The University of British Columbia Library
Data Management Working Group Report

August 2010

Executive summary

In January 2010, the Library formed a working group to consider and advise the eStrategy Committee on whether UBC Library should be involved in the process of data management, to investigate current data management practices and initiatives at UBC, to research current data initiatives in comparable university libraries and identify potential partners for data management initiatives. The Terms of Reference for the group is appended in Appendix 1.

The group was co-led by Eugene Barsky and Bronwen Sprout, and reported to the eStrategy committee. Nineteen library professionals volunteered to serve in this very diverse group, demonstrating a strong interest amongst UBC archivists and librarians about this topic. Group members conducted a literature review and interviewed UBC researchers and administrators to find out about current data management practices at UBC.

The group's findings demonstrate that there is an institutional need for data management both practically and as mandated by granting agencies, and suggest that the Library can position itself as a bridge between researchers and users by providing metadata and infrastructure support. In terms of the Library's efforts, data management aligns well with other digital initiatives such as cIRcle and the new Digital Initiatives unit, as outlined in the Library's strategic plan in the areas of *Accelerate Research* and *Engage with Community*. As our literature review revealed, data management is a relatively new area for academic libraries. There is an opportunity to establish ourselves as leaders in Canada in data management, and to collaborate nationally and internationally on data management issues. In light of the significant opportunities for the Library in this area, the group recommends that a limited term, internal position be created for a data management librarian to continue the group's work and to further university-wide engagement.

Background

Ubiquitous computing has had a significant impact on the process and products of scientific research. Instruments are gathering ever-increasing volumes of data, and new computing techniques are enabling researchers to query data in new ways. Right now in Canada, and in particular in UBC, the vast majority of research data is not being preserved. Currently, UBC Library does not have an active strategy for curating the data produced through research carried out at the institution. However, as part of its 2009 strategic plan, UBC has made a number of commitments to research excellence including "develop[ing] a campus strategy for making UBC research accessible in digital repositories, especially open access repositories." (University of British Columbia President's Office, 2009)

The Library's eStrategy committee established a working group to advise the eStrategy committee regarding research data curation at UBC, in particular as it relates to the potential role for UBC Library.

Literature review

The group focused on researching the data management requirements for major funding agencies, major journals and publishers, comparable research institutions, and COPPUL.

Major findings

- Academic libraries have begun to experiment in data curation and some best practices are beginning to emerge. Keys to success include:
    - Surveying researchers' data needs in advance of a curation project
    - Leveraging librarians' expertise to work with researchers to establish standards for metadata and preservation
    - Ensuring that cyber infrastructure is scalable and funded for the future
    - Considering where the data is best hosted. Many of our colleagues are using their IRs to store and curate their datasets and research.

- Major funding agencies in Canada, the US, the UK and Australia all require that projects must either provide a plan for data management as part of their funding applications, or agree to abide by the agency's pre-existing policy.
- Most journal publishers do not yet have a policy in place for raw data preservation – regarding this as the responsibility of authors. An exception is the biomedical publishers who are requiring raw data to be made available for other users.
- COPPUL is just beginning to examine this issue via the Digital Curation and Preservation Task Group. Bronwen Sprout is UBC Library's representative to the COPPUL Digital Preservation group.

Moreover, librarians possess a unique set of skills for metadata and data organization; however, we might lack other skills, e.g. computer science or infrastructure building background which will need to be acquired through academic or on-the-job training. The complete literature review is available in Appendix 2.

Interviews with Researchers

We interviewed several UBC researchers about their current practices and future needs regarding research data management. Our findings indicate that current storage and distribution capabilities vary widely between departments, with no consistent standards and practices in place. Some researchers are even utilizing resources at other universities to satisfy their needs for data computation and storage, as their needs are not being met through UBC. In none of these cases were metadata files being created; researchers typically do not have the time or inclination to produce such documentation on their own. One professor noted that graduate students would benefit greatly from a storage and management program for their research data. Such a program could also be used as a data management teaching tool for students. The set of questions posed to researchers is available in Appendix 3.

Interviews with IT managers

Due to the decentralized nature of IT management at UBC, a variety of administrators were approached to gain insight into common practices and potential areas that the library might address regarding data management. A presentation was made to UBC IT Managers, a formal committee representing disciplines across campus. A representative from the Office of Planning and Institutional Research and UBC Okanagan IT Research staff were interviewed. Preliminary findings include the following: data management at UBC is conducted at a departmental, research group, or individual researcher level; file management, metadata creation, and cataloguing standards are irregularly applied and there is appetite for the library to provide leadership in this area; privacy and security are major concerns for researchers and administrators with open sharing of data a tertiary concern; consortia/collaborative endeavours for data storage, such as WestGrid, exist and could provide either a model or the foundational infrastructure to build data services upon. The set of questions posed to IT managers is available in Appendix 4.

Recommendations

We recommend that the Library, working with the Director, Library Digital Initiatives, create a short-term (secondment) leadership opportunity for a UBC librarian or archivist to further explore the issues and follow up on the data management group. We recommend either a full time position for 6 months or 0.5 part time position for 12 months that will be backfilled. Specifically, we recommend this position be responsible for:

- Establishment of a pilot project, complete with a project charter to test the waters for a Library role in data management. We recommend a discipline-based approach to a pilot project to consider the differing data needs/practices amongst the various disciplines at UBC.

- Creation of training programs and/or study groups to enhance the knowledge base of a wider pool of staff members about metadata and preservation best practice

- Development of a toolkit for researchers to assist in their efforts at data management

- Continue to develop the connections with IT managers and explore opportunities with other University administrators to identify a possible role for the library in data management.

- Structurally, the newly formed UBC-wide Scholarly Communications Steering Committee may provide a venue for these activities. There is also a natural connection with cIRcle and the Digital Initiatives unit.