

# **UCLA**

## **Presentations**

### **Title**

Keynote Address: "Local or Global? Making Sense of the Data Sharing Imperative"

### **Permalink**

<https://escholarship.org/uc/item/2qv025hv>

### **Author**

Borgman, Christine L.

### **Publication Date**

2012-04-04

### **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/>

# Local or global? Making sense of the data sharing imperative

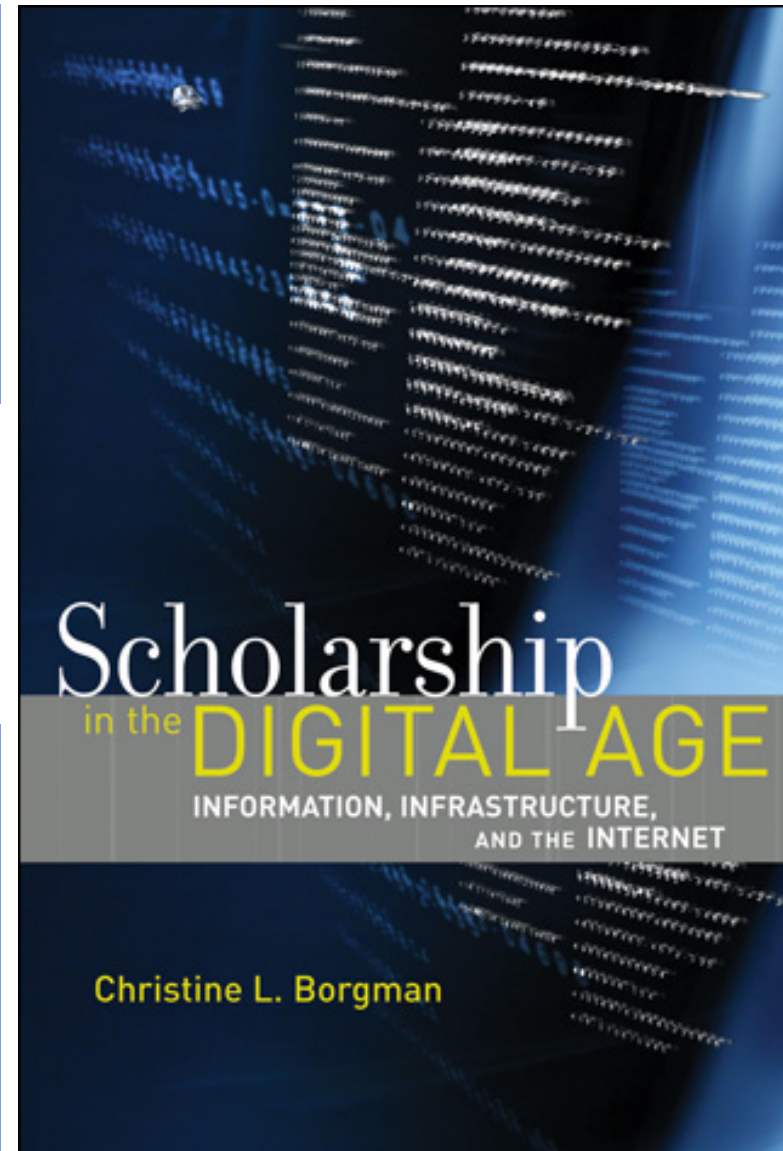
Christine L. Borgman

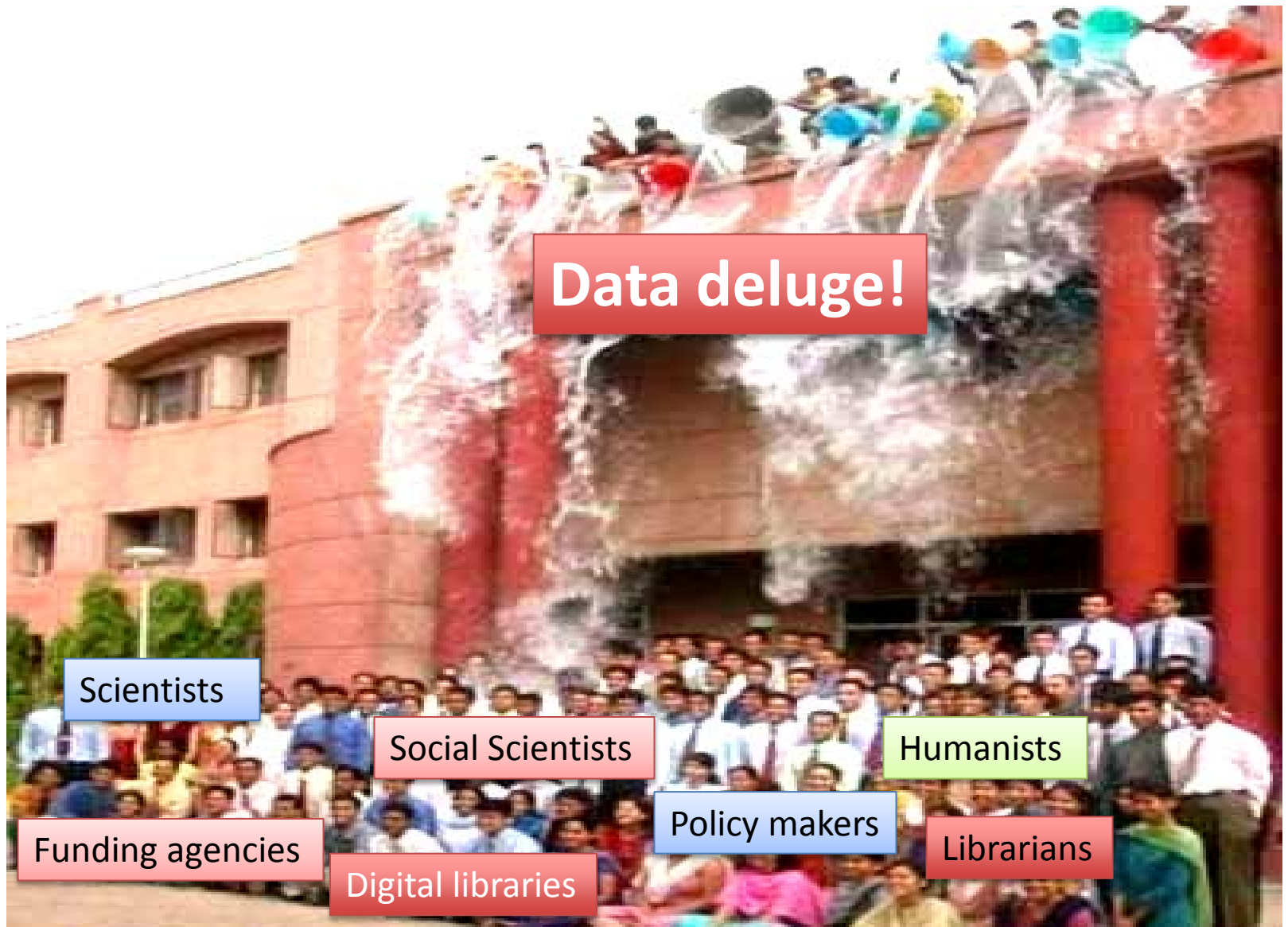
Professor & Presidential Chair in Information Studies  
University of California, Los Angeles

Keynote Presentation

University of Massachusetts and New England  
Area Librarians eScience Symposium

April 4, 2012







# Data sharing imperatives

- National Science Foundation

- Data sharing requirements
- Data management plans



National Science Foundation  
WHERE DISCOVERIES BEGIN

- Wellcome Trust

- Data sharing requirements
- Data management plans

Supported by  
**wellcome**trust

- Economic and Social Research Council

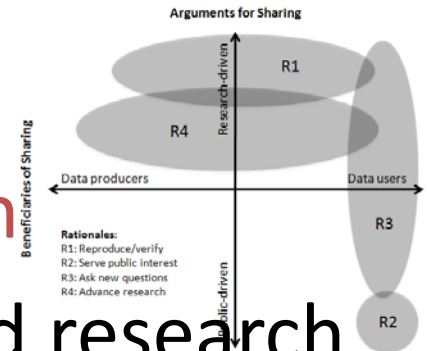
- Data sharing requirements
- Data reuse
- Data deposit



# Why share research data?

## Rationales

1. To reproduce or to verify research
2. To make results of publicly funded research available to the public
3. To enable others to ask new questions of extant data
4. To advance the state of research and innovation

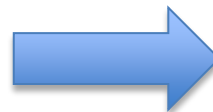
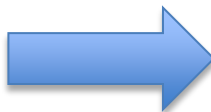


# 1. Reproduce or verify research



Benzoic Acid	%yield		IR Peaks (cm <sup>-1</sup> )		Solid (C) or Oil (O) Product	Mp (°C)
	Gross	Recrystallization	N-H	C=O		
Sodium benzoate		2.58	3327	1638	White C	79-89
Sodium benzoate			3337	1640&1600	O	
Sodium benzoate			3326	1642&1601	O	
Sodium benzoate	37.8		3274	1640	O	
p-nitro	51.84	10.59	3423	1693	Yellow C	152-157
m-nitro	37.38	5.43	3334	1694	Green C	152-157
Benzoic acid		7.44	3293	1642	White C	152-154
m-bromo		47.4	3316	1702	Green paste	
p-bromo		14.53	3344	1638	Pink C	164-166
p-chloro		29.69	3340	1638	Yellow C	
m-chloro		74.53	3410	1637	tan paste	
o-chloro		17.31	3422	1654	Tan C	
3,5-dinitro		44.53	3297	1647	Tan C	139-141
p-hydroxy		3.751	3401	1643	yellow/green C	210
p-amino		8.475	3411	1645	Dark O	
o-methoxy		42.49	3412	1646	Yellow O	

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



<http://serc.carleton.edu/cismi/broadaccess/groupwork.html>

# Scientific Gold Standard



REPLICATION—THE CONFIRMATION OF RESULTS AND CONCLUSIONS FROM ONE STUDY obtained independently in another—is considered the scientific gold standard.

Jasny, B. R., Chin, G., Chong, L. & Vignieri, S. (2011). Again, and again, and again. *Science*, 334(6060): 1225.







Victoria Stodden,  
Columbia

- Deductive sciences
  - Check the proof
- Experimental sciences
  - Redo the field work
- Computational sciences
  - Start with the dataset
  - Reconstruct workflow

# Reproducibility?

<b>Analytic validity</b>	Do different labs, techniques, and platforms measure the same thing?
<b>Repeatability</b>	Can other scientists access the data and protocols, repeat the analyses, and get the same results?
<b>Replication</b>	Do many different data sets and their combination (meta-analysis) get consistent results?
<b>External validation</b>	Do different data sets by different teams, preferably prospectively and with large-scale evidence, get consistent results?
<b>Clinical validity</b>	Does the discovered information predict clinical outcomes?
<b>Clinical utility</b>	Does the use of the discovered information improve clinical outcomes?





# Data, Replication, and Interpretation

- Unit of replication
  - One paper
  - One dataset
  - One program of research
- Provenance
  - Chain of custody
  - Transformations from original state
- Tacit knowledge
  - Domain knowledge
  - Research methods
  - Research skills



[http://chicagoist.com/2008/10/09/a\\_gourmet\\_oasis\\_provenance\\_food\\_and.php](http://chicagoist.com/2008/10/09/a_gourmet_oasis_provenance_food_and.php)

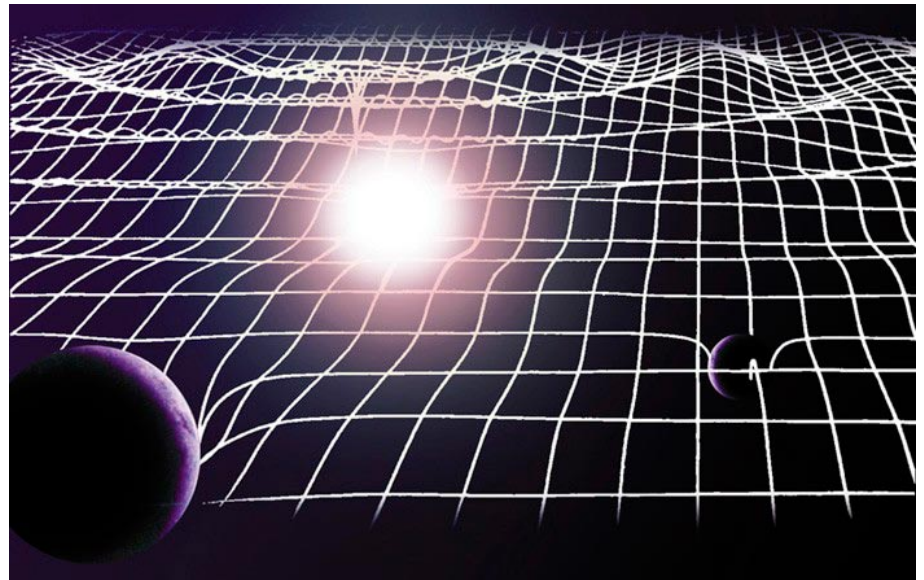
# Reproducibility rationales

- Resolve disputes
- Confirm scientific claims
- Protect public interest



# Resolve disputes?

- Gravitational waves
- Valid experiments were those that
  - Detected waves
  - Failed to detect waves



Collins, H. M. (1975). The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, 9: 205-24.

Collins, H. M. (1998). The meaning of data: Open and closed evidential cultures in the search for gravitational waves. *American Journal of Sociology*, 104(2): 293-338.

# Gravitational waves, 2011

Black hole twins spew gravitational waves: Physics World, April 2011

Astronomers could be on the cusp of detecting gravitational waves after four decades of trying, according to a team of Polish astrophysicists. They say that if current gravitational-wave detectors are upgraded to search for binary black-hole systems, gravitational waves would be expected "within the first year of operation". If correct, it would open up a new window to the cosmos, allowing astronomers to see the universe with fresh eyes. ...

However, a team of researchers, led by Chris Belczynski of the Los Alamos National Laboratory, report that these projects have taken the wrong option, saying that double black hole systems may be far more common than previously thought. The reason is related to stars' metallicity, which is the fraction of elements that are heavier than helium. The lower the metallicity the less mass is lost at the end of the star's life and therefore the black holes that form are more likely to survive to become a black hole binary.

# Confirm scientific claims

**12 Feb 2004: Landmark paper**

Woo Suk Hwang from Seoul National University and his colleagues announced that they have cloned human embryos and harvested stem cells from one of them (Woo Suk Hwang et al. *Nature* **303**, 1669-1674; 2004). The work makes a significant step towards stem-cell therapies for disease. Other groups have claimed to clone human embryos, but supporting evidence has been sketchy. This paper provides further supporting evidence.

- [Cloned human embryos yield stem cells](#)

- What data do peer reviewers need?
- How are data used in peer review?
- What is the responsibility of peer reviewers to reproduce research?

The screenshot shows the top of a Nature journal article page. The header features the 'nature' logo and the tagline 'International weekly journal of science'. Below the logo is a navigation bar with links for 'nature news home', 'news archive', 'specials', 'opinion', 'features', 'news blog', and 'nature'. The article title is 'Cloned human embryos yield stem cells'. The publication date is '12 Feb 2004'. The author is 'Woo Suk Hwang et al.'. The article is categorized as 'News'. A 'Timeline of a controversy' section is visible, titled 'A chronology of Woo Suk Hwang's stem-cell research.' It discusses concerns about ethics and fraud. A photo of Woo Suk Hwang is shown with the caption 'Woo Suk Hwang faces questions.' The article is dated '31 October 2006'. A link to 'Hwang takes the stand at fraud trial' is provided at the bottom.

nature International weekly journal of science

nature news home news archive specials opinion features news blog nature

Published online 19 December 2005 | Nature | doi:10.1038/news051219-3

News

## Timeline of a controversy

### A chronology of Woo Suk Hwang's stem-cell research.

Concerns about ethics, errors (accidental or intentional) and possible fraud have dogged the stem-cell researcher Woo Suk Hwang, from Seoul National University in South Korea, since his landmark 2004 *Science* paper on stem cells from a cloned human embryo. Here [news@nature.com](mailto:news@nature.com) describes how events have unfolded from that initial paper - with the most recent events presented first (you may want to read from the bottom-up the first time you read this). Keep checking back for updates over the coming weeks.

**31 October 2006**

A confident and defiant Hwang takes the stand for the first time in court. The defence denies allegations of fraud and embezzlement, and has prepared a case against the charge of violating the bioethics law for the next hearing. A verdict may be handed down by the end of the year.

• [Hwang takes the stand at fraud trial](#)

Comments on this story

Stories by keywords

- [stem cells](#)

This article elsewhere

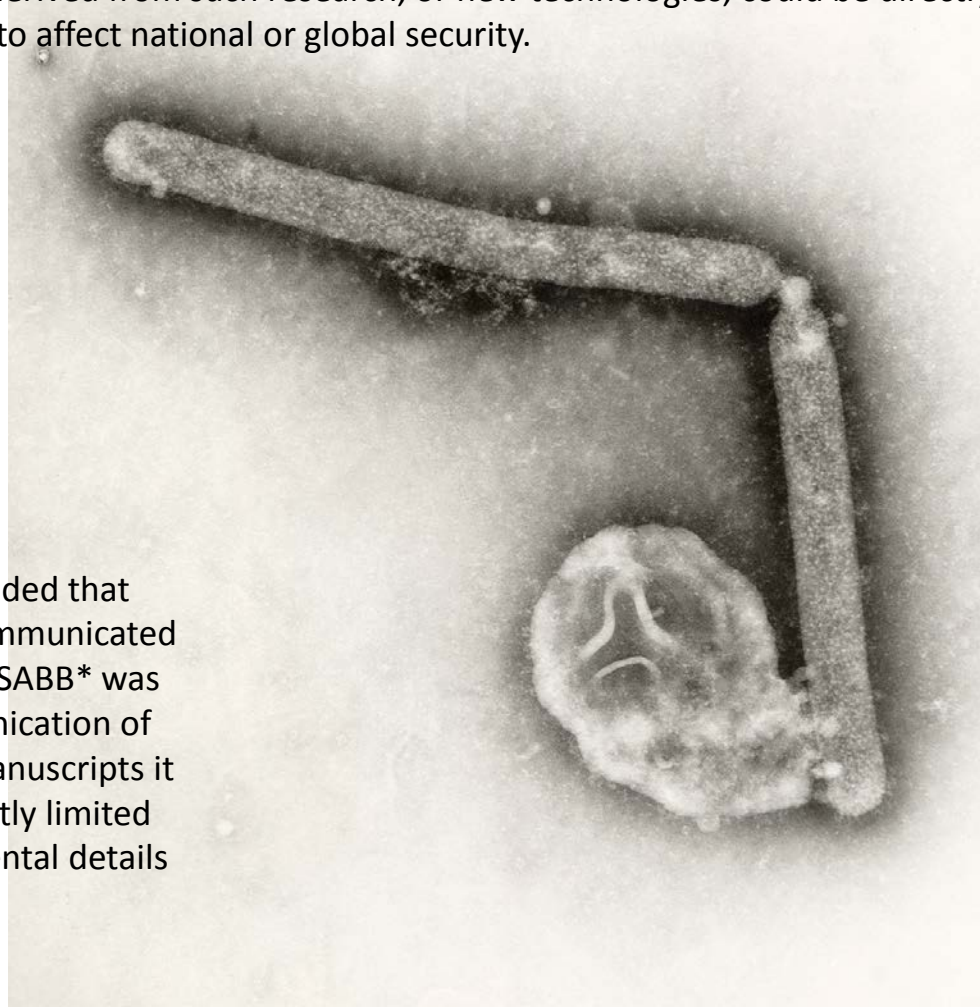
- [Blogs linking to this article](#)
- [Add to Connotea](#)
- [Add to Digg](#)
- [Add to Facebook](#)
- [Add to Newsvine](#)
- [Add to Del.icio.us](#)
- [Add to Twitter](#)

## Avian influenza A/H5N1 virions.

Efforts to describe or define life-sciences research of particular concern have focused on the possibility that knowledge or products derived from such research, or new technologies, could be directly misapplied with a sufficiently broad scope to affect national or global security.

We found the potential risk of public harm to be of unusually high magnitude.

We therefore recommended that the work not be fully communicated in an open forum. The NSABB\* was unanimous that communication of the results in the two manuscripts it reviewed should be greatly limited in terms of the experimental details and results.



This is an **unprecedented recommendation** for work in the life sciences ... Our concern is that publishing these experiments in detail would provide information to some person, organization, or government that would help them to develop similar mammal-adapted influenza A/H5N1 viruses for harmful purposes.

K I Berns et al. *Science* 2012;335:660-661

\*U.S. National Science Advisory Board for Biosecurity





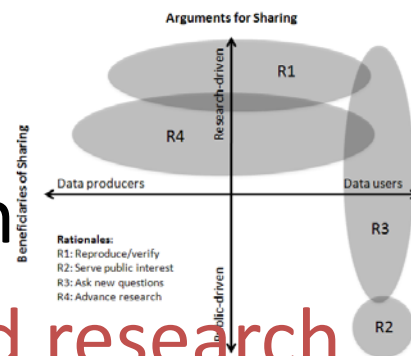
Sharing data ↔ reproducibility?



# Why share research data?

## Rationales

1. To reproduce or to verify research
2. To make results of publicly funded research available to the public
3. To enable others to ask new questions of extant data
4. To advance the state of research and innovation



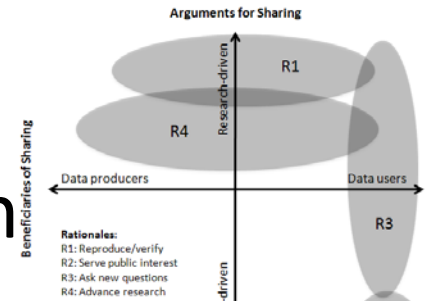
## 2. Public monies serve the public good



# Why share research data?

## Rationales

1. To reproduce or to verify research
2. To make results of publicly funded research available to the public
3. To enable others to ask new questions of extant data
4. To advance the state of research and innovation



# 3. Others can ask new questions



data



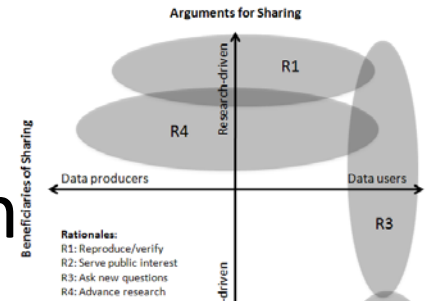
discovery

<http://annualreport.ucdavis.edu/2008/images/photos/discovery.jpg>

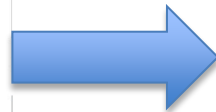
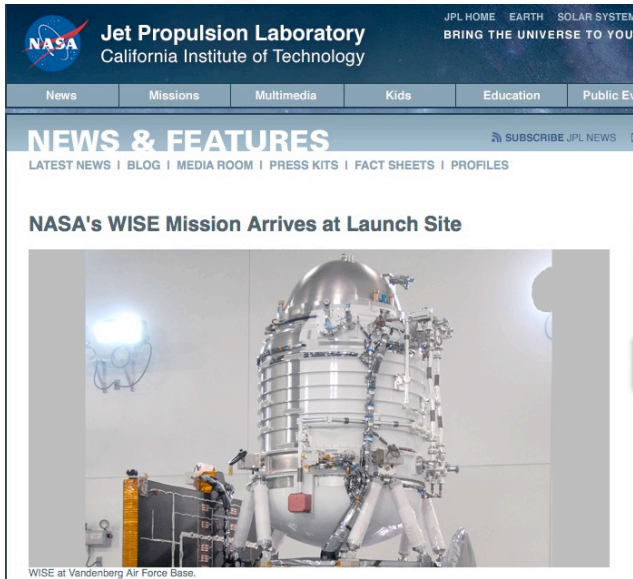
# Why share research data?

## Rationales

1. To reproduce or to verify research
2. To make results of publicly funded research available to the public
3. To enable others to ask new questions of extant data
4. To advance the state of research and innovation



# 4. Data curation advances research



International Virtual Observatory Alliance



# The Conundrum of Sharing Research Data

*If the rewards of the data deluge are to be reaped, then researchers who produce those data must share them, and do so in such a way that the data are interpretable and reusable by others.\**

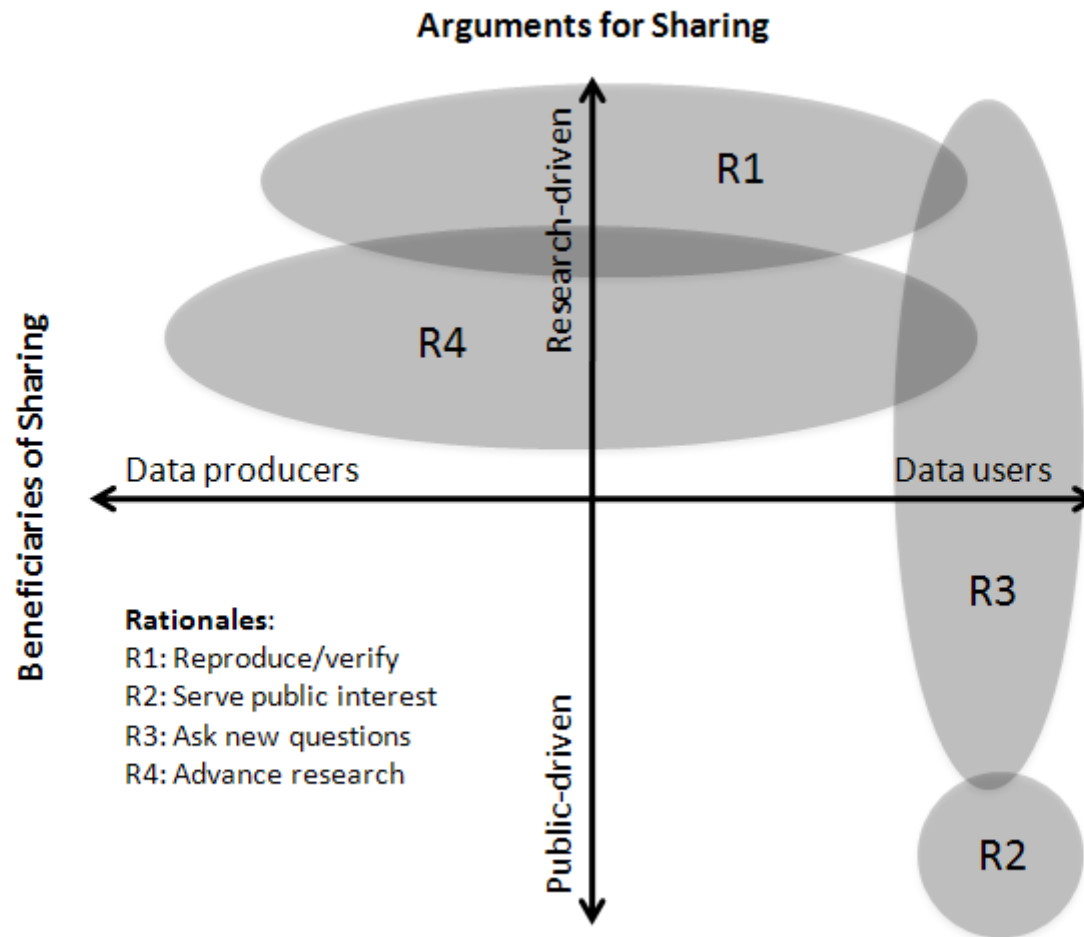
\*Borgman, C. L. (2012, forthcoming). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*.







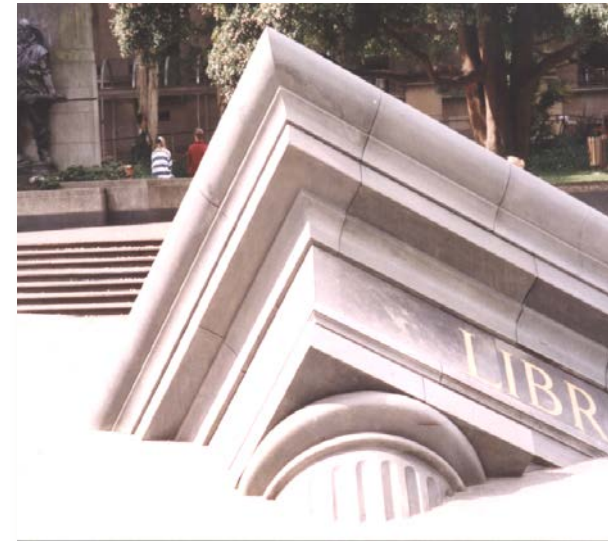
# Rationales for Sharing Research Data





# Conclusions

- Rationales for data sharing are implicit
  - To reproduce or to verify research
  - To make results of publicly funded research available to the public
  - To enable others to ask new questions of extant data
  - To advance the state of research and innovation
- Incentives to share are implicit
- “Data” remains a complex construct
- Librarians and archivists are key





# 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.
  - *Knowledge and Data Transfer: the Formation of a New Workforce. # 1145888*. C.L. Borgman, PI; S. Traweck, Co-PI.
- Microsoft External Research: Tony Hey, Lee Dirks, Catherine van Ingen, Catherine Marshall
- Sloan Foundation: *The Transformation of Knowledge, Culture, and Practice in Data-Driven Science: A Knowledge Infrastructures Perspective. # 20113194*. C.L. Borgman, PI; S. Traweck, Co-PI. Joshua Greenberg, program director
- Project website: <http://knowledgeinfrastructures.gseis.ucla.edu/index.html>