June 9, 2015

Data, data everywhere — but how to manage and govern?

Christine L. Borgman, University of California, Los Angeles

Available at: https://works.bepress.com/borgman/366/
Data, Data, Everywhere – How to Manage and Govern?

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

Berkman Center for Internet & Society
Harvard University
June 9, 2015
Day after day, day after day,
We stuck, nor breath nor motion;
As idle as a painted ship
Upon a painted ocean.

Water, water, every where,
And all the boards did shrink;
Water, water, every where,
Nor any drop to drink.

*The Rime of the Ancient Mariner*

**Samuel Taylor Coleridge, 1798**
Data Management and Governance

• Data
• Data collected *by* our community
• Data collected *about* our community
• UCLA Data Governance Task Force
Big Data

Volume

Data Size

Data Complexity

Speed of Change

Data Sources

Variety

Long tail of data

Volume of data

Number of researchers

Slide: The Institute for Empowering Long Tail Research
Open Data: Free

• A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike

Open Data Commons. (2013).
Open Data: Useful

• Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, and sustainability.

Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

Precondition:

Researchers share data
Data collected *by* our community

- **Data types**
  - Research data
  - University analytics for teaching and learning

- **Policy and management responses**
  - Mandates of funders and journals
  - Research data management services
  - Release and retention practices
  - Laws and policies
    - Human subjects regulations
    - Open records laws
    - HIPAA, FERPA, PII...
Open access policies

• Australian Research Council
  – Code for the Responsible Conduct of Research
  – Data management plans
• National Science Foundation
  – Data sharing requirements
  – Data management plans
• U.S. Federal policy
  – Open access to publications
  – Open access to data
• European Union
  – European Open Data Challenge
  – OpenAIRE
• Research Councils of the UK
  – Open access publishing
  – Provisions for access to data
Dataverse

Harvard Dataverse
A collaboration with Harvard Library, Harvard University IT, and IQSS

Search this dataverse...

Dataverses (1,080)
Datasets (58,142)
Files (273,489)

Dataverse Category
Organization or Institution (94)
Research Project (72)
Researcher (51)

Replication Data for: Crowdsourcing Accountability in a Nigerian Election
Jun 8, 2015 - Catie Snow Bailard Dataverse

Bailard, Catie Snow, 2015, "Replication Data for: Crowdsourcing Accountability in a Nigerian Election",
http://dx.doi.org/10.7910/DVN/HPGNEA, Harvard Dataverse, V1 [UNF:6:P5c93FSwnO0dZaL+8kPSsA==]

Replication data for Crowdsourcing Accountability in a Nigerian Election
Data collected about our community

• Student records
  – Registrar
  – Course management systems
  – ID card based services: library, dorms, food, health...
  – Internet services: email, social media, music, ...

• Faculty records
  – Publications
  – Grants
  – Teaching evaluations
  – Service activities
  – Financial, medical
  – Internet services

Gustave Dore, Ancient Mariner Illustration, 1798
On 1-4 June, 2014, a group of educators, scientists, and legal/ethical scholars assembled at the Asilomar Conference Grounds in Pacific Grove, California. Their task was to develop a framework to inform decisions about appropriate use of data and technology in learning research for higher education. A modified Chatham House Rule guided their deliberations, which produced the convention presented here.

This convention reflects general principles rather than the views of individual participants.

The Asilomar Convention for Learning Research in Higher Education

Individuals, nations, and international agencies of all kinds increasingly rely on the promise of education to improve the human condition. Contemporary technology has created unprecedented opportunities to create radical improvements in learning and educational achievement, but also conditions under which information about learners is collected continuously and often invisibly. For these reasons, collection and aggregation of evidence to pursue learning research must proceed in ways that respect the privacy, dignity, and discretion of learners.

Virtually all modern societies have strong traditions for protecting individuals in their interactions with large organizations, especially for purposes of scientific research, yet digital media present problems for the inheritors of those traditions. Norms of individual consent, privacy, and autonomy, for example, must be more vigilantly protected as the environments in which their holders reside are transformed by technology. Because the risks associated with data exposure are growing
Student Privacy Bill of Rights

News

In a March 2014 Washington Post article, EPIC unveiled the Student Privacy Bill of Rights, an enforceable student privacy and data security framework.

In line with the President’s Consumer Privacy Bill of Rights, which is based largely based on the well-established Fair Information Practices (FIPs), schools, districts, and EdTech and other cloud-based service providers should adhere to the following practices when collecting student data. These rights should transfer from parents or legal guardians to students once the student is eighteen or attending college.

1. Access and Amendment: Students have the right to access and amend their erroneous, misleading, or otherwise inappropriate records, regardless of who collects or maintains the information.

   - There are gaps in current laws and proposed frameworks concerning students' access and amendment to their data. Schools, companies, government agencies, and other entities that collect any student information should provide student access to this information. This includes access to any automated decision-making rule-based systems (i.e., personalized learning algorithms) and behavioral information.
Bibliometrics, Scientometrics, Informetrics, Webometrics...


---


## Indicators from Publication Data

Searches for author: Christine Borgman, Christine L. Borgman, CL Borgman (excluding other C Borgman authors) on July 28, 2014

<table>
<thead>
<tr>
<th>Source</th>
<th>Publications</th>
<th>Citations received</th>
<th>H-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar (Google)</td>
<td>380</td>
<td>7766</td>
<td>39</td>
</tr>
<tr>
<td>Web of Science (Thomson-Reuters)</td>
<td>145</td>
<td>1629</td>
<td>20</td>
</tr>
<tr>
<td>Scopus (Elsevier)</td>
<td>77</td>
<td>1314</td>
<td>14 (after 1995)</td>
</tr>
</tbody>
</table>
Mapping Scholarship

Ten Simple Rules for the Care and Feeding of Scientific Data


Published: April 24, 2014 • DOI: 10.1371/journal.pcbi.1003542 • Featured in PLOS Collections

Introduction

Rule 1. Love Your Data, and Help Others Love It, Too

Rule 2. Share Your Data Online, with a Permanent Identifier

Rule 3. Conduct Science with a Particular Level of Rigor

Figures

Published April 24, 2014;
Altmetrics data as of June 9, 2015
• How should UCLA collect, organize, and use research analytics about our community?
• Who should have access to these data?
  – Within UCLA?
  – In partnership with public and private entities?
• What are the governance principles?
• What are the governance processes?
Privacy and Information Security

Information security protects all information and infrastructure

- Individuals (e.g., web sites visited, research being conducted and related data)
  - Information about individuals (e.g., student or patient records; or SSNs)
    - Confidential information (e.g., intellectual property, security info)

Autonomy privacy: ability of individuals to conduct activities without observation

Information privacy: protects information about individuals

Information:

Infrastructure (e.g., computers and networks)

University of California Privacy and Information Security Committee
Proposed Data Governance Approach

Scope

• Data held by the campus about its students, faculty, and staff that includes personal information – even if not PII as defined by existing law or policy

• Where competing privacy interests, goals, University values, or obligations in the application or use of this data exist and for which no statutory provision, common law, or University policy is directly applicable (esp. IRB)

Goals

• Resolve legitimate disagreements and provide a path forward

• Promote transparency

• Promote open discussion

• (Leverage existing structures and expertise)

Data Governance Task Force Site: https://ccle.ucla.edu/course/view/datagov
Triggers for review

• When data are used to make decisions about people
• When data are collected about people without their knowledge or consent
• When data about people are used in unexpected ways without subjects’ knowledge or consent
  – New applications of data or systems
  – Mining, analysis, and aggregation
• When data are shared with external entities
  – Private sector partners
  – Public sector partners
  – Other universities

Data Governance Task Force Site:
https://ccle.ucla.edu/course/view/datagov
Board on Privacy and Data Protection

Executive Vice Chancellor and Provost*

Voting members
- Faculty Chair – Appointed by EVC + Senate
- Administrative Vice Chair – Vice Provost, IT
- 6 faculty members
- 6 administrative members
- 1 undergraduate student representative
- 1 graduate student representative

Non-voting members
- UCLA Chief Privacy Officer
- Chief Information Security Officer
- Designee of the EVC and Provost
- Designee from Audit & Advisory Services

* decision-making authority
Executive Vice Chancellor and Provost*  

Oversight Committee on Audit, IT Governance, Compliance and Accountability*  

Board on Privacy and Data Protection  

IT Planning Board  

Academic Senate*
Board on Privacy and Data Protection

UCLA Chief Privacy Officer

- Training and awareness
- Governance support
- Privacy breach analysis
- Policy development and interpretation
- Data use questions
- UC privacy and information security report recommendations implementation
The Office of the UCLA CPO becomes the triage point for incoming requests.

* Board on Privacy and Data Protection

* UCLA Chief Privacy Officer

* Institutional Review Board

* decision-making authority
Discussion questions

• Problem: data or uses of data not covered by existing laws or policies (e.g., FERPA, HIPAA, PII)

• How to scope the data governance problem?
  – By subjects of data collection?
  – By uses of data?
  – By agency collecting the data?
  – By partners involved?

• What are appropriate criteria, values, practices?

• What are workable governance processes?
Acknowledgements

• Kent Wada, UCLA Chief Privacy Officer and Chief Information Security Officer
• James F. Davis, UCLA Associate Vice Provost for Information Technology
• UCLA Privacy and Data Protection Board
• UC Initiative on Privacy and Information Security

Data Governance Task Force Site:
https://ccle.ucla.edu/course/view/datagov