

# **UCLA**

## **Presentations**

### **Title**

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

### **Permalink**

<https://escholarship.org/uc/item/809146r2>

### **Author**

Borgman, Christine L.

### **Publication Date**

2015-06-09

### **Copyright Information**

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

# Data, Data, Everywhere – How to Manage and Govern?

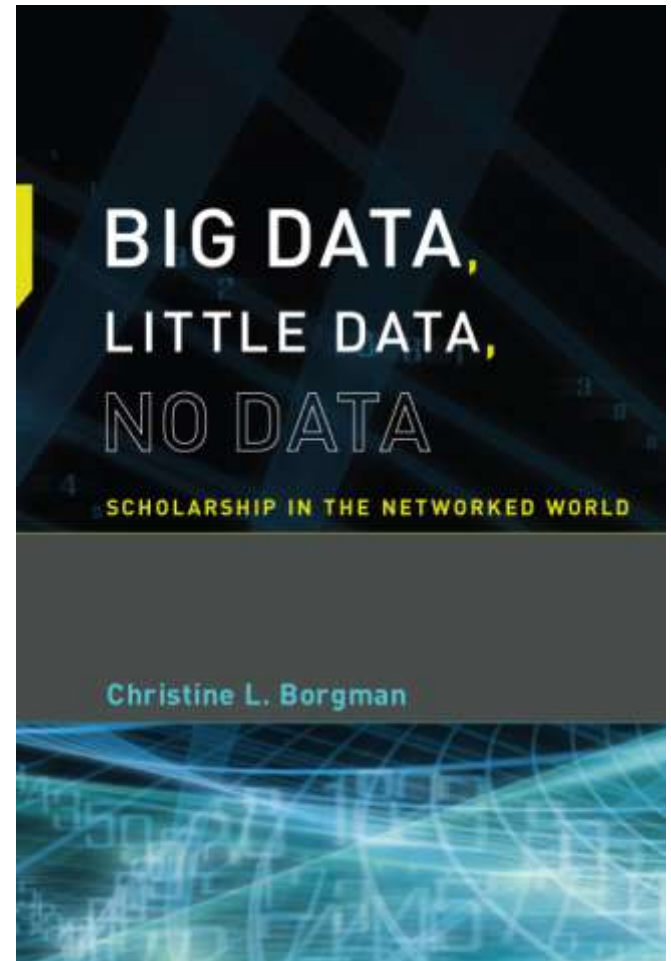
Christine L. Borgman

Professor and Presidential Chair in Information Studies  
University of California, Los Angeles

@scitechprof

Berkman Center for Internet & Society  
Harvard University

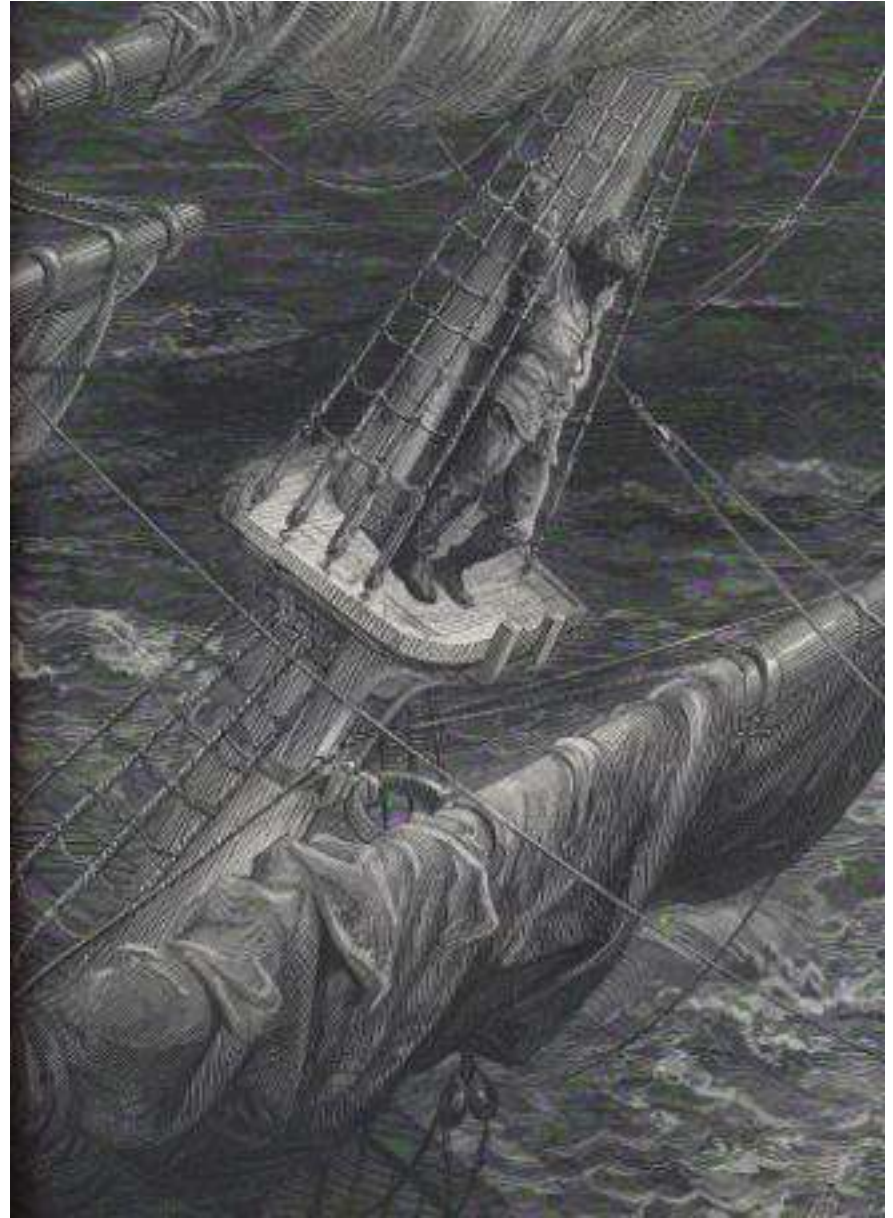
June 9, 2015



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.

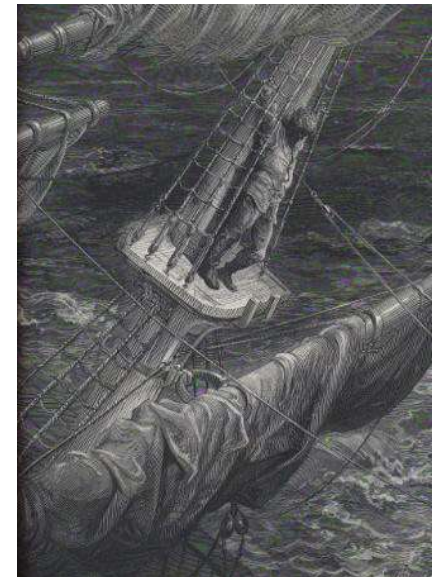
*The Rime of the Ancient Mariner*  
[Samuel Taylor Coleridge, 1798](#)



Gustave Dore, Ancient Mariner Illustration, 1798

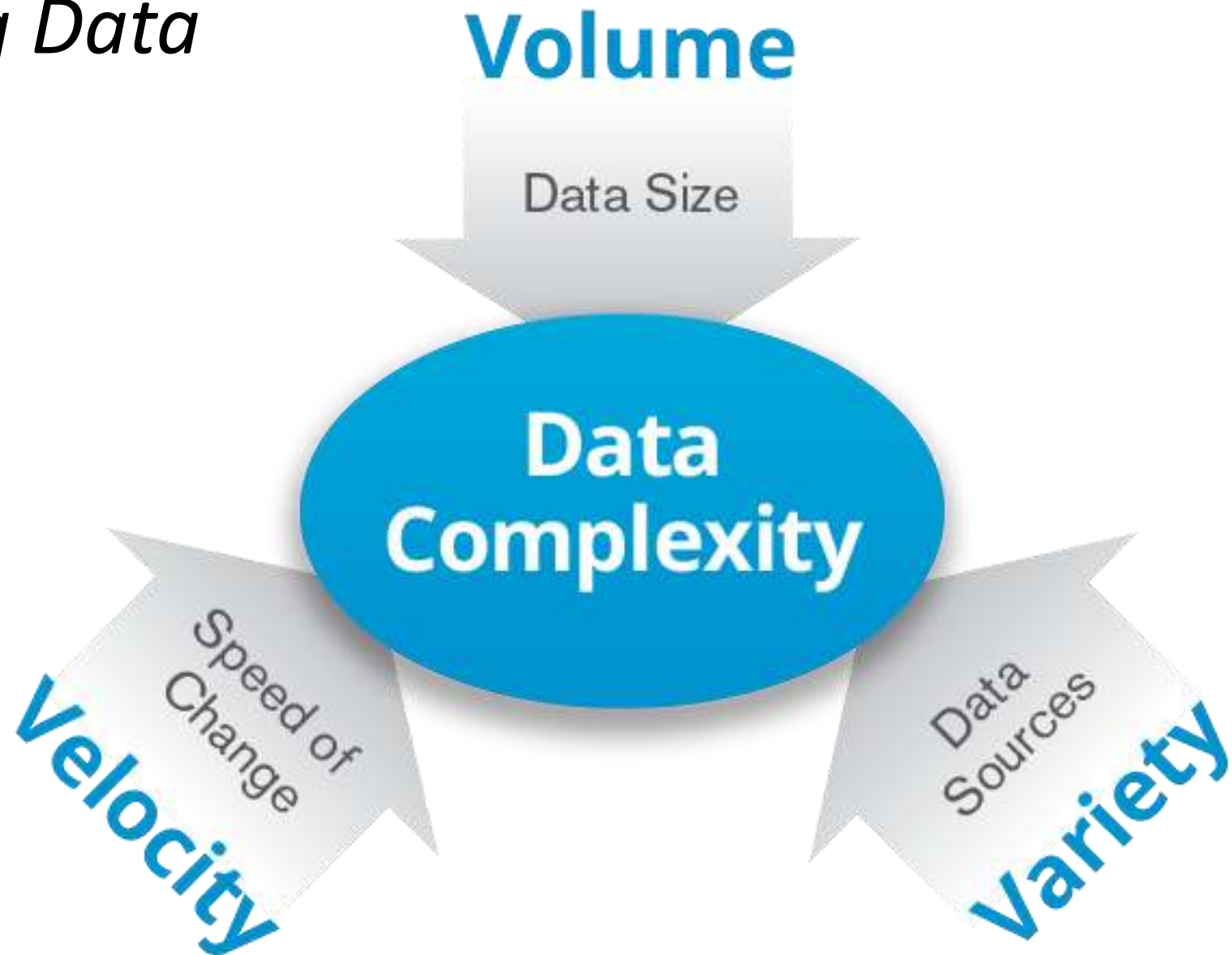
# Data Management and Governance

- Data
- Data collected *by* our community
- Data collected *about* our community
- UCLA Data Governance Task Force

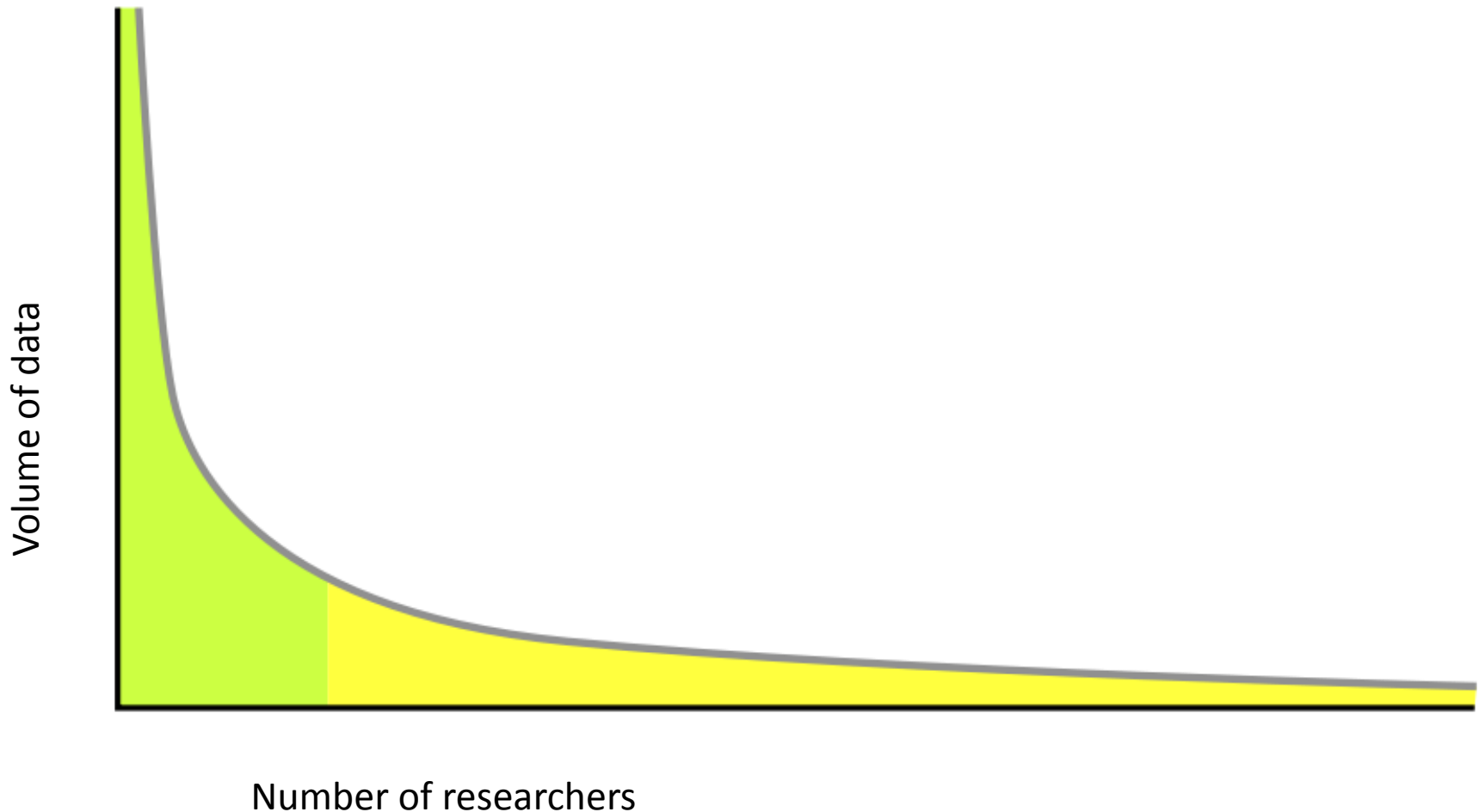


Gustave Dore, Ancient Mariner Illustration, 1798

# *Big Data*



# Long tail of data



# 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).



State Library and Archives of Florida, 1922. Flickr commons photo



```
/FontMatrix matrix def
/FontBBox[2048 -1164 1 index div -628 2 index div 4096 3 index div 2062 5 -1 rol
/sfnts [<
7472756500090000000000000637674200000000000000009C000007DA6670676D0000000000000878
019901AC01C101C501C901E101F601F601F60222022202280236023F024302460267028502850294
01160125011800EA00EA00AE0000003E05BB008A04D70053003FFF8CFFD500150028002200990062
B200402F2B59B002602D2C21B0C051580C6423648BB81555621BB200802F2B59B002602D2C0C6423
B8FF80B30809341CB8FF8040E80809343609352F4A0959045809A7090626062B082C0B280C281328
D4401E091202550C0C047F180118401112025518401D1D02551810251E073C2C04002FED3FEDC42B
170003020913141500215902071B25090B1E0505262500180C0D0D0655180210100655180CB8FFF8
1F1E4A62A1A4FB432E7900020044FFE40405058E001F002D024BB1020243545840262F4010100255
2F2B2BDD2B2BC01112392F2BCDD0CD003FED5DC45D5D2B323FED12392F5D5D5DCD31301B4019125F
11242517012B50100110302A2912110608070A101B011BB80152B38F2D012DB802F6B2012A2BB801
025D5972103C103C1112173911123939111239011112393912393931304379407A4B573A4524351C
2D602D702D04802D902D02B02D01002D102DC02DD02D042D6037A67F182B10F65D5D71723CFD3C10
FFEAB40C0C02550CB8FFE2B40D0D02550CB8FFD6B4101002550CB8FFDEB50F0F02550C05BD03E200
2F2F2F3FDDCD3F3F10ED10ED313001B00D4B5458BF0046FFE80045FFE8002FFFE80030FFE8B51A18
1010065529B8FFF2B70F0F0655292935341112392F2B2B2BDD2B2B2BC01112392F2B2B2BCD2B2B2B
3236353427260200D07E6B76CF7FCF7A677DCC53356B429F82617E694703AF9E87AF7BFC80A58BAD
961A9C1E9621982A9E2BA816A61AAB1CAD2BB916BE2BCD2BDA2BEC2BFB2B2A202D732573288F1397
201A01601A701A021A120B003FC45D5DED5D5D2F3FC45DED5D5D5D1217393F012F2B2BCD2F2BCDD4
363702902126775C4656201F5F92CBBD75546C2115170D21211C9E62455761FEDE2D2D9B7B364D33
393FDD5DCD31301B40350127600D5D36202760277027B02704340B371F3A20481F4820051A08134F
FFE0400A1339082013391B201339012B2B2B2B002B012B2B59595913211523220615141713133635
C0B215391AB8FFF0401315393608153928301439293014391108163909B8FFE0401B163929401139
0000000900010000000000000000100000721FE4500571000FB74FADF10000001000000000000000
003403630044036300040363B2242F1FBA034E006D0800400E1F7F027F037F047F050430440112BF
014A001F000D0126400B1F0DC61F0D571F0D371F0D410D019E0141000D00420141000D001E014100
021E0024455258B90024021E4459594BB8020153205C58B9010F00224544B1222245445958B90C00
2B2B2B2B2B2B2B2B2B2B2B2B00017375007373004569440073730173742B2B2B2B2B732B00732B
00>] def
/CharStrings 27 dict dup begin
/.notdef 0 def
    /space 1 def
        /comma 2 def
            /period 3 def
                /W 4 def
                    /a 5 def
                        /b 6 def
                            /c 7
```



# Open Data: Useful

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



Organization for Economic Cooperation and Development. (2007).  
*OECD Principles and Guidelines for Access to Research Data from Public Funding*.  
<http://www.oecd.org/dataoecd/9/61/38500813.pdf>



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



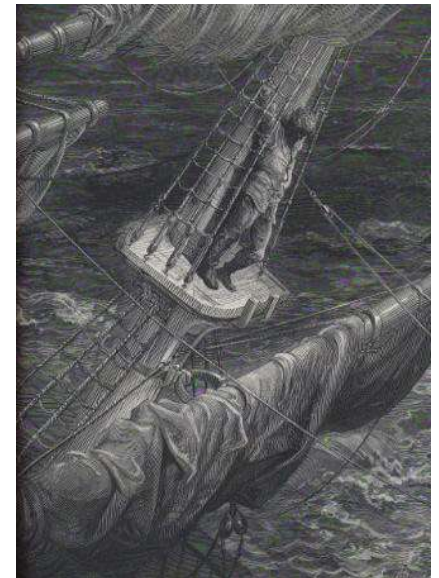
Research Data Sharing  
without barriers

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



Australian Government

National Health and Medical Research Council



National Science Foundation  
WHERE DISCOVERIES BEGIN

Supported by  
**wellcome**trust

Policy RECommendations for Open Access to Research Data in Europe







Harvard DataVerse

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

 Metrics

1,321,822 Downloads



World Agroforestry Centre  
- ICRAF DataVerse



Population Services  
International (PSI)  
DataVerse



International Food Policy  
Research Institute (IFPRI)  
DataVerse



Murray Research Archive  
DataVerse



Search this dataverse...

 Find

[Advanced Search](#)

 Add Data

☒  **Dataverses (1,080)**

1 to 10 of 59,222 Results

 Sort

« < Previous **1** 2 3 4 5 Next > »

☒  **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



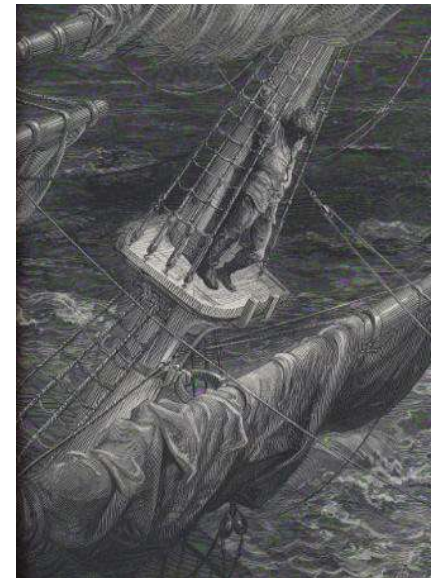
Jun 8, 2015 - Catie Snow Bailard DataVerse

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:P5c93FSwnOOdZaL+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



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.

Share this page:



Search epic.org

### Support EPIC

EPIC relies on support from individual donors to pursue our work.

[Defend Privacy. Support EPIC.](#)

### Subscribe to the EPIC Alert

The EPIC Alert is a by-monthly newsletter highlighting emerging privacy issues.

email address

GO

### #Privacy

#### Tweets



**StealthCoin Fans**  
@stealthcoinfans

13h

# Bibliometrics, Scientometrics, Informetrics, Webometrics...

data—associating stored genes with nonidentifying numbers—to protect privacy.<sup>19</sup> Other guidelines recommend anonymization in contexts such as electronic commerce,<sup>20</sup> internet service provision,<sup>21</sup> data mining,<sup>22</sup> and national security data sharing.<sup>23</sup> Academic researchers rely heavily on anonymization to protect human research subjects, and their research guidelines recommend anonymization generally,<sup>24</sup> and specifically in education,<sup>25</sup> computer network monitoring,<sup>26</sup> and health studies.<sup>27</sup> Professional statisticians are duty-bound to anonymize data as a matter of professional ethics.<sup>28</sup>

Market pressures sometimes compel businesses to anonymize data. For example, companies like mint.com and wesabe.com provide web-based personal finance tracking and planning.<sup>29</sup> One way these companies add value is by aggregating and republishing data to help their customers compare their spending with that of similarly situated people.<sup>30</sup> To make customers comfortable with this type of data sharing, both mint.com and wesabe.com promise to anonymize data before sharing it.<sup>31</sup>

Architecture, defined in Lessig's sense as technological constraints,<sup>32</sup> often forces anonymization, or at least makes anonymization the default choice. As one example, whenever you visit a website, the distant computer with which you communicate—also known as the web server—records some information

19. Roberto Andorno, *Population Genetic Databases: A New Challenge to Human Rights, in ETHICS AND LAW OF INTELLECTUAL PROPERTY 39* (Christen Lenz, Nils Hoppe & Roberto Andorno eds., 2007).

20. ALEX BRISON & LARRY DUBOV, *MASTER DATA MANAGEMENT AND CUSTOMER DATA INTEGRATION FOR A GLOBAL ENTERPRISE* 358–39 (2007).

21. See *id.* Part I.A.3.b.

22. CLK GUPTA, *INTRODUCTION TO DATA MINING WITH CASE STUDIES* 432 (2006).

23. MARKLE FOUND. TASK FORCE, *CREATING A TRUSTED NETWORK FOR HOMELAND SECURITY* 144 (2003), available at [http://www.markle.org/downloads/marklet\\_report2\\_full\\_report.pdf](http://www.markle.org/downloads/marklet_report2_full_report.pdf).

24. See THE SAGE ENCYCLOPEDIA OF QUALITATIVE RESEARCH METHODS 196 (Lisa M. Given ed., 2008) (entry for "Data Security").

25. LOUIS COHEN ET AL., *RESEARCH METHODS IN EDUCATION* 189 (2003).

26. See Roaming Pang et al., *The Devil and Pocket Trust: Anonymization*, 36 COMP. COMM. REV. 29 (2006).

27. INST. OF MED., *PROTECTING DATA PRIVACY IN HEALTH SERVICES RESEARCH* 178 (2000).

28. European Union Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concepts of Personal Data*, 01248/07/EN WP 136, at 21 (June 20, 2007) [hereinafter 2007 Working Party Opinion], available at [http://ec.europa.eu/justice\\_home/euipracy/docs/wpdocs/2007/wp136\\_en.pdf](http://ec.europa.eu/justice_home/euipracy/docs/wpdocs/2007/wp136_en.pdf).

29. See Eric Benderoff, *Spend and Save the Social Way—Personal Technology*, SEATTLE TIMES, Nov. 5, 2008, at A9.

30. See Carolyn Y. Johnson, *Online Social Networking Meets Personal Finance*, N.Y. TIMES, Aug. 7, 2007, available at <http://www.nytimes.com/2007/08/07/technology/07iht-debt.1.7013213.html>.

31. See, e.g., Wesabe, *Security and Privacy*, <http://www.wesabe.com/page/security> (last visited June 12, 2010); Mint.com, *How Mint Personal Finance Management Protects Your Financial Safety*, <http://www.mint.com/privacy> (last visited June 12, 2010).

32. LESSIG, *supra* note 18, at 4.

Aad, G., T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abidinov, et al. 2012. "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC." *Physics Letters [Part B]* 716 (1):1–29. doi:10.1016/j.physletb.2012.08.020.

Abbate, Janet. 1999. *Inventing the Internet*. Cambridge, MA: MIT Press.

Accomazzi, Alberto. 2010. "Astronomy 3.0 Style." *Astronomical Society of the Pacific Conference Series* 433: 273–281.

Accomazzi, Alberto, and Rahul Dave. 2011. "Semantic Interlinking of Resources in the Virtual Observatory Era." *Astronomical Society of the Pacific Conference Series* 442: 415–424. doi: arXiv:1103.5958.

Acropolis Museum. 2013. "The Frieze." <http://www.theacropolismuseum.gr/en/content/frieze-0>.

Agosti, Maristella, and Nicola Ferro. 2007. "A Formal Model of Annotations of Digital Content." *ACM Transactions on Information Systems* 26 (1). doi:10.1145/1292591.1292594.

Agre, Philip E. 1994. "From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing." *Computer Supported Cooperative Work* 3 (2):167–195. doi:10.1007/BF00773446.

Ahn, Christopher P., Rachael Alexandroff, Carlos Allende Prieto, Scott F. Anderson, Timothy Anderton, Brett H. Andrews, Éric Aubourg, et al. 2012. "The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey." *Astrophysical Journal* 203:21. doi:10.1088/0067-0049/203/2/21.

Akyildiz, I. F., W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. "Wireless Sensor Networks: A Survey." *Computer Networks* 38 (4):393–422. doi:10.1016/S1389-1286(01)00302-4.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57, 1701.

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.

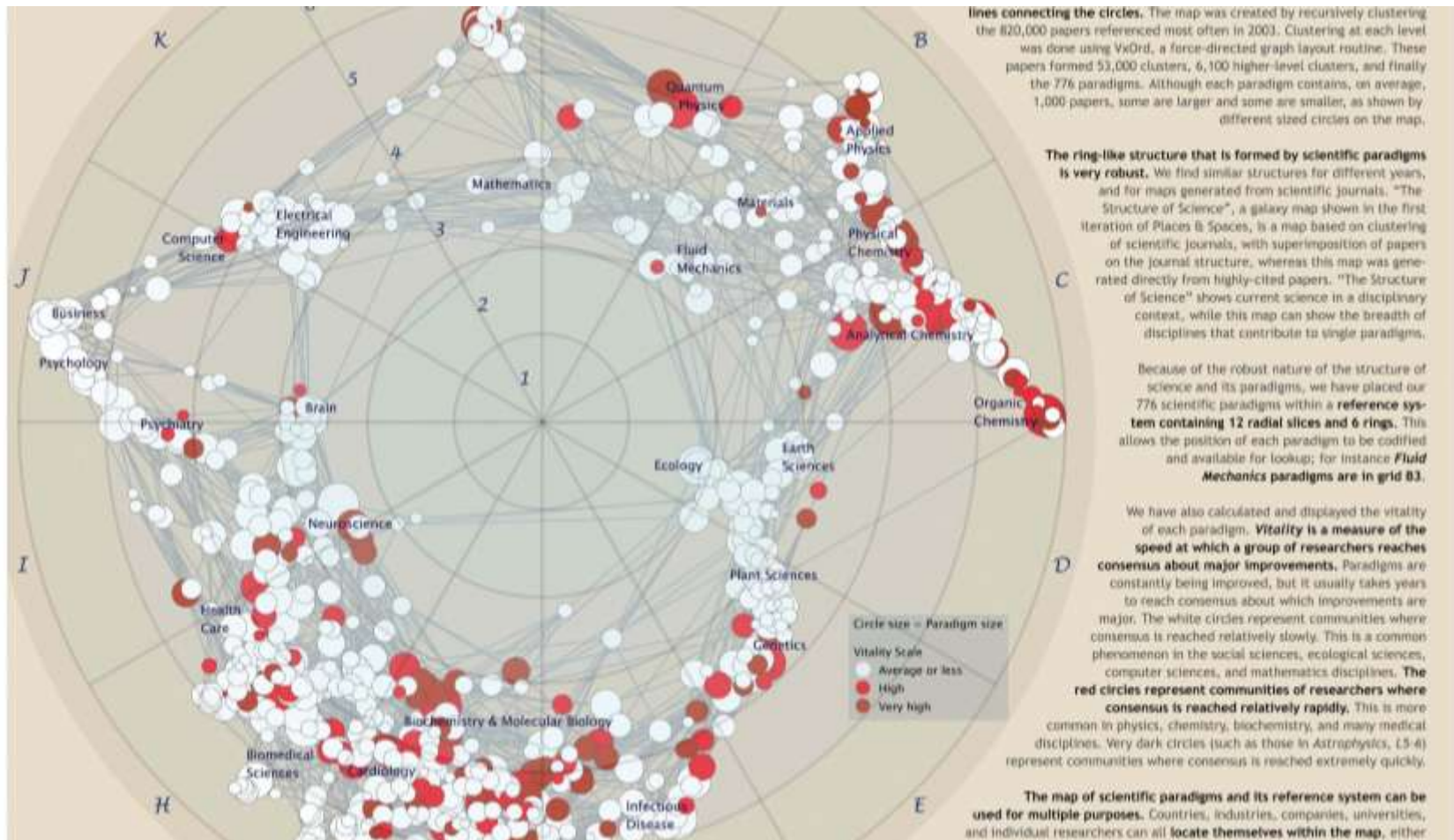


# Indicators from Publication Data

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

Source	Publications	Citations received	H-index
Google Scholar (Google)	380	7766	39
Web of Science (Thomson-Reuters)	145	1629	20
Scopus (Elsevier)	77	1314	14 (after 1995)

# Mapping Scholarship



Börner, K. (2010). *Atlas of Science: Visualizing What We Know*. Cambridge, Mass: The MIT Press.



eScholarship  
University of California

eScholarship provides Open-Access scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide. Powered by the California Digital Library.

[My Account](#)

[My Saved Items \(0\)](#)

[GO](#)

[About eScholarship](#)

[Deposit/Publish](#)

[Browse Content](#)

[Help](#)

## UC Faculty Adopt Open Access Policy

Charting a new direction  
in scholarly communication

[» Learn more](#)

**UC** **OPEN  
ACCESS  
POLICY**

[Learn more](#) [»](#)

[DEPOSIT](#)

[WAIVER](#)

[FAQ](#)

### Popular Research This Month

1. [Ebola Virus Disease:  
Essential Public He...](#)
2. [An Introduction To Green  
Marketing](#)
3. [How Languages are Learned](#)
4. [A Spectral Analysis of World  
GDP Dynamic...](#)
5. [Making Data Count: A Data  
Metrics Pilot ...](#)

[» SEE MORE](#)

## Manage Your Publications

### Books

Publish online.  
Print on demand.

### Journals

Peer review and  
MS management  
made simple.

### Working Papers

Share your research  
before it gets stale  
(or scooped).

### Previously Published Works

Open your  
publications to  
the world.

### Conferences

Plan your meeting.  
Publish the results.

### eScholarship by the Numbers

<b>Views Since 2002:</b>	23,216,720
<b>Publications:</b>	75,712
<b>Research Units:</b>	338
<b>Journals:</b>	70



Powered by The California Digital Library [Terms and Conditions](#) [Privacy Policy](#) [Help](#) [Contact Us](#) © 2012 The Regents of The University of California





# Altmetrics

## Ten Simple Rules for the Care and Feeding of Scientific Data

Alyssa Goodman, Alberto Pepe , Alexander W. Blocker, Christine L. Borgman, Kyle Cranmer, Merce Crosas, Rosanne Di Stefano, Yolanda Gil, Paul Groth, Margaret Hedstrom, David W. Hogg, Vinay Kashyap, Ashish Mahabal, Aneta Siemiginowska, Aleksandra Slavkovic

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

**191**  
Saves

**3**  
Citations

**22,098**  
Views

**360**  
Shares

[Article](#)[Authors](#)[Metrics](#)[Comments](#)[Related Content](#)[Download PDF](#) [Print](#)[Share](#)

Included in the  
Following Collection

*PLOS Computational Biology:*  
[Ten Simple Rules](#)

Subject Areas



### 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

### Figures



Published April 24, 2014;  
Altmetrics data as of June 9, 2015

Home / Featured / Kent Wada and Christine Borgman Lead Data Governance Task Force

## KENT WADA AND CHRISTINE BORGMAN LEAD DATA GOVERNANCE TASK FORCE

February 10, 2015 by Stefanie Pietkiewicz

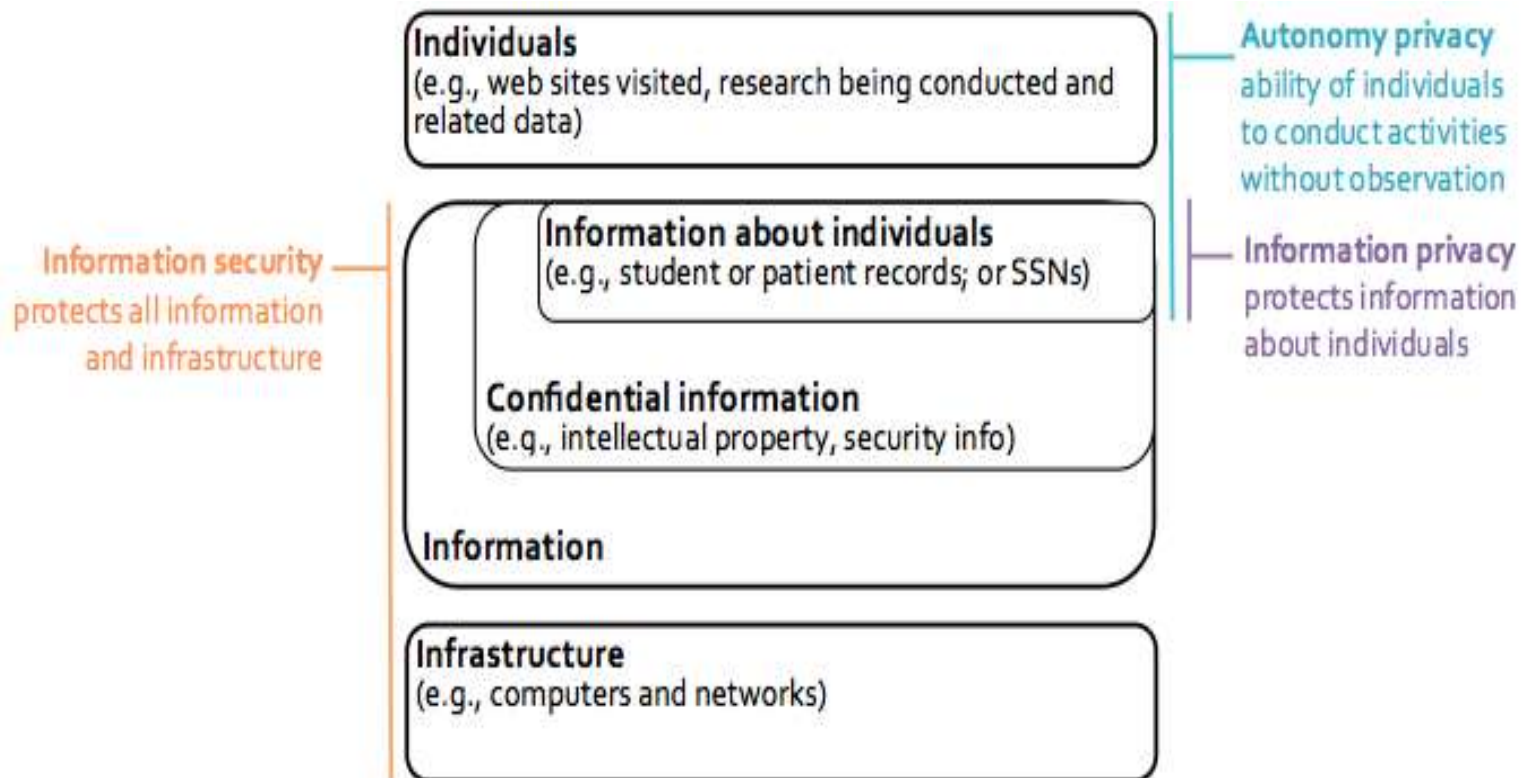


Kent Wada and Christine Borgman

- 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



# 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)

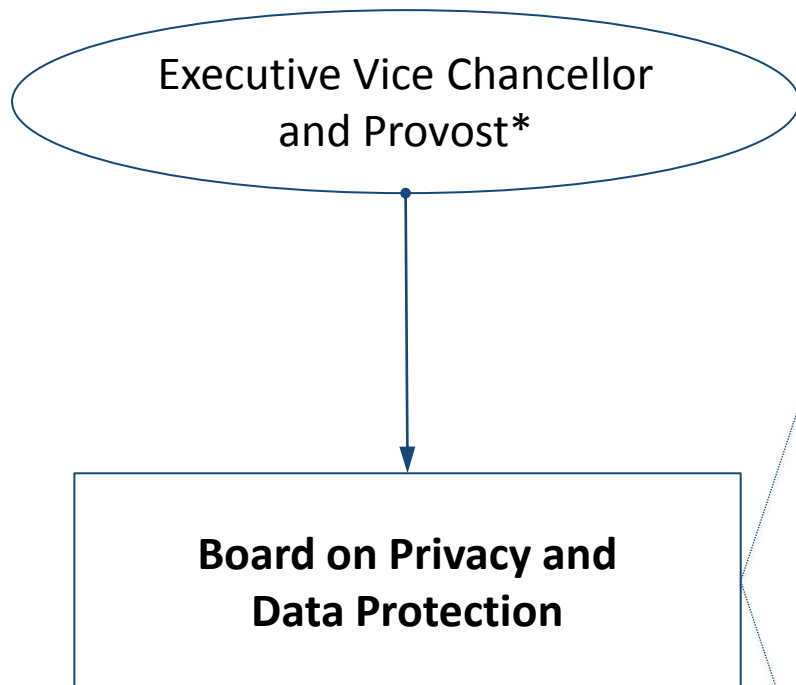
## Goals

- Resolve legitimate disagreements and provide a path forward
- Promote transparency
- Promote open discussion
- (Leverage existing structures and expertise)



# 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



#### Voting members

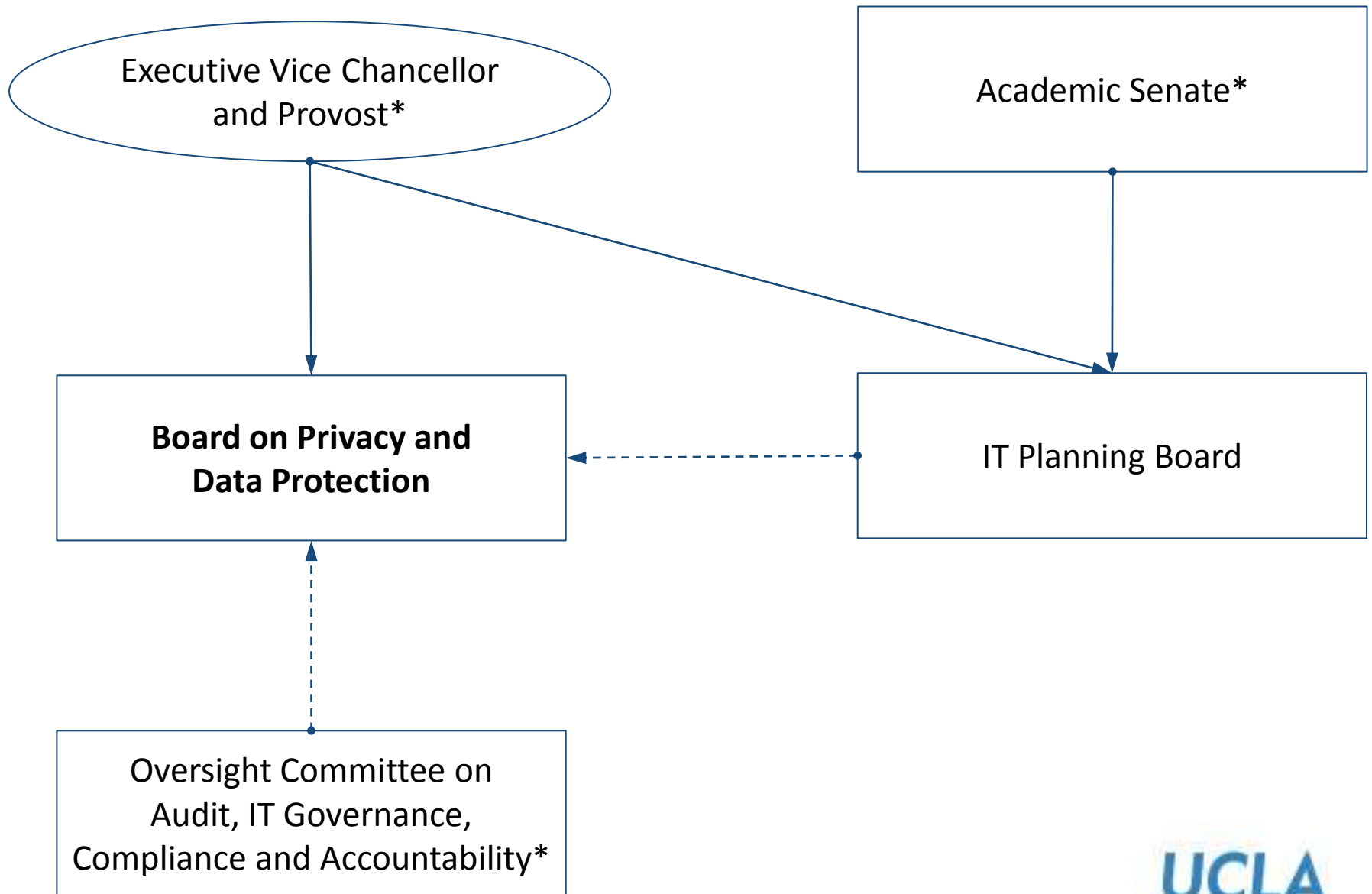
- *Faculty Chair – Appointed by EVC + Senate*
- *Administrative Vice Chair – Vice Provost, IT*
- 6 faculty members
- 6 administrative members
- 1 undergraduate student representative
- 1 graduate student representative

#### 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





\* decision-making authority

**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

**UCLA**

# 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>

