May 27, 2010

The Digital Future is Now: What the Humanities can Learn from eScience

Christine L Borgman, University of California, Los Angeles

Available at: https://works.bepress.com/borgman/237/
The Digital Future is Now: What the Humanities can Learn from eScience

Christine L. Borgman
Professor & Presidential Chair in Information Studies UCLA

Libraries in the Digital Age
Zadar, Croatia (by Videoconference)
27 May 2010

Slides posted under Creative Commons Attribution Non-Commercial License. Some rights reserved.
http://creativecommons.org/licenses/by-nc/3.0/
science

Wikipedia article on Science, generated in Wordle.net
Wikipedia article on Humanities, generated in Wordle.net
Digital Humanities is not a unified field but an array of convergent practices that explore a universe in which:

a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and

b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences.
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
We face a critical juncture. Leaving it to "them" is unfair, wrongheaded, and irresponsible. Them is us.
• Scholarly information infrastructure
• Science {and, or, vs} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
Infrastructure
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
• Scale
• Language and communication
• Space and time
• Social networking

• Scholarly information infrastructure
• Science {and, or, vs.} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
The Scholarly Journal, 1665-2010

PHILOSOPHICAL TRANSACTIONS: GIVING SOME ACCOMPT OF THE PRESENT Undertakings, Studies, and Labours OF THE INGENIOUS IN MANY CONSIDERABLE PARTS OF THE WORLD

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane;
Printers to the Royal Society.

The science of well-being: integrating neurobiology, psychology and social science

Papers of a Discussion Meeting organized and edited by F. A. Huppert, N. Baylis and B. Revenne

Philosophical Transactions of The Royal Society Biological Sciences

The world’s longest running international science journal
## Why do scholars publish?

<table>
<thead>
<tr>
<th>Function</th>
<th>Print</th>
<th>Digital</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Legitimization</strong></td>
<td>• Peer review</td>
<td>• Peer review</td>
</tr>
<tr>
<td>Authority, quality, priority,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trustworthiness</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dissemination</strong></td>
<td>• Publisher</td>
<td>• Publisher</td>
</tr>
<tr>
<td>Awareness, diffusion, publicity</td>
<td>• Pre-print distribution</td>
<td>• Pre-print distribution</td>
</tr>
<tr>
<td></td>
<td>• Copy</td>
<td>• Post on Web</td>
</tr>
<tr>
<td></td>
<td>• Mail</td>
<td>• Deposit</td>
</tr>
<tr>
<td><strong>Access, preservation, curation</strong></td>
<td>• Library</td>
<td>• Library</td>
</tr>
<tr>
<td>Availability, discovery, retrieval,</td>
<td></td>
<td>• Publisher</td>
</tr>
<tr>
<td>persistence</td>
<td></td>
<td>• Repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Homepage</td>
</tr>
</tbody>
</table>
Citation distribution of scientific literature

Graphs by Jillian Wallis, UCLA
Citation distribution of social scientific literature

# of Citations (half-life)

Date of Publication
Citation distribution of humanities literature
Physics

- Astrophysics (astro-ph new, recent, find)
  - Includes: Cosmology and Extragalactic Astrophysics; Earth and Planetary Astrophysics; Galaxy Astrophysics; High Energy Astrophysical Phenomena; Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- Condensed Matter (cond-mat new, recent, find)
  - Includes: Disordered Systems and Neural Networks; Materials Science; Mesoscale and Nanoscale Physics; Other Condensed Matter; Quantum Gases; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- General Relativity and Quantum Cosmology (gr-qc new, recent, find)
- High Energy Physics – Experiment (hep-ex new, recent, find)
- High Energy Physics – Lattice (hep-lat new, recent, find)
- High Energy Physics – Phenomenology (hep-ph new, recent, find)
- High Energy Physics – Theory (hep-th new, recent, find)
- Mathematical Physics (math-ph new, recent, find)
- Nuclear Experiment (nucl-ex new, recent, find)
- Nuclear Theory (nucl-th new, recent, find)
- Physics (physics new, recent, find)
  - Includes: Accelerator Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and Molecular Clusters; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics Education; Physics and Society; Plasma Physics; Popular Physics; Space Physics
- Quantum Physics (quant-ph new, recent, find)

Mathematics

- Mathematics (math new, recent, find)
  - Includes (see detailed description): Algebraic Geometry; Algebraic Topology; Analysis of PDEs; Category Theory; Classical Analysis and ODEs; Combinatorics; Commutative Algebra; Complex Variables; Differential Geometry; Dynamical Systems; Functional Analysis; General Mathematics; General Topology; Geometric Topology; Group Theory; History and Overview; Information Theory; K-Theory and Homology; Logic; Mathematical Physics; Metric Geometry; Number Theory; Numerical Analysis; Operator Algebras; Optimization and Control; Probability; Quantum Algebra; Representation Theory; Rings and Algebras; Spectral Theory; Statistics; Symplectic Geometry

Nonlinear Sciences

- Nonlinear Sciences (nlin new, recent, find)
  - Includes (see detailed description): Adaptation and Self–Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons
arXiv monthly submission rate statistics

Monthly Submission RATE for arxiv.org

First 17.9 years (17 Jun '89 total = 543,887)

Blue – Number of new submissions received during each month since Aug '91

Statistics of submissions to and use of arXiv:

- monthly submission totals
- monthly submission totals sorted by subject areas (updated 31 Dec '08)
- today's usage (this server only)
### arXiv web server usage for 18 Jun '09 (arxiv.org site only)

<table>
<thead>
<tr>
<th>Hour</th>
<th>Total Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>48509</td>
</tr>
<tr>
<td>01</td>
<td>49631</td>
</tr>
<tr>
<td>02</td>
<td>59139</td>
</tr>
<tr>
<td>03</td>
<td>70031</td>
</tr>
<tr>
<td>04</td>
<td>65818</td>
</tr>
<tr>
<td>05</td>
<td>64423</td>
</tr>
<tr>
<td>06</td>
<td>59854</td>
</tr>
<tr>
<td>07</td>
<td>60020</td>
</tr>
<tr>
<td>08</td>
<td>64486</td>
</tr>
<tr>
<td>09</td>
<td>66495</td>
</tr>
<tr>
<td>10</td>
<td>68302</td>
</tr>
<tr>
<td>11</td>
<td>62572</td>
</tr>
<tr>
<td>12</td>
<td>55418</td>
</tr>
<tr>
<td>13</td>
<td>50648</td>
</tr>
<tr>
<td>14</td>
<td>5252</td>
</tr>
<tr>
<td>15</td>
<td>47325</td>
</tr>
<tr>
<td>16</td>
<td>47325</td>
</tr>
<tr>
<td>17</td>
<td>45848</td>
</tr>
<tr>
<td>18</td>
<td>18281</td>
</tr>
<tr>
<td>19</td>
<td>1069747</td>
</tr>
</tbody>
</table>

Total number of connections = 1069747 (+93 local & administrative connections)

Current local time is Thu, 18 Jun 09 19:38:43 EDT

Statistics of submissions to and use of arXiv:

- monthly submission totals
- monthly submission totals sorted by subject areas (updated 31 Dec '08)
- today's usage (this server only)
• Scholarly information infrastructure
• Science {and, or, vs} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives
• Learning

• Call to Action
Data in digital scholarship

- Scholarly capital
  - Human capital
  - Instrumentation
  - Data
- Leverage research investments
  - Replicate, verify research findings
  - Ask new questions with extant data
    - Computational biology, chemistry
    - Longitudinal and comparative research
    - Mining large bodies of text and observations
- Collaborative research
  - Data creation
  - Data sharing, reuse

http://datalib.ed.ac.uk/GRAPHICS/blue_data.gif
What are data?

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: x-rays
- 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

http://www.census.gov/population/cen2000/map02.gif
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)
The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson  06.23.08
• Scholarly information infrastructure
• Science {and, or, vs} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
Research laboratories

http://dcc2.bumc.bu.edu/gcrcweb/T%2010c-%20Core%20laboratory.JPG

http://www.americaslibrary.gov/assets/aa/edison/aa_edison_subj_e.jpg

http://www.una.edu/compcenter/img/computerlab.jpg

http://student.some.ox.ac.uk/mcr/images/library%20interior.jpg
“Applied computer science is now playing the role that mathematics did from the 17th through the 20th centuries: providing an orderly, formal framework and exploratory apparatus for other sciences”

– G. Djorgovski
## Scientific and Technical Publications

Scientific papers are based on analyses of, or presentations of, the SDSS data. Data Release papers describe the specific process for each data release. Technical papers describe the SDSS instrumentation, calibration, software, strategy, and targeting algorithms. Technical papers may include some SDSS data for illustrative purposes. This list represents the definitive list of SDSS papers submitted to peer-reviewed journals. Other Publications Based on SDSS Data is a list of publications in journals and astro-ph which use public SDSS data.

### Scientific Publications

<table>
<thead>
<tr>
<th>Title</th>
<th>First Author</th>
<th>astro-ph</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>A MaxBCG Catalog of 12,875 Galaxy Clusters from the Sloan Digital Sky Survey</td>
<td>B. Koester</td>
<td>0701268</td>
<td>ApJ accepted</td>
</tr>
<tr>
<td>Luminosity dependence of the spatial and velocity distributions of galaxies: Semi-analytic models versus the Sloan Digital Sky Survey</td>
<td>C. Li</td>
<td>0701218</td>
<td>MNRAS submitted</td>
</tr>
<tr>
<td>Clustering Analyses of 300,000 Photometrically Classified Quasars--II. The Excess on Very Small Scales</td>
<td>A. Myers</td>
<td>0612191</td>
<td>ApJ accepted</td>
</tr>
<tr>
<td>Clustering Analyses of 300,000 Photometrically Classified Quasars--I. Luminosity and Redshift Evolution in Quasar Bias</td>
<td>A. Myers</td>
<td>0612190</td>
<td>ApJ accepted</td>
</tr>
<tr>
<td>The UV Properties of SDSS Selected Quasars</td>
<td>G. Trammell</td>
<td>0611549</td>
<td>AJ accepted</td>
</tr>
<tr>
<td>The Peculiar SN 2003hk: Do Some Type Ia Supernovae Explode as Deflagrations?</td>
<td>M. M. Phillips</td>
<td>0611295</td>
<td>PASP submitted</td>
</tr>
<tr>
<td>Broad Absorption Line Variability in Repeat Quasar Observations from the SDSS</td>
<td>B. Lundgren</td>
<td>0610658</td>
<td>ApJ submitted</td>
</tr>
<tr>
<td>3.6-7.9 um Photometry of L and T Dwarfs and the Prevalence of Vertical Mixing in their Atmospheres</td>
<td>S. Leggett</td>
<td>0610214</td>
<td>ApJ accepted</td>
</tr>
<tr>
<td>Galaxy Colour, Morphology, and Environment in the Sloan Digital Sky Survey</td>
<td>N. Ball</td>
<td>0810171</td>
<td>MNRAS submitted</td>
</tr>
<tr>
<td>A New Survey for Giant Arcs</td>
<td>J. Hennawi</td>
<td>0610061</td>
<td>AJ submitted</td>
</tr>
<tr>
<td>Using the Galactic Dynamics of M7 Dwarfs to Infer the Evolution of their Magnetic Activity</td>
<td>A. West</td>
<td>0809001</td>
<td>AJ 132:2507 (2006)</td>
</tr>
</tbody>
</table>
• Scholarly information infrastructure
• Science {and, or, vs} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
collaboration is everything
An Ecology of CENS Data

CENS Data Practices Research Group:
Christine Borgman, Jillian Wallis, Alberto Pepe, Matthew Mayernik, Andrew Lau, David Fearon, Katie Shilton
Sharing CENS research data

• **Research questions:**
  • What are CENS data?
  • When, how, and with whom will they share data?
  • What contextual information is necessary to interpret the data?
  • What resources exist to provide metadata?

• **Application of results:**
  • Architecture to capture, manage, and provide access to CENS data
  • Leverage data resources for research and learning
“Temperature is temperature.”

“There are hundreds of ways to measure temperature. ‘The temperature is 98’ is low-value compared to, ‘the temperature of the surface, measured by the infrared thermopile, model number XYZ, is 98.’ That means it is measuring a proxy for a temperature, rather than being in contact with a probe, and it is measuring from a distance. The accuracy is plus or minus .05 of a degree. I [also] want to know that it was taken outside versus inside a controlled environment, how long it had been in place, and the last time it was calibrated, which might tell me whether it has drifted.."
What are CENS Data?

Sensor Collected Application Data
- Roll/pitch/yaw
- Rudder angle
- Motor speed
- Heading
- Wind duration
- Wind direction
- Wind speed
- Soil moisture
- LandSat images
- Rainfall
- ORP
- Calcium
- Ammonium
- CO2
- Temperature
- pH
- Mercury
- Organism presence
- Organism concentration
- Nutrient concentration

Sensor Collected Proprioceptive Data
- Water depth
- Ammonium
- Phosphate
- Water temp
- Chlorophyll
- Chloride
- Nutrient concentration

Sensor Collected Performance Data
- Awake time
- Fault detection
- Neighbor table
- Packets transmitted
- Packets received
- Battery voltage
- Routing table

Hand Collected Application Data
- GPS/location
- Mosscam
- CDOM
- Time
- Water depth
- PAR
- Flow
- Bird calls

Figure by Jillian Wallis, UCLA
Artifacts and the Scientific Life Cycle

Datasets

Publications

Preservation

Experiment Design

Deployment Plans

Analysis

Calibration

Capture

Integration

Deviation

Figure by Jillian Wallis, UCLA
• Scholarly information infrastructure
• Science {and, or, vs} humanities

http://therange.files.wordpress.com/2008/12/artvsscience.jpg

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
Scholars’ incentives to share

- Open science, scholarship
- Recognition
- Collaboration
- Reciprocity
- Coercion

Image source:
www.buffaloworks.us/images/sharing%20orangs.jpg
Incentives *not* to share

- Rewards for publication
- Effort to document data
- Competition, priority of claims
- Intellectual property
  - Control over data and sources
  - Access to data and sources
  - Definitions of data

Image source: www.buildingsrus.co.uk/.../target1.htm
Why openness matters

• Interoperability trumps all
  – Import and export in open formats
  – Mixup and mashup
  – Add value
  – Avoid lock in

• Discoverability of related
  – Documents
  – Data
  – Assorted digital objects

• Usability and reusability
  – For research
  – For learning

http://pzwart.wdka.hro.nl/mdr/research/lliang/mdr/mdr_images/opencontent.jpg/
• Scholarly information infrastructure
• Science {and, or, vs} humanities

• Publication practices
• Data
• Research methods
• Collaboration
• Incentives

• Call to Action
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
Call to Action: Lessons from Science

- Publication practices: *Increase speed and scope of dissemination through online publishing and open access*
- Data: *Define, capture, manage, share, and reuse data*
- Research methods: *Adapt practices to ask new questions, at scale, with a deluge of data*
- Collaboration: *Find partners whose expertise complements yours, listen closely, and learn*
- Incentives: *Identify best practices for documenting, sharing, and licensing humanities content*
- Generally: *Err toward openness, reusability, and generalizability*
Call to Action: Research questions / problems

• What are data?
• What are the infrastructure requirements?
• Where are the social studies of digital humanities?
• What is the humanities laboratory of the 21st century?
• What is the value proposition for digital humanities in an era of declining budgets?

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
  • Murtha Baca, Getty Research Institute
  • William Dutton, Oxford Internet Institute
  • Johanna Drucker, UCLA-Info Studies
  • Bernard Frischer, U Virginia, IATH
  • Amy Friedlander, CLIR
  • Richard Furuta, Texas A&M, Computer Sci
  • Kimberly Garmoe, UCLA-History
  • Anne Gilliland, UCLA-Info Studies
  • Jason Hewitt, UCLA-Info Studies
  • Marina Jirotka, Oxford University
  • Clifford Lynch, Coalition for Networked Info
  • Ann O’Brien, Loughborough Univ, Info Sci
  • Susan Parker, UCLA Library
  • Todd Presner, UCLA Dig Humanities
  • Allen Renear, UIUC-GSLIS
  • David Robey, Oxford University
  • Harold Short, King’s College, London
  • Paul Spence, King’s College, London
  • Joshua Sternfeld, UCLA, Info Studies
  • Sarah Thomas, Bodleian Library
  • Sharon Traweek, UCLA, History
  • Anne Trefethen, Oxford eResearch
  • John Unsworth, UIUC-GSLIS
  • Sarah Watstein, UCLA
  • Maureen Whalen, Getty Trust
  • Robert Winter, UCLA

UCLA Dress Rehearsal
Tony Aponte, Evan Baker, Cheryl Bartel, Marta Brunner, Sharon Farb, Bob Freel, Kimberly Garmoe, Claire Hamasu, Jason Hewitt, Lisa Kemp Jones, Patrick Keilty, Julie Kwan, Andrea Lynch, Matthew Mayernik, Kathryn Mika, Sylva Manoogian, Rikke Ogawa, Ryoko Onishi, Sonja Palfner (Darmstadt), Susan Parker, Ellen Pearlstein, Katie Shilton, Jillian Wallis, Elizabeth Waraksa, Sarah Watstein, Neal Williams, Maureen Whalen (Getty Trust)