

This paper was presented at the conference:

## **Oxford eResearch 2008**

11-13 September 2008

University of Oxford

Conference website:

<http://www.oii.ox.ac.uk/microsites/eresearch08/index.cfm>

Conference papers collection:

<http://ora.ouls.ox.ac.uk/objects/uuid%3A64aa6f39-7e81-4d42-a008-ee2d7524bd67>

Conference organisers:

Oxford Internet Institute

University of Oxford

1 St Giles, Oxford OX1 3JS

<http://www.oii.ox.ac.uk/>

email: [enquiries@oii.ox.ac.uk](mailto:enquiries@oii.ox.ac.uk)

Oxford e-Research Centre

University of Oxford

7 Keble Road, Oxford OX1 3QG

<http://www.oerc.ox.ac.uk/>



## Will the Data Deluge Improve or Impair the Quality of Scholarship?

Christine L. Borgman  
Professor & Presidential Chair in Information Studies  
University of California, Los Angeles

Keynote Presentation to Oxford eResearch 08  
*Keynote Panel on e-Research, Virtual Witnessing, and the Quality of Scholarship*  
Oxford Internet Institute, Oxford, UK, 12 September 2008

The “data deluge” started in the sciences, with highly instrumented research in physics, astronomy, chemistry, and molecular biology, and has spread to most other fields. Embedded sensor networks are now essential technologies for the environmental sciences. Internet communication provides rich traces of human activity for social scientists to study. Educators are logging interactions with simulated experiments, collaborative tools, and embedded assessments. Scholars in the humanities, not to be outdone, are among those mining texts and modeling cities.

While the data deluge presents immense opportunities for eResearch, it is also the area of eScholarship most subject to hype. *Wired Magazine* recently pronounced that science no longer needs theory, models, metadata, ontologies, or “the scientific method” – mining the data deluge replaces all of them [1]. That glib claim obscures the epistemological problems of ascertaining “facts” and classifying them in appropriate cognitive and cultural structures. It also obscures the challenges of distributing trust in digital environments. Making sense of long strings of bits is difficult enough in the few fields that have agreed upon data structures, metadata, and ontologies. In most fields, individual investigators still manage their data in spreadsheets or databases of local origin, with names of data elements and variables assigned anew for each research project and instrument. Essential information about instrument calibrations, field conditions, and data transformations may be stored elsewhere, if documented at all [2].

Although many funding agencies now require data management plans and the deposit of data<sup>i</sup>, much remains to be negotiated about definitions and about the scope of data to be managed, preserved, and shared – Raw, processed, published data? Physical specimens? Codebooks? Interview instruments? Field notebooks? Blogs and wikis describing the research process? In what condition must data be released, in what time frame (6 months, 2 years, after publication)? How long must data be kept (until the end of the project, 5 years, 10 years, forever)? What requirements for de-identification of human subjects apply? In the few fields for which data repositories exist, such as genomics, seismology, and social surveys, a plan for depositing data there at the end of a project may be sufficient. In most fields, data are likely to remain the responsibility of the investigator or the campus. If these opportunities and challenges can be addressed adequately, the data deluge may improve scholarship. If researchers are coerced into depositing data without adequate assurances of data integrity, interpretability, curation, and trust, then the data deluge may only impair the quality of scholarship.

References:

1. Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Retrieved from [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) on 22 July 2008.
2. Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.

---

<sup>i</sup> Enforcement of data release rules often is inconsistent, however.