

# UCLA

## Presentations

### Title

The Digital Archive: The Data Deluge arrives in the Humanities

### Permalink

<https://escholarship.org/uc/item/2cc978tf>

### Author

Borgman, Christine L.

### Publication Date

2010-05-10

### 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/>

# The Digital Archive: The Data Deluge arrives in the Humanities

Christine L. Borgman

Professor & Presidential Chair  
in Information Studies, UCLA

“Time Will Tell, But Epistemology Won't: In Memory of Richard Rorty”

A Celebration of Richard Rorty's Archive

UC-Irvine, May 14, 2010

UC-Irvine, May 14, 2010

Deluge!!!



Data!!

Historians

Literary Theorists

Philosophers

Classicists

I'm finding that something as simple as constructing my maps of related concepts are not easily applied to primary sources in digital libraries. So what use are the digital libraries, if all they do is put digitally unusable information on the web? The digital libraries don't offer a platform for traditional note taking, much less for larger scale analysis, either quantitative or qualitative.





## UCISpace Home

### Browse UCISpace

- Collections
- Titles
- Dates
- Authors
- Subjects
- **This Collection**
  - Titles
  - Dates
  - Authors
  - Subjects

### My Account

- Login

### About UCISpace

### Contact Us

## Richard Rorty born digital files, 1988-2003

Search within this collection and its subcollections:

[Advanced Search](#)

Files retrieved from Richard Rorty's 3.5-inch floppy disks during the processing of his papers.

Richard Rorty (1931-2007) was a pragmatist philosopher, critical theorist, and public intellectual who is commonly described as one of the most important thinkers of his era. Included in this collection are electronic word-processing files, created between 1988 and 2003. The files were retrieved from Rorty's 3.5" floppy disks during processing of his personal papers and converted to PDF format in order to facilitate preservation and access.

Included are letters, many drafts of writings, lecture notes, syllabi, and exams. Also included are bibliographies of his work, a few administrative files, and a couple of documents relating to his children. This collection also contains drafts of writings done by some of his peers and colleagues. Researchers will notice some data loss and corruption has occurred in some of the earliest files.

Access to Rorty's born digital files is provided through UCISpace @ the Libraries. Interested researchers should fill out the researcher application and email it to [spcoll@uci.edu](mailto:spcoll@uci.edu) or fax it to (949)824-2472. Access may be granted in less than 5 business days.

These digital files are part of the larger collection, the Richard Rorty Papers, which are described in a finding aid in the Online Archive of California. Access to these non-digital materials is provided in the Special Collections and Archives Reading Room at the UC Irvine Libraries.

Richard Rorty donated his papers to the Critical Theory Archive at the UC Irvine Libraries in 2006. Among the boxes of materials were more than 70 floppy disks containing about 1,100 word-processing files. These born-digital files were copied from the disks and saved in pdf format for long-term preservation and access. They are made available exclusively through the virtual reading room in UCISpace @ the Libraries.

### Copyright

*This material is provided for private study, scholarship, or research. Transmission or reproduction of any material protected by copyright beyond that allowed by fair use requires the written permission of the copyright owners. The authors or their heirs retain their copyrights to the material. Contact the University of California, Irvine Libraries, Special Collections and Archives for more information ([spcoll@uci.edu](mailto:spcoll@uci.edu)).*

### Subcollections in this collection

- Richard Rorty born digital administrative and professional files, 1989-1999 [12]
- Richard Rorty born digital biographical files, 1992-2001 [22]
- Richard Rorty born digital correspondence, 1992-2000 [59]
- Richard Rorty born digital family files, 1998-2001 [2]
- Richard Rorty born digital research files, 1992-2002 [45]
- Richard Rorty born digital teaching files, 1988-2002 [237]
- Richard Rorty born digital writings, 1988-2003 [642]



UCISpace Home >> Richard Rorty born digital fil... >> Richard Rorty born digital res... >> **Advanced Search**

### UCISpace Home

#### Browse UCISpace

- Collections
- Titles
- Dates
- Authors
- Subjects
- **This Subcollection**
  - Titles
  - Dates
  - Authors
  - Subjects

#### My Account

- Login

#### About UCISpace

#### Contact Us

## Advanced Search

Conjunction	Search type	Search for
	Full Text ▾	derrida
AND ▾	Full Text ▾	1999
AND ▾	Full Text ▾	
Results/page 10 ▾	Sort items by Relevance ▾	in order descending ▾
<input type="button" value="Go"/>		



UCISpace Home

Browse UCISpace

- Collections
- Titles
- Dates
- Authors
- Subjects
- **This Subcollection**
  - Titles
  - Dates
  - Authors
  - Subjects

My Account

- Login

About UCISpace

Contact Us

Advanced Search

Conjunction	Search type	Search for
AND	Full Text	derrida
AND	Full Text	1999
AND	Full Text	

Results/page  Sort items by  in order

Your query "((derrida) AND (1999))" produced 1 result(s).

**Search Results for Subcollection: Richard Rorty born digital research files, 1992-2002**

---

Now showing items 1-1 of 1

1

Gumbrecht, Hans Ulrich. "Socrates und Uncle Sam: Wie anders Richard Rorty die Rolle des Philosophen verkoerpert:" draft  
 Gumbrecht, Hans Ulrich (*University of California, Irvine Libraries. Special Collections and Archives. Critical Theory Archive. Irvine, California 92623-9557.*  
*http://special.lib.uci.edu. email: spcoll@uci.edu, 2010-02-17*)  
 Date Created: 1998-10-13T15:47:20PDT

Now showing items 1-1 of 1

1



UCIspace Home

Browse UCIspace

- Collections
- Titles
- Dates
- Authors
- Subjects
- **This Subcollection**
  - Titles
  - Dates
  - Authors
  - Subjects

My Account

- Login

About UCIspace

Contact Us

Gumbrecht, Hans Ulrich. "Socrates und Uncle Sam: Wie anders Richard Rorty die Rolle des Philosophen verkoerpert:" draft

Item File(s)

Files	Size	Format	View
gumbrechts article on rorty.pdf	48.44Kb	PDF	View/Open

The following license files are associated with this item:

- Original License

Item Description

**Title:** Gumbrecht, Hans Ulrich. "Socrates und Uncle Sam: Wie anders Richard Rorty die Rolle des Philosophen verkoerpert:" draft

**Author(s):** Gumbrecht, Hans Ulrich

**Contributor(s):** Rorty, Richard

**Date Created:** 1998-10-13

**Permanent Link to This Record:** <http://hdl.handle.net/10575/119>

**Type:** Archives and Manuscripts

---

**Subject:** Pragmatism

---

Philosophy

---

Rorty, Richard

---

**Language:** ger

**File Format:** application/pdf

**Rights:** This material is provided for private study, scholarship, or research. Transmission or reproduction of any material protected by copyright beyond that allowed by fair use requires the written permission of the copyright owners. The authors or their heirs retain their copyrights to the material. Contact the University of California, Irvine Libraries, Special Collections and Archives for more information ([spcoll@uci.edu](mailto:spcoll@uci.edu)).

**Access Rights:** To access this item, interested researchers should fill out the researcher application ( [http://www.lib.uci.edu/libraries/collections/special/using/patron\\_registration.pdf](http://www.lib.uci.edu/libraries/collections/special/using/patron_registration.pdf) ) and email it to [spcoll@uci.edu](mailto:spcoll@uci.edu) or fax it to (949)824-2472. Access may be granted in less than 5 business days.

**Publisher:** University of California, Irvine Libraries. Special Collections and Archives. Critical Theory Archive. Irvine, California 92623-9557. <http://special.lib.uci.edu>. email:







Search this blog

GO

[About This Blog](#)**Our Twitter Feed**

[Nominatios Open for Bobbitt Poetry Prize: The Library of Congress is accepting nominations from publishers for th...](#) <http://bit.ly/d9qZz0>

[Go to our Twitter Page](#)

**Categories**

- > [Abraham Lincoln](#)
- > [American Folklife Center](#)
- > [Audiovisual](#)
- > [Blogging](#)
- > [Books](#)
- > [Capitol Hill](#)
- > [Cataloging](#)
- > [Collections](#)
- > [Concerts](#)
- > [Congress](#)
- > [Copyright](#)
- > [Curators](#)
- > [Education](#)
- > [Events](#)
- > [Exhibitions](#)
- > [film](#)
- > [folksonomies](#)
- > [History](#)
- > [Kluqe Center](#)
- > [Law Library](#)
- > [LC Web Site](#)
- > [Libraries](#)
- > [Manuscripts](#)
- > [Maps](#)
- > [Music](#)
- > [National Book Festival](#)
- > [New Visitors Experience](#)
- > [News](#)

[Print](#) [Subscribe](#) [Share/Save](#)

## How Tweet It Is!: Library Acquires Entire Twitter Archive

April 14th, 2010 by [Matt Raymond](#)

[**UPDATE:** We posted an [FAQ](#) on April 28]

Have you ever sent out a "tweet" on the popular Twitter social media service? Congratulations: Your 140 characters or less will now be housed in the Library of Congress.

That's right. Every public tweet, ever, since Twitter's inception in March 2006, will be archived digitally at the Library of Congress. That's a LOT of tweets, by the way: Twitter processes more than 50 million tweets every day, with the total numbering in the billions.

We thought it fitting to give the initial [heads-up](#) to the Twitter community itself via our own feed [@librarycongress](#). (By the way, out of sheer coincidence, the announcement comes on the same day our own number of feed-[followers](#) has surpassed 50,000. I love serendipity!)

We will also be putting out a press release later with even more details and quotes. Expect to see an emphasis on the scholarly and research implications of the acquisition. I'm no Ph.D., but it boggles my mind to think what we might be able to learn about ourselves and the world around us from this wealth of data. And I'm certain we'll learn things that none of us now can even possibly conceive.

Just a few examples of important tweets in the past few years include the first-ever tweet from Twitter co-founder Jack Dorsey (<http://twitter.com/jack/status/20>), President Obama's tweet about winning the 2008 election (<http://twitter.com/barackobama/status/992176676>), and a set of two tweets from a photojournalist who was arrested in Egypt and then freed because of a series of events set into motion by his use of Twitter (<http://twitter.com/jamesbuck/status/786571964>) and (<http://twitter.com/jamesbuck/status/787167620>).

Twitter plans to make its own announcement today on its [blog](#) from "[Chirp](#)," the Official Twitter Developer Conference, in San Francisco. (**UPDATE:** Here's their [post](#).)

So if you think the Library of Congress is "just books," think of this: The Library has been collecting materials from the web since it began harvesting congressional and presidential campaign websites in 2000. Today we hold more than 167 terabytes of web-based information, including legal blogs, websites of candidates for national office, and websites of Members of Congress.

We also operate the National Digital Information Infrastructure and Preservation Program [www.digitalpreservation.gov](http://www.digitalpreservation.gov), which is pursuing a national strategy to collect, preserve and make available significant digital content, especially information that is created in digital form only, for current and future generations.



software studies initiative

HOME | ABOUT | EVENTS | PROJECTS | PUBLICATIONS | IMAGES

CULTURAL ANALYTICS

### ArtHistory.viz | Mining 20,000,000+ Images of Art

ArtHistory.viz uses cultural analytics approaches to explore patterns in art, architecture, and visual culture. Techniques include image processing and computer vision, statistical analysis, information visualization, and scientific visualization.



flickr.com/photos/culturevis

ESPAÑOL | PORTUGUÊS



Lev Manovich  
Director

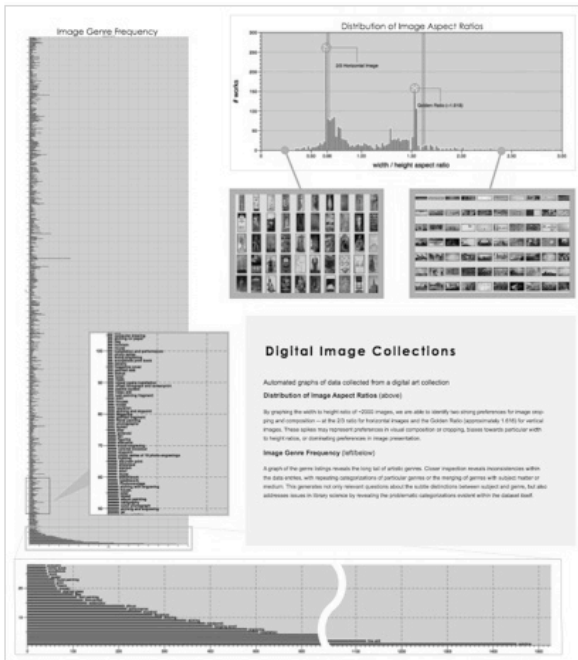
Noah Wardrip-Fruin  
Associate Director, North  
California

Benjamin Bratton  
Associate Director, South  
California

Jeremy Douglass  
Postdoctoral Scholar

Cicero Inacio da Silva  
Software Studies Brazil

Tara Zepel  
Graduate Researcher



Gu, Jia. Art Objects as Data Points.  
Presentation at Calit2 Summer Researchers Poster Session. 2008 (pdf)

# Unprecedented Data Availability

**AmeriFlux**  
About AmeriFlux  
Objectives  
Organization  
Strategic Plan  
Participants/Sites  
Participant Information  
Site Information  
List Server  
Standards  
Core/Desired Measures  
Guidelines/SOPs  
Data  
Submission Guidelines  
Available Data  
Data Access  
Key Resources  
Equipment  
Sources/Tips  
Publication Lists

**LTER**  
The US Long Term Ecological Research Network  
LTER Site Home Pages  
Go  
Search  
About LTER  
LTER Network  
What's New  
XML  
network is a collaborative effort...  
the LTER program in 1980 to...  
network publications, and

**CUAHSI**  
The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)  
0 600 1,200 Miles  
Hydrology Data Access System  
HDAS provides uniform access to several national-scale water resource databases:  
USGS NWIS  
AmeriFlux Network  
How to use it?  
To chart or download time series, please follow these steps:  
1. Zoom in on area of interest until you see station points on the map  
2. Click on the icon

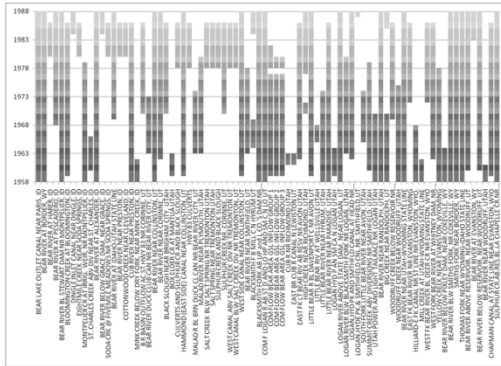
**National Weather Service**  
Hydrologic Information Center  
Home News Organization  
Current Flooding  
-WFO Products  
-River Locations in Flood  
-Flood Summary  
Outlooks  
-Hydrologic  
-Meteorological  
Current Hydrologic Conditions  
-WFO Products  
-River Conditions  
-Drought  
Archive & Historic Information  
3613 Total Gauges  
5 Locations in Flood  
296 Gauges: Observations older than 12 hours  
17 Gauges: Out of Service  
Last map update: Sun, Sep, 17, 2006 at 09:50:13 pm  
0 Gauge  
0 Gauge  
11 Gau  
3284 G

**National Climatic Data Center**  
National Climatic Data Center  
U.S. Department of Commerce  
Search Field: Search NCDC  
Climate Data  
Weather/Climate, Events, Information & Assessments  
Satellite  
Upper Air  
Marine  
Paleoclimatology  
HOW ARE WE DOING?  
FIRSTGOV  
Disclaimer

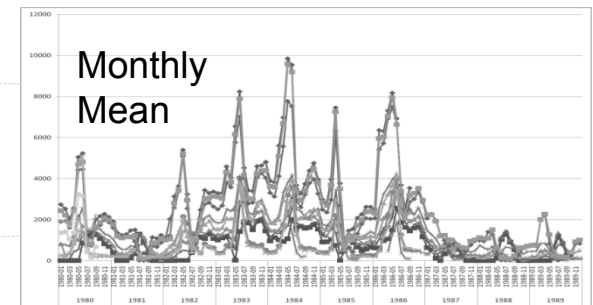
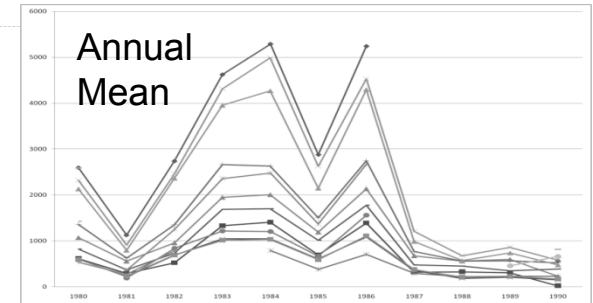
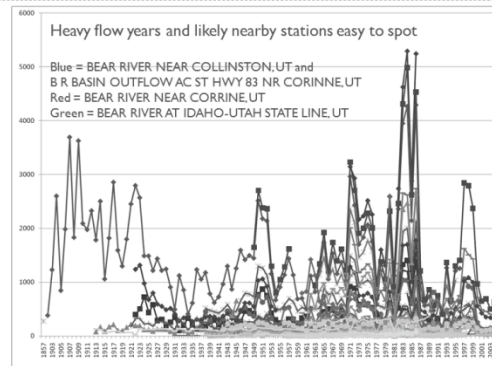
Slide courtesy of Catherine van Ingen, Microsoft

# Bear River – A First Look

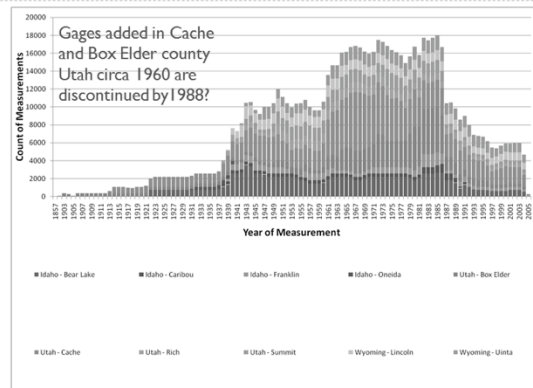
Stream gage comings and goings  
1958-1988



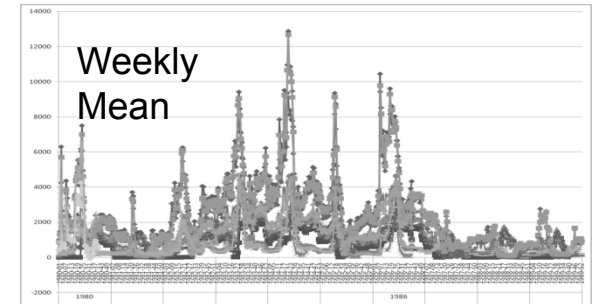
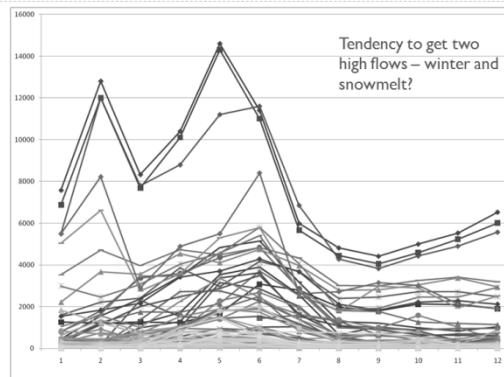
Annual Mean Daily Discharge



Daily Average Discharge Availability by State, County

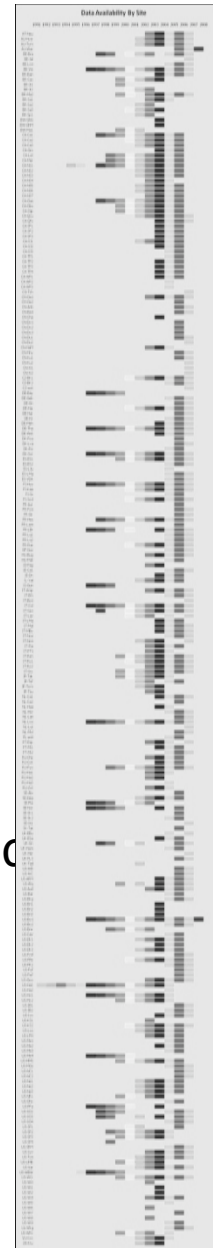
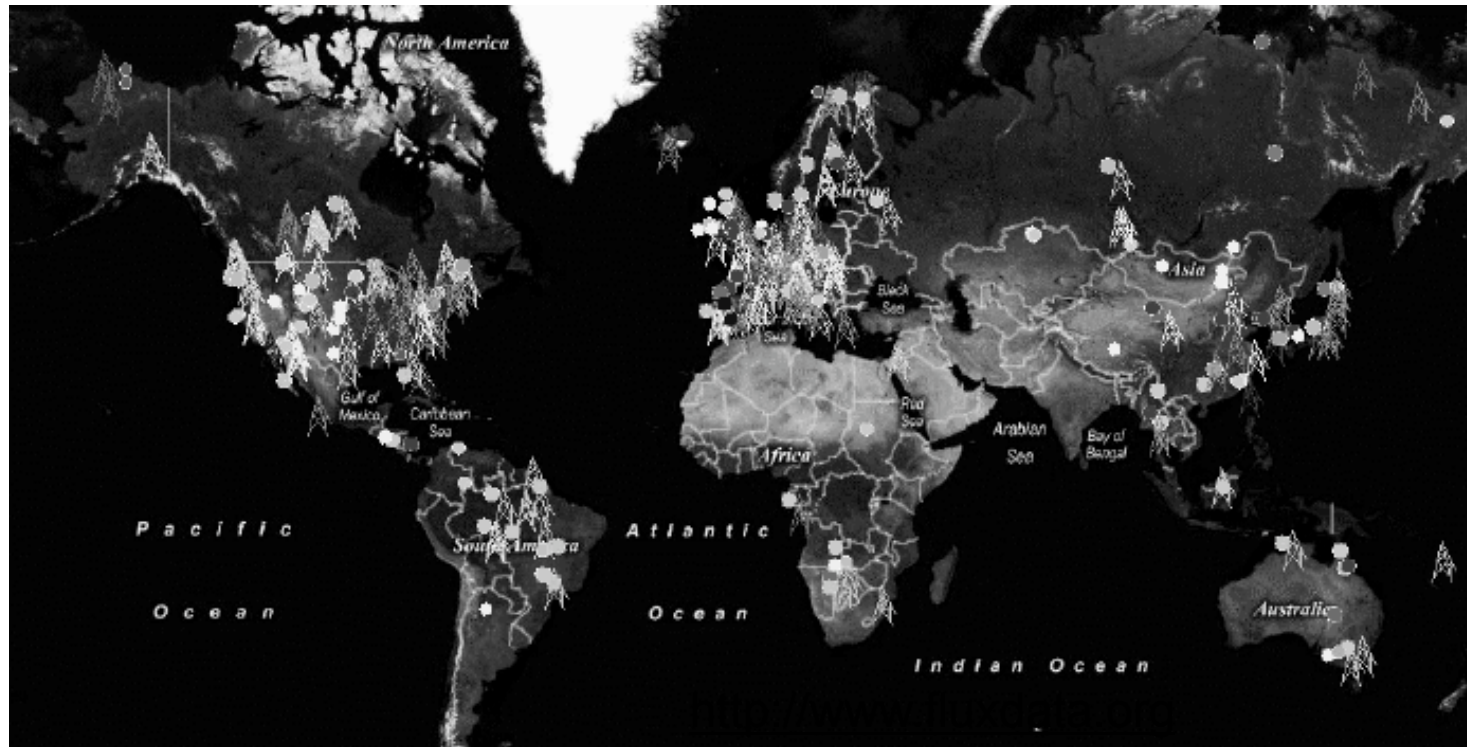


Monthly Mean Daily Discharge by Month



Slide courtesy of Catherine van Ingen, Microsoft Research

# FLUXNET : A Network of Networks



- 467 towers world wide
- 967 site-years of sensor data from 253 towers
- ~20 sensor measurements per tower; 20 derived science variables
- 145 ancillary variables
- Original data set assembled and processed in 2007
- 20x larger than previous synthesis dataset
- 75 paper teams with over 200 scientists

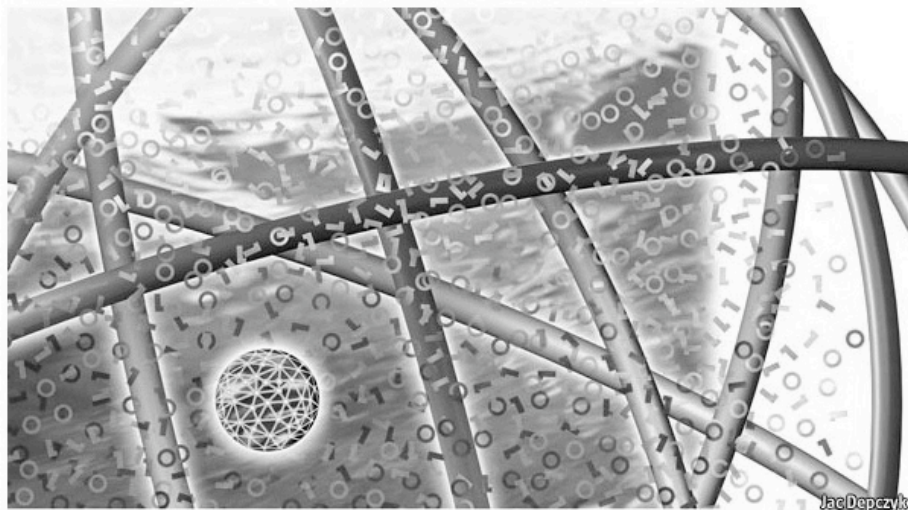
Slide courtesy of Catherine van Ingen, Microsoft Research

A special report on managing information

## Data, data everywhere

Information has gone from scarce to superabundant. That brings huge new benefits, says Kenneth Cukier (interviewed here)—but also big headaches

Feb 25th 2010 | From *The Economist* print edition



WHEN the Sloan Digital Sky Survey started work in 2000, its telescope in New Mexico collected more data in its first few weeks than had been amassed in the entire history of astronomy. Now, a decade later, its archive contains a whopping 140 terabytes of information. A successor, the Large Synoptic Survey Telescope, due to come on stream in Chile in 2016, will acquire that quantity of data every five days.

Such astronomical amounts of information can be found closer to Earth too. Wal-Mart, a retail giant, handles more than 1m customer transactions every hour, feeding databases estimated at more than 2.5 petabytes—the equivalent of 167 times the books in America's Library of Congress (see [article](#) for an explanation of how data are quantified). Facebook, a social-networking website, is home to 40 billion photos. And decoding the human genome involves analysing 3 billion base pairs—which took ten years the first time it was done, in 2003, but can now be achieved in one week.

- Comment (30)
- E-mail
- Print
- Reprints & permissions
- Recommend (343)
- Share
- Buy PDF

### Advertisement

The Economist Debate Series

## Is fair trade more important than free trade?

May 4th-14th

[JOIN THIS LIVE DEBATE](#)



The Economist

### Related Items

#### In this special report

- Information has become superabundant
- Too much data
- Information is changing business
- How internet companies profit from online data
- Governments are becoming more open about data
- New ways of showing data
- The uses of information about information
- New regulations
- How machines deal with information
- Sources and acknowledgments
- Offer to readers

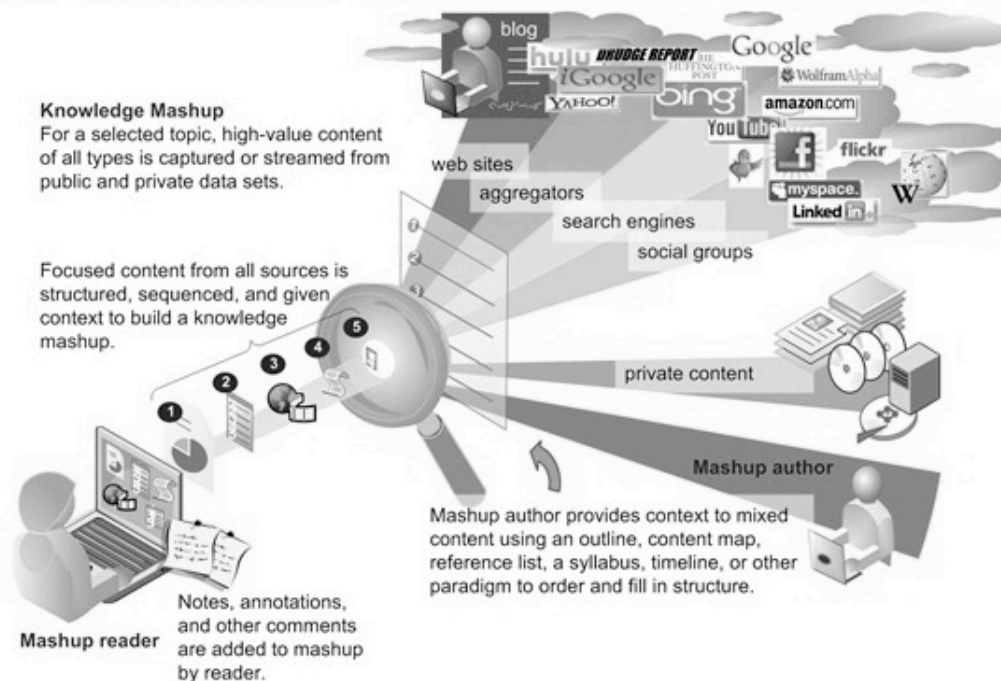
#### From *The Economist*

- Too much data

## The Knowledge Mashup

It's the world according to YOU. It's about you getting the information you want when you want it. It's about accessing content from open and collaborative sources, then filtering and focusing that content to meet unique documentation, training, and other educational needs. It's personalized, real-time information delivered directly to your computer devices. That's the promise of Web 2.0 moving to Web 3.0—automatically accessing and controlling the best information you deem as relevant to your needs. In your world, you are the one who provides context and meaning in the cloud of overwhelming data and disparate information. You are the existentialist of the Information Age who defines real knowledge amongst the chaos and chatter.

That's the idea behind the Knowledge Mashup. Find the best pieces of information, tag that information, and then structure and sequence it for specific needs. Use a simple object-oriented approach that identifies articles, videos, pictures, and people as independent objects and then give these objects context and navigation by overlaying them with some type of structural paradigm (a list, content map, outline, search filter).



The idea of a knowledge mashup is to bring together disparate but germane resources for a specific topic. The original author of knowledge mashup can tailor these resources and give context to educate, communicate, and impart information. It allows for usage of the best content that is controlled by an author or teacher in providing context and organization. It is a filter on the fire hose on the Internet to direct just the information needed for a specific product, process, instructional aid, or discipline. Among other benefits, it allows the author to be a "guide by the side" of the reader in finding information rather than the all-knowing "sage on the stage."<http://mashstream.com/>






**Business mashups enable individual efficiency and innovation**

Business mashups offer organizations a way to rapidly adapt to changing business needs. In a 2008 survey conducted by IBM, a majority of CEOs rated their organization's ability to manage change 22% lower than their expected need for change. Business mashups can give people the tools needed to adapt to change, develop new insights, and act on new business opportunities.

With IBM® Mashup Center, organizations can unlock and transform diverse sources of information into mashable assets such as feeds and widgets. These assets can then be dynamically assembled into new applications that address daily business challenges. Enabling business users in this way can reduce application backlog, lower development and increase the reuse of existing assets to reach more people cost-effectively.

- LEARN IT!** →
- SEE IT!** →
- TRY IT!** →
- BUY IT!** →

**We're here to help**



Easy ways to get the answers you need.

- Request a quote
- ✉ E-mail IBM

Or call us at:  
**877-426-3774**  
 Priority code:  
**109HH03W**

**IBM is the winner!**

VARBusiness has awarded a 2008 Tech Innovator Award to IBM Mashup Center.

→ Learn more

**News and announcements**  
→ New version of IBM Mashup Center announced

**Featured video**  
▶ Business mashups explained! Watch the animated video

**Featured webcast**  
→ Join us to learn about the new features planned for IBM Mashup Center v1.1

**Featured offer**  
→ Register for the IBM Mashup Center eKit

**IBM Mashup Center**  
→ Learn more about this easy-to-use business mashups platform

# World News Mashup



# MEDIA MASHUP

WITH **npr** AND *The New York Times*

OSCON · July 24, 2009 · San Jose, California

[http://alligatorfarm.files.wordpress.com/2009/07/sheep\\_clothing.jpg](http://alligatorfarm.files.wordpress.com/2009/07/sheep_clothing.jpg)



Sources: Denver Public Trustee, DenverGIS; Graphic by Brad Heath and Ron Coddington, USA TODAY

## Recent News

**Information Management:** Information Management talks about EMML's strength even in the ever changing IT industry.

**Open Mashup Alliance:** OMA releases whitepaper, 'Profitability, Predictability & Performance through Enterprise Mashups'

**Integration Developer News:** 'Open Mashup Alliance: New Language Will Boost Adoption'

**InfoWorld:** 'Led by JackBe, software makers form mashup alliance'

**ZDNet:** 'Creating a unified model for enterprise mashups'

**Application Development Trends:** 'Alliance Aims to Provide Mashup Interoperability'

**SD Times:** 'Mashup alliance seeks universal language'

**ZDNet:** 'Enterprise mashup proponents start organizing'

**SysCon:** 'Open Mashup Alliance Makes Debut'

**SD Times:** 'Will EMML be mother of all mashups?'

[what is a mashup?](#)

[why is there a need for a 'mashup language'?](#)

[what is the Open Mashup Alliance \(OMA\)?](#)

[why is EMML different from other languages?](#)

[why should my organization be interested in the OMA?](#)

[where can I learn more about enterprise mashups?](#)

---

## what is a mashup?

According to the [Wikipedia](#), a mashup 'combines data or functionality from two or more external sources to create a new service. The term mashup implies easy, fast integration, frequently using open APIs and data sources to produce results that were not the original reason for producing the raw source data. An example of a mashup is the use of cartographic data to add location information to real estate data, thereby creating a new and distinct Web services that was not originally provided by either source'. Practically speaking, Enterprise Mashups combine and remix data from databases, spreadsheets, websites, Web Services, RSS/Atom feeds, and unstructured sources that deliver actionable information for better decision-making.

## why is there a need for a 'mashup language'?

An open enterprise mashup markup language can greatly improve the portability of mashup designs and the interoperability of mashup solutions, thereby reducing the risk and cost to organizations that are adopting enterprise mashups.

A [Domain Specific Language \(DSL\)](#) designed specifically for mashups can also improve the quality of mashups of well by accomodating the important design characteristics that make mashups easier to create and reuse.

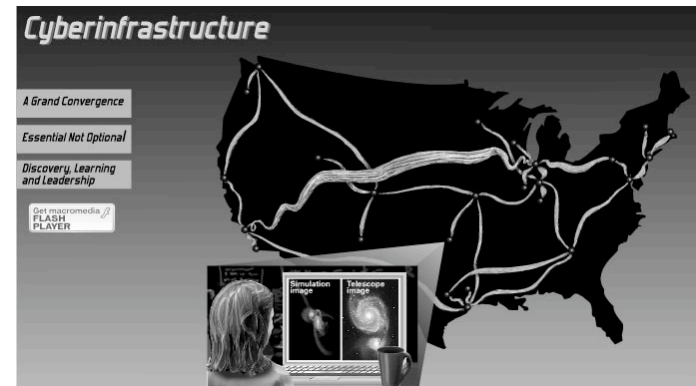
## what is the Open Mashup Alliance (OMA)?

The Open Mashup Alliance (OMA) is an organization chartered to promote the adoption of mashup solutions in the enterprise through the evolution of enterprise mashup standards like on open enterprise mashup markup language.

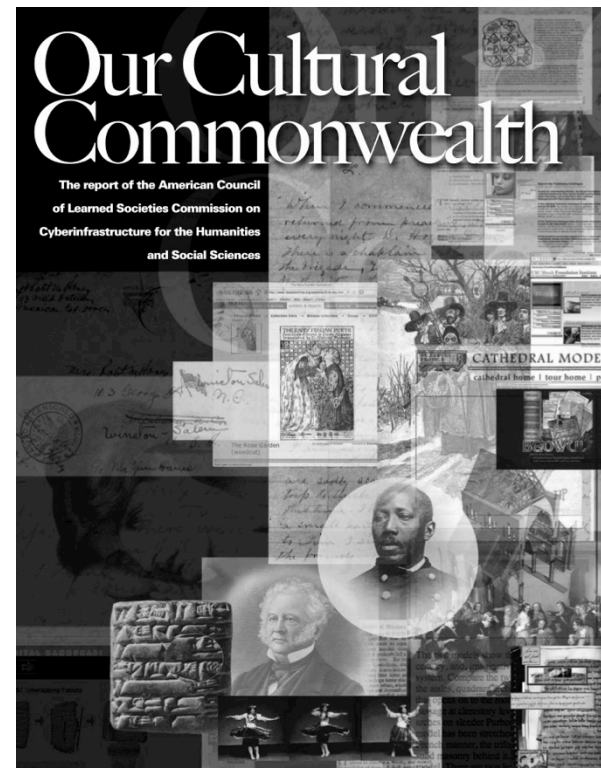
The OMA will evolve and promote an open schema and language for mashups called Enterprise Markup Mashup Language (EMML). EMML has been developed specifically to facilitate the creation of mashup services and mashup applications.

# Scholarly Information Infrastructure

- Cyberinfrastructure, eScience, eSocial Science, eHumanities, ...eResearch
- Goal: enable new forms of scholarship that are
  - information-intensive
  - data-intensive
  - distributed
  - collaborative
  - multi-disciplinary



[http://www.nsf.gov/news/special\\_reports/cyber/images/noflashintro.jpg](http://www.nsf.gov/news/special_reports/cyber/images/noflashintro.jpg)



# Asking Questions and Building a Research Agenda for Digital Scholarship

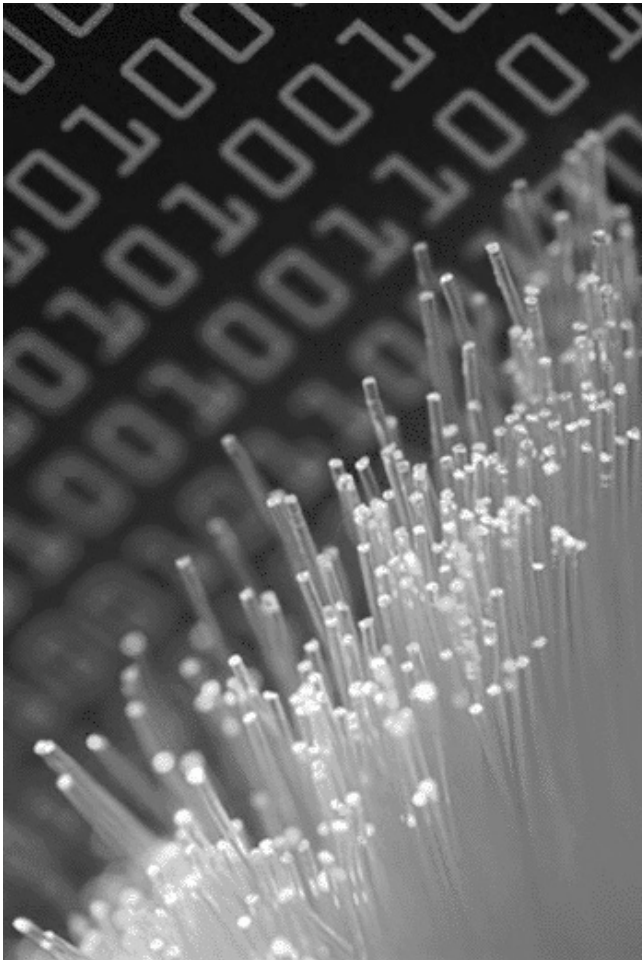
*Amy Friedlander*

---

- Scale
- Language and communication
- Space and time
- Social networking

Friedlander, A. (2009). Asking Questions and Building a Research Agenda for Digital Scholarship. In *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. Washington, DC, Council on Library and Information Resources. CLIR Publication No. 145: 1-15. <http://www.clir.org>

# What are data?



[http://datalib.ed.ac.uk/GRAPHICS/blue\\_data.gif](http://datalib.ed.ac.uk/GRAPHICS/blue_data.gif)

## Categories of data\*

- Observational
- Computational
- Experimental
- Records

## Objective or subjective?

- Facts
- “alleged evidence” (Buckland, 2006)

\*Long-Lived Data, NSF, 2005

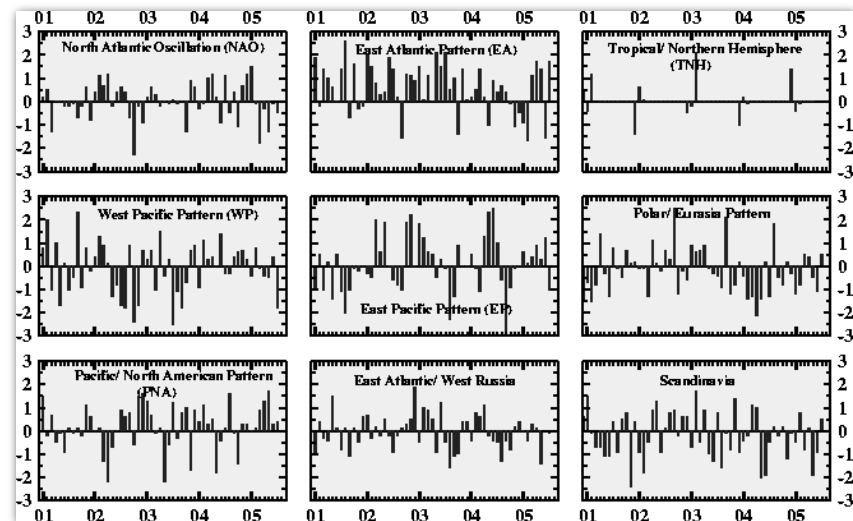
# Scientific Data

- Examples

- Ecology: weather, ground water, sensor readings, historical record
- Medicine: xrays
- Chemistry: protein structures
- Astronomy: spectral surveys
- Biology: specimens
- Physics: events, objects
- Documentation: Lab and field notebooks, spreadsheets

- Sources

- Generate own data
- Acquire from collaborators, other scientists
- Data repository



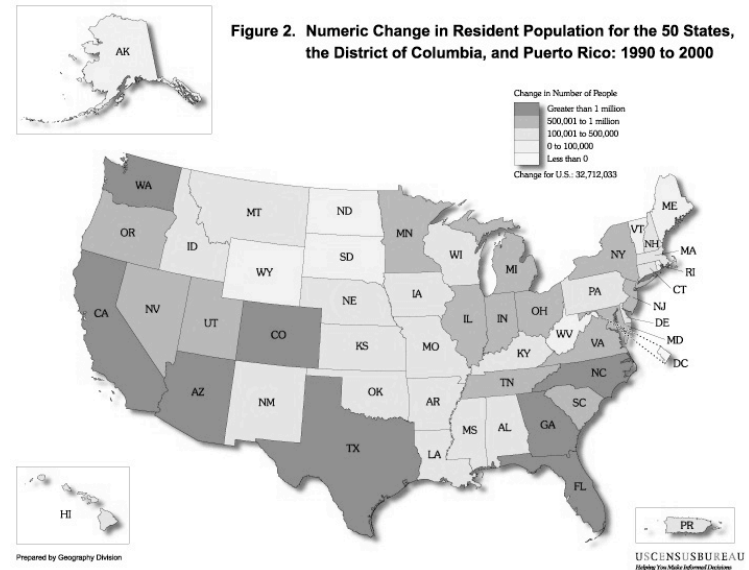
# Social Scientific Data

- Examples

- Opinion polls
- Surveys, interviews
- Mass media
- Laboratory experiments
- Field experiments
- Demographic records
- Census records
- Voting records
- Economic indicators

- Sources

- Generate own data
- Acquire from other scholars
- Data repositories: Social Surveys
- Government records
- Corporate records





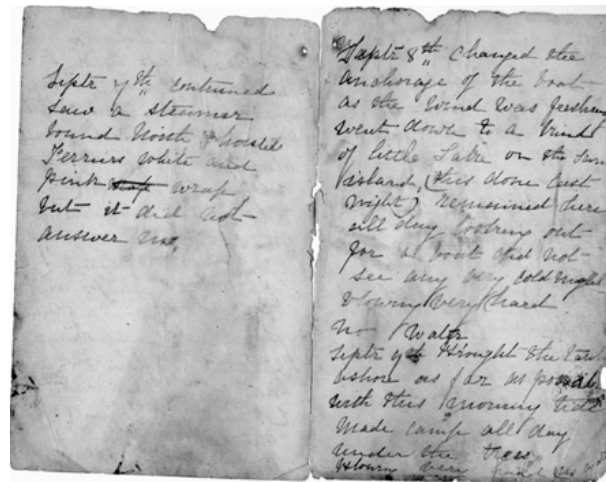
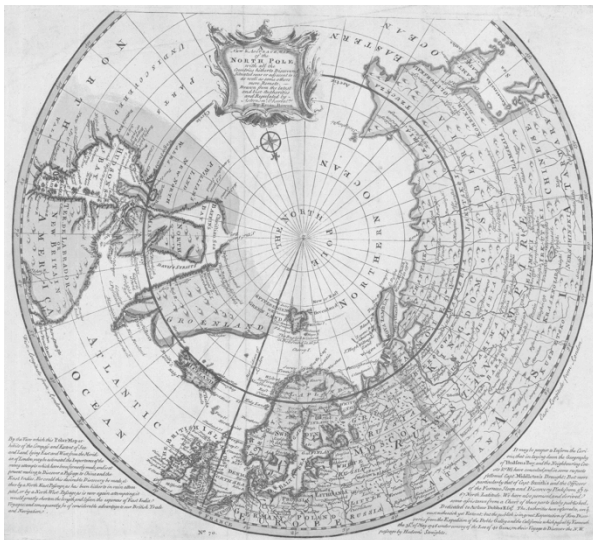
# Humanities and arts data

- Examples

- Newspapers
- Photographs
- Letters
- Diaries
- Books, articles
- Birth, death, marriage records
- Church records
- Court records
- School and college yearbooks
- Maps...

- Sources

- Libraries, archives, museums
- Public records
- Corporate records, mass media
- Acquire from other scholars
- Data repositories: Beazley, Arts & Humanities Data Service (UK)



[http://recollections.nma.gov.au/ejournal\\_library/images/vol\\_1\\_no\\_1/national\\_treasures\\_watson\\_diary\\_w450/files/13015/National%20Treasures%20-%20WatsonDiary.jpg](http://recollections.nma.gov.au/ejournal_library/images/vol_1_no_1/national_treasures_watson_diary_w450/files/13015/National%20Treasures%20-%20WatsonDiary.jpg)



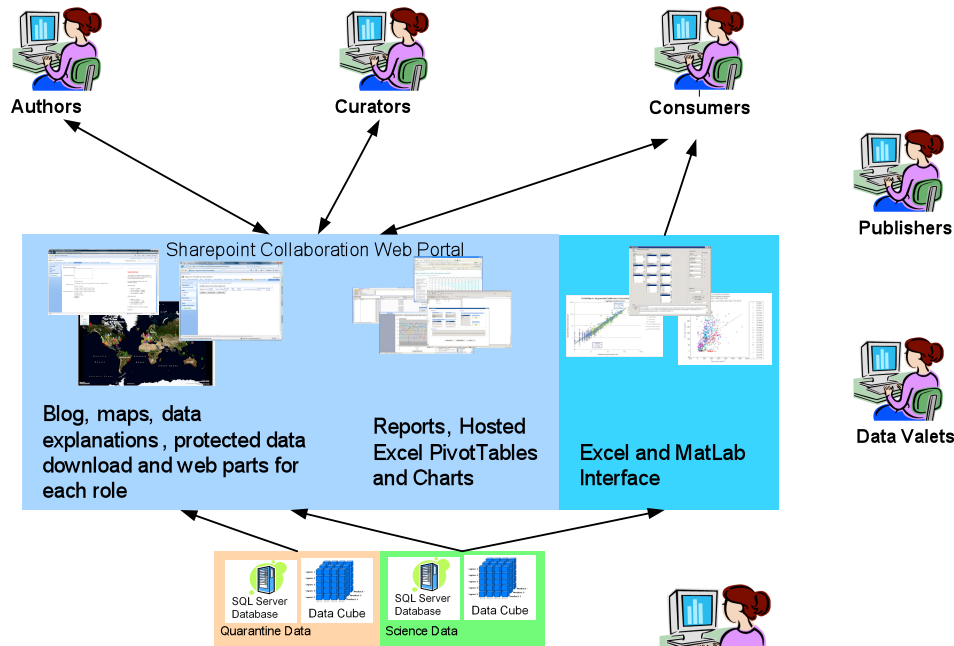
Rev. Charles Lutwidge Dodgson [Lewis Carroll]  
Xie Kitchin  
ca. 1875  
Albumen print  
15.7 x 12.7 cm

<http://www.hrc.utexas.edu/collections/photography/>

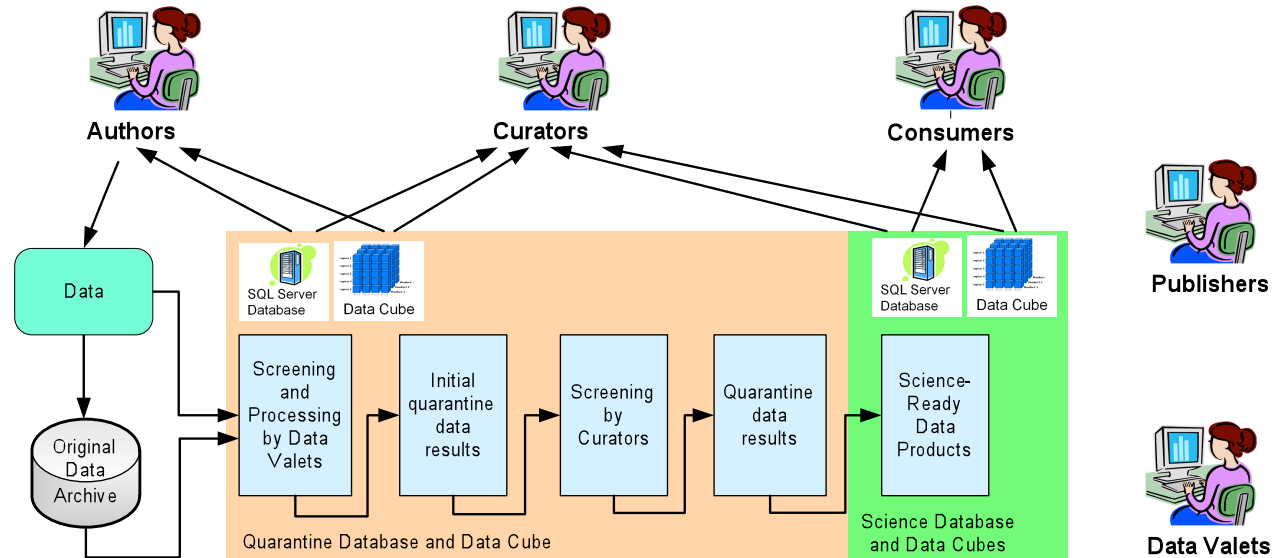
[http://libweb5.princeton.edu/visual\\_materials/maps/websites/northwest-passage/arctic1747-full.jpg](http://libweb5.princeton.edu/visual_materials/maps/websites/northwest-passage/arctic1747-full.jpg)

# Enabling Virtual Conversations

## Collaboration Centric View



## Data Centric View



Slide courtesy of Catherine van Ingen, Microsoft Research

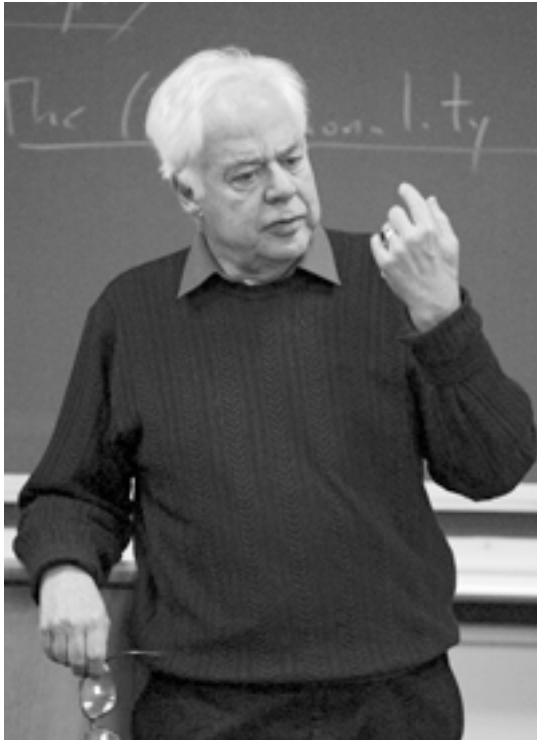


# Data access comparison\*

- Data-Centric Science
  - Data usage
    - Instant access
    - Robotic mining
    - Mashup with other sources
    - Multi-variate analyses
  - Tools
    - Open standards
    - Common platforms
    - Mining and visualization
  - Intellectual property
    - Open access license
    - Share and share alike
    - Derivative works encouraged
- Archives-Centric Humanities
  - Data usage
    - Delayed access
    - Manual mining
    - Inspect individual archive
    - Deep interpretation
  - Tools
    - Collection-level finding aids
    - Content specific methods
  - Intellectual property
    - Registered scholars
    - Licensed for local use
    - On-site access

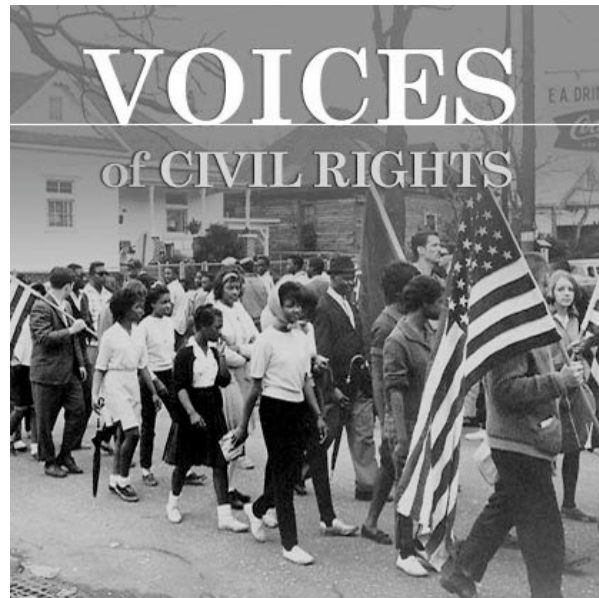
\*As noted in the discussion, this slide is a deliberately sharp contrast to make a point; the comparisons are, of course, nuanced and complex.

# Rights to Use and Reuse



# INTELLECTUAL PROPERTY:

**Rights & the Modern University:  
How to Move Forward into the 21st Century**



Who can read?  
Who can learn?  
Who can speak?

<http://interactive.usc.edu/members/sanderson/IPforum.jpg>

<http://www.loc.gov/exhibits/civilrights/images/civilrights-homeimage-previ.jpg>

I'm finding that something as simple as constructing my maps of related concepts are not easily applied to primary sources in digital libraries. So what use are the digital libraries, if all they do is put digitally unusable information on the web? The digital libraries don't offer a platform for traditional note taking, much less for larger scale analysis, either quantitative or qualitative.



UCLA doctoral student in history, personal communication, June 8, 2009 (used with permission)

# Summary

- The data deluge is real
- Data value lays in
  - Scale
  - Aggregation
  - Analytical tools
  - Distributed access
  - Standard platforms
- Data access depends on
  - Open tools and services
  - Intellectual property agreements



Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4). Retrieved from <http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html> on 14 April 2010.

Image: Christine L. Borgman, 1995 30



# Acknowledgements & Thanks



- 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; Thomas 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)

- Microsoft External Research



Special thanks for critical commentary on an earlier version of this talk

UCLA Department of Information Studies:

Evan Baker, David Fearon, Monica Garcia, Matthew Mayernik, Ellen Mulligan, Katie Shilton, Derek Quezada, Jillian Wallis, Laura Wynholds

And for advice on intellectual property issues for archives from Maureen Whalen, Getty Trust