OPINION

Whose Role Is It Anyway?: A Library Practitioner's Appraisal of the Digital Data Deluge

by Marisa L. Ramírez

Research Data Access and Preservation

Editor's Note:
Though this opinion piece is not drawn from the recent RDAP Summit, to which the Special Section in this issue is devoted, it does add to the discussion on the topic.

EDITOR'S SUMMARY

As the volume of digital data multiplies exponentially and the use of digital repositories to capture academic research expands, the demands on academic librarians are also increasing. Librarians are expected to serve as liaisons between data authors, managers, scientists and end users, while providing a full range of curation services. Little has been offered from the perspective of archival and records management, despite archivists' traditional role as keepers and stewards of scholars' data. Archival science focuses on appraising, selecting and describing data, managing data retention and attending to source, authenticity and preservation. Professional archivists have considerable expertise in handling volumes of research data, and archival methods can add efficiency to digital data management. Greater collaboration between academic library liaisons and archivists is urged, recognizing and integrating the skills of each profession to best advantage for the most effective approach to comprehensive data curation and management of digital repositories.

KEYWORDS

archival science
digital repositories
archivists
academic libraries
professional competencies

Marisa L. Ramírez is the digital repository librarian at California Polytechnic State University (Cal Poly), San Luis Obispo. She can be reached at mramir14<at>calpoly.edu.

The digital data universe is predicted to surpass 1,800 exabytes by the end of 2011, due in part to the increasing affordability of powerful devices designed to create, capture and store digital data. At this rate, the amount of digital data is predicted to eclipse Avogadro's number by the year 2026 [1]. The identification, collection and preservation of digital data created as a result of research is an important issue, particularly because the sharing and reuse of raw research outputs offer great potential for subsequent recombination, analysis, insight and discovery.

The capture and curation of these resources present many challenges to librarian practitioners. Some of the most salient include the following:

- **Appraisal and Selection**
  - Deep disciplinary knowledge is needed to appraise data.
  - Manually appraising data sets is very time consuming and expensive, and automated approaches are in their infancy.
- **Retention**
  - It is unclear what criteria should be used to determine how long research data should be kept.
- **Description**
  - Specialized knowledge, particularly from the data author, is required for the creation and application of ontologies and metadata.
- **Preservation**
  - Research data sets can be complex and dynamic, relying on integration with software and associated visualizations.

- Prioritization and stewardship of extensive and diverse digital assets can be difficult, given that computational elements and outputs are frequently heterogeneous.

  ■ **Authenticity**
  - Data provenance – the tracking of all context and transformations the data has gone through – is key to verifying the authenticity and reliability of data files.

  ■ **Compliance**
  - It is not yet known how best to track and apply regulations, policies or protocols that govern the retention, access and reporting of digital data assets.

The Role of Libraries and Librarians

The library has a role to play in data management, in the "collection, organization, description, curation, archiving and disseminating of scholarship" [2, p.5] in the digital realm. As such, many academic research libraries are creating and using distributed repository systems to "support access to digital objects of e-research," ranging from documents to data sets and many other types of items.

Moreover, the academic librarian is identified as a key agent involved in the stewarding of research data. Recent information science literature encourages academic library practitioners, particularly librarian liaisons, to "skill up" for new roles to support the complex scientific systems and research protocols. These new roles include the data authors (the scientists and students that produce digital data), data managers (a partner in the data curation processes), data scientists (primarily comprising computer scientists and software engineers, as well as librarians) and data users who are in the larger academic and education communities. The librarian liaison is identified as integral to effective data curation activities and is described as the "best-qualified set of staff for … data set collecting work because of their relationships with faculty, departments and research centers across campus,"engaged in "data identification, mediation, selection and appraisal, and preparation." [3, p, 59]

Whose Role Is It, Anyway?

Regrettably, recent discourse on this topic neglects to address the value that other traditions of information practice, such as archives and records management, can offer to advance the discussion on the curation of digital data. It has been archivists, not librarians, who historically have served as "keepers of the record," seeking to balance the stewardship and protection of collections with the pragmatics of managing an ever-growing corpus of paper and electronic information. Librarians would be best served to embrace input from archivists to remain relevant and vital to scholars' data stewardship practice.

The archival methodology is steeped in a rich tradition of curatorial activities, with a particular focus on appraisal, selection, description and retention of content. These practices have value in the digital domain and should be consulted when dealing with hairy data management issues.

Archival studies, at their core, are interdisciplinary in nature, incorporating ideas from science and the humanities, which are adapted to the archival profession's needs. As such, archivists are well positioned to advise on the alignment of curatorial practices across and within disciplines. They are not strangers to navigating the complex legal, cultural and political waters and can provide a welcome perspective on addressing compliance issues for digital data, particularly serving in an advisory role when questions of ethics and legality are called into play.

Archivists contend with curatorial challenges of preserving heterogeneous, mutable and manipulatable electronic files developed by university office workers. Because digital information is no longer in a fixed medium, but instead can be a dynamic interrelated set of discrete files, archivists are confronting challenges of provenance, authenticity and preservation of e-records so that the user can view the data as the original user saw it. Archival theory and practice can help address questions surrounding the original-view, program emulation and preservation of e-research.

Archivists understand how backlogs of unprocessed materials result in the physical and intellectual inaccessibility of information. A progressive archival strategy to address backlogs was introduced by Green and Meissner [4], which emphasizes "more product, less process" to more efficiently

manage the description and organization of such collections, thus providing expedited intellectual access to information housed in university repositories. This approach can be applied to the digital data, particularly as the body of research data is projected to proliferate.

Incorporating perspectives from the archives, including related fields of records management and museum studies, into the data curation dialogue would not only be logical, but would also assist in ensuring sound basis for e-science curatorial activities. According to Newton et al [3] librarians must be able to articulate the value of curating data sets in repositories; they must understand system capabilities; and they must be able to cultivate relationships in order to become more in tune with the research that is underway by their faculty. These activities are the hallmark of an archivist's role within the information profession, so it logically follows that one would best confer and collaborate with archivists to tackle such challenges.

So Where from Here?

This is an exciting time for information professionals. We find ourselves at the nexus of information, technology and expertise, and this is an opening for us to introduce and apply our skills and knowledge to new domains.

Professionals in the archival field possess valuable expertise that can be leveraged in order to more effectively capture and preserve digital research outputs. Much of this expertise already exists within the university archives setting, despite recent literature leading us to believe otherwise.

Given their familiarity with information literacy and instruction, library liaisons may best serve an immediate need as data literacy advocates, providing their faculty constituents with information on granting agency requirements and raising awareness about the general data management issues.

Nonetheless, there needs to be an acknowledgement and integration of archival expertise into the broader data dialogue. There are some projects currently underway that aim to do just that. For example, University of Michigan's iSchool has developed an archives and records management curriculum that provides instruction on issues that bridge both the analog and digital realms and provides context for digital curation in the profession. There are also research efforts underway to unite digital practices, regardless of organizational context. Closing the Digital Curation Gap (CDCG) collaboration is an IMLS-funded project to develop shared curatorial best practices for professionals in such settings as libraries, archives, museums and other cultural institutions and information centers.

From a practitioner's point of view, our profession will have the opportunity to provide more robust, responsible and inspired curatorial services if an ecumenical approach is taken to address issues surrounding the management and preservation of digital content. This approach would integrate the most relevant theory across information disciplines, providing sound footing to support digital curatorial practice. Lest we forget, digital curation challenges are germane to all information professionals, and collectively we possess the expertise to ensure fragile digital data will be available for generations to come. ■

Resources Mentioned in the Article

[1] Ganz, J. (2008). *The diverse and exploding digital universe.* Framingham, MA: International Data Corporation. Retrieved April 4, 2011 from www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf.

[2] Jones, E. (2008). E-science talking points for ARL deans and directors. *Association of Research Libraries.* Retrieved April 4, 2011, from www.arl.org/bm~doc/e-science-talking-points.pdf.

[3] Newton, M.P., Miller, C.C., & Stowell Bracke, M. (2011). Librarian roles in institutional repository data set collecting: Outcomes of a research library task force. *Collection Management, 36,* 53-67.

[4] Greene, M.A., & Meissner, D. (2005). More product, less process: Revamping traditional archival processing. *American Archivist, 68,* 208-263.