#### UCLA Presentations

**Title** Drowning in the Data Deluge: Digital Library Challenges for Asia (Keynote)

#### Permalink

https://escholarship.org/uc/item/3824c6fx

**Author** Borgman, Christine L.

Publication Date 2011-10-25

#### **Copyright Information**

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

### Drowning in the Data Deluge: Digital Library Challenges for Asia



Christine L. Borgman Dineh Moghdam Patti K.Corbett SCHOLARLY COMMUNICATION AND BIBLIOMETRICS







Christine Borgman, Professor & Presidential Chair in Information Studies University of California, Los Angeles (UCLA) Keynote Presentation, International Conference on Asian Digital Libraries Beijing, 25 October 2011





## Digital libraries and data

- Thinking about infrastructure
- Infrastructure trends, 1970s-2010s
- Rethinking the infrastructure





Kyoto, 2009



## Digital libraries and data

• Thinking about infrastructure







Kyoto, 2009

## Cyberinfrastructure and eResearch

- Technology
- People
- Policy
- Scholarship
  - information-intensive
  - data-intensive
  - distributed
  - collaborative
  - multi-disciplinary



http://www.nsf.gov/news/special\_reports/cyber/images/noflashintro.jpg



Supercomputer the Lenovo Shenteng 7000 in the Computer Network Information Center: http://english.cas.cn/Re/IbW/





Star, S. L. & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. Information Systems Research, 7(1): 111-134.

Figure: Bowker, G. C., Baker, K., Millerand, F., Ribes, D., Hunsinger, J., Klastrup, L. & Allen, M. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In Hunsinger, J., Klastrup, L. & Allen, M. (Eds.). International Handbook of Internet Research. Dordrecht, Springer Netherlands: 97-117.



## Digital libraries and data

#### Infrastructure trends, 1970s-2010s



Ephesus, 2007





http://www.artsjournal.com/ bookdaddy/2008/07/

http://americanhistory.si.edu/collections/comphist/objects/ trs80l.jpg



http://blogs.sun.com/dennisding/entry/ sun open standards definition checklist





http://pingmag.jp/2007/03/23/ infosthetics-form-follows-data/



@姚晨

姆们都是有微博身份的

@羅培智SOLUO



@naijuy俞剑

@玄溟灏墨

@孟非



have a



### Infrastructure trends: 1970s-2010s

- I. Closed to open
- 2. Static to dynamic
- 3. Readers to authors
- 4. Publications to data



A Basic Text

Christine L. Borgman Dineh Moghdam Patti K.Corbett







## I. Closed to open: 1970s-1990s

- Closed network
  - Research community
  - Bibliographic services
  - NREN
- Closed standards
  - Cataloging rules
  - MARC formats



http://americanhistory.si.edu/collections/comphist/objects/ trs80l.jpg

Integrated library automation systems



## I. Closed to open: 1990s-

- Open network
  - Academic and research communities
  - Commercial and public services
  - Commodity internet
- Open standards
  - Internet protocols
  - WC3 standards
  - Operating systems



http://www.gsd.harvard.edu/inside/building\_services/Sustainability.html



## I. Closed to open: Today

- Mixed open and closed network
  - Academic community
  - Public / private services
  - Commodity internet
  - App world



http://blogs.sun.com/dennisding/entry/ sun\_open\_standards\_definition\_checklist

- Mixed open and closed standards
  - Internet protocols
  - WC3 standards
  - Operating systems
  - Application platforms



http://www.flickr.com/photos/gaetanlee/157465326/



### 2. Static to dynamic





## 2. Static to dynamic: 1970s-1990s

- Static content
  - Published documents
  - Databases grow incrementally
- Static context
  - Search results based on query
  - Same search yields same results
    - Anyone
    - Anywhere



http://www.artsjournal.com/ bookdaddy/2008/07/



# 2. Static to dynamic: 1990s-

- Dynamic content
  - Multiple versions of documents
  - Websites change constantly
- Dynamic context



http://www.flazh.de/en/404.htm

- Search results vary by profile and location
- Same search yields different results
  - Anyone
  - Anywhere

http://www.dlib.org/dlib/may10/xiaodong/05xiaodong.print.html



## 2. Static to dynamic: Today

- Content
  - Static
  - Dynamic
- Context
  - Static
  - Dynamic
- Continuity



## The Scholarly Journal, 1665-2011

TRANSACTIONS: GIVING SOME ACCOMPT

OF THE PRESENT Undertakings, Studies, and Labours

INGENIOUS

IN MANY CONSIDERABLE PARTS OF THE WORLD

Vol I. For Anno 1665, and 1666.

In the SAVOT, Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and Fames Allefry in Duck-Lane, Printers to the Royal Society.

#### THE ROYAL ISSN 0962-8436 SOCIETY IN THIS ISSUE The science of well-being: 29 Sentember 2004 volume 359 integrating neurobiology, psychology number 1449 pages 1329-1451 and social science Papers of a Discussion Meeting organized and edited by F. A. Huppert, N. Baylis and B. Keverne Philosophical Transactions of The Royal Society The world's longest running **Biological Sciences** international science journal

### 3. Readers to authors: 1970s-2000s

- Reader services
  - Libraries
  - Finding information
  - Local clientele
- Author services
  - Publishers
  - Producing information
  - Global clientele



Ephesus, 2007



## 3. Readers to authors: Today

- Author DIY
  - Copy editing
  - Publication-ready formatting
  - Negotiating copyright
  - Supplemental materials
  - Institutional deposit
  - Maintaining access to data...
- Universal authorship
  - Email, blogs, tweets, Facebook...



http://www.yiwu-marketguide.com/images/power-tools.jpg



http://www.guzer.com/pictures/suprise\_suprise.jpg

## 4. Publications to data: 1970s-1990s

- Publications
  - Types: Journals, books, papers...
  - Role: product at end of project
  - Social structure: Peer review, citation
- Research data
  - Types: heterogeneous
  - Role: process
  - Social structure: embedded in practice



## 4. Publications to data: 2000s-

- Publications
  - Types: Journals, books, papers...
  - Role: products throughout project
  - Social structure: Peer review, citation
  - Dissemination: libraries, blogs, tweets...
- Research data
  - Types: heterogeneous
  - Role: process and product
  - Social structure: embedded in practice
  - Dissemination: publications, websites, repositories



## 4. Publications to data: Today

- Publications
  - Peer review: Publishers
  - Cataloging: Libraries
  - Access: Libraries and search engines
  - Credit by citation: Authors
- Research data
  - Peer review: Journals? Repositories?
  - Metadata: Authors? Libraries? Repositorie
  - Access: Authors? Libraries? Repositories?
  - Credit by citation: Reusers of data?





## Digital libraries and data

• Rethinking the infrastructure





## Digital libraries and data: Grand challenges

- I. Take back information retrieval
- II. Engage the information life cycle
- III. Distribute the architecture
- N. Match policy to incentives



Kanazawa Airport, Japan, 2009

## I. Take back information retrieval

- Today
  - Generic search engines
  - Bibliographic services
  - Discrete data collections
- Tomorrow
  - Discoverability
    - Generic search
    - Specialized search
  - Organization and retrieval
    - Aboutness
    - Linking related objects
  - Continuity
    - Reproducibility
    - Trust

http://cof-interspiritual-mindfulness.blogspot.com/2011/07/ from-daily-om-most-fulfilling-road.html



http://kissaneasylum.typepad.com/workforce\_development/2009/



## II. Engage the life cycle

- Research design
- Data
- Metadata
- Writing
- Curation
- Use and reuse
- Reproducibility



Pepe, A., Mayernik, M. S., Borgman, C. L. & Van de Sompel, H. (2010). From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. Journal of the American Society for Information Science and Technology, 61(3): 567–582.



## III. Distribute the architecture

- Data scales fastest
- Select and filter
- Link related objects
- Use surrogates for discovery
- Move computation to the data
- Share access and assets

assets NET Services GetDocumentRequest GetDocumentRequest Cettoren Front End Services Event-Driven Front End Services

http://geekswithblogs.net/EltonStoneman/archive/2008/12/02/cloud-distributed-computing.aspx



## IV. Match policy to incentives

- Data curation: means or end?
  - Reuse
  - Reproducibility
- Data management is
  - Expensive
  - Poorly rewarded
  - Highly inconsistent
- Selection matters



London, 2005

Stewardship matters



## Digital libraries and data: Grand challenges

- I. Take back information retrieval
- II. Engage the information life cycle
- III. Distribute the architecture
- N. Match policy to incentives



Kanazawa Airport, Japan, 2009



### Digital libraries and data: Asian opportunities

- Internationalization of universities
- Cultural industry as economic pillar
- Collaborative research
- Entrepreneurship





solutionfocusedpolitics.wordpress.com



## Advice from Jim Gray

- Solve the problem of someone who is in pain
- May all your problems be technical



Feature Story

Microsoft Researcher Jim Gray Receives Turing Award for Helping to Transform Databases into Dynamic Tools Used by Millions

Gray's database research paved the way for ATM machines, computerized airline reservations and e-commerce. The Microsoft senior researcher this month received the prestigious A.M. Turing Award-the "Nobel Prize of computer science"-for his contributions.



## Acknowledgements

- CENS Data Practices team at UCLA David Fearon, Matthew Mayernik, Katie Shilton, Jillian Wallis, and Laura Wynholds; Alberto Pepe
- Research funding:
  - 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





