ingen, Catherine Marshall

• Conundrum paper comments: CENS Data Practices team at UCLA – David Fearon, Matthew Mayernik, Katie Shilton, Jillian Wallis, and Laura Wynholds; Paul Uhlir of the National Academies.