

**The University of British Columbia Library  
Data Management Working Group Report - Appendixes**

August 2010.

**Appendix 1 - Terms of Reference**

**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 being lost. One of the major stumbling blocks for this new e-Science is the need to find strategies for the long-term storage, preservation, and access to research data.

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)

**Purpose**

The Library's eStrategy committee has 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.

**Members:**

<ul style="list-style-type: none"><li>· Eugene Barsky (co-lead)</li><li>· Tom Brittnacher</li><li>· Hilde Colenbrander</li><li>· Margaret Friesen</li><li>· Kathy Hornby</li><li>· Paul Joseph</li><li>· Joy Kirchner</li><li>· Mary Luebbe</li><li>· Joerg Messer</li></ul>	<ul style="list-style-type: none"><li>· Katherine Miller</li><li>· Shawna Parlongo</li><li>· Sarah Romkey</li><li>· Barbara Sobol</li><li>· Bronwen Sprout (co-lead)</li><li>· Lea Starr</li><li>· Rudi Traichel</li><li>· Erwin Wodarczak</li></ul>
--	--

**Responsibilities**

To consider and advise the eStrategy committee on:

- o whether UBC Library needs to be involved in the process of data management
- o current data management practices and initiatives in UBC (Vancouver and Okanagan) – UBC IT, UBC Office of Research Services and others
- o current data initiatives in comparable university libraries and identify potential partners on campus and in other Canadian and international libraries for data management initiatives
- o possible pilot project for data management

and to provide a report to the eStrategy committee on its findings and recommendations as above.

### **Constitution**

The group's co-chairs will be Eugene Barsky and Bronwen Sprout. Membership will be made up of interested librarians and library staff.

### **Frequency of meetings**

The group will meet once a month for approximately six months.

## Appendix 2 - Literature Review:

After reviewing the professional literature it is clear that academic libraries do have a role to play in curating data generated by researchers at their various institutions. Over the years librarians have been instrumental in establishing and maintaining institutional repositories and social science data collections. While doing so, we have developed expertise in areas such as data selection, metadata standards, access and search functionality, managing the data lifecycle and digital preservation. We can and should leverage this knowledge to expand our services to include research data created by researchers at our own institutions.

The keys to success are:

- surveying local researchers' data preservation needs and practices
- establishing strong partnerships with researchers to establish metadata standards and preservation best practices
  
- ensuring that our cyberinfrastructure is both scalable and provided with long-term commitment and funding

### Selected Annotated Bibliography

Inge Angevaere. *Taking Care of Digital Collections and Data: 'Curation' and Organisational Choices for Research Libraries*. Vol.19, 2009.

Focuses on issues surrounding digital preservation and curation for research libraries. Stresses the fragility of digital data and the care and expense required to preserve it long-term. Suggests three options for libraries: Store data on existing cyber infrastructure and hope for the best; find a third party to host their data; build a proper digital archive in-house. Suggests that research libraries may revive their perceived value by curating researchers' digital datasets themselves. Ends with a do & don't list for research libraries' digital initiatives. Do: survey researchers' needs and practices; ensure your cyber infrastructure is scalable; establish selection criteria; provide researchers with metadata instruction and support; focus on access. Don't: Store data on local computers or outdated storage devices; neglect to back up data frequently; entrust your institution's data to a custodian without a proven track record.

Christine Borgman, Jillian Wallis and Noel Enyedy. *Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries*. Vol.7. Springer Science & Business Media B.V, 2007.

Addresses the "data deluge" threatening to overwhelm researchers in the sciences. Focuses on habitat ecology, an example of "little science." Details the value added services that libraries can provide to a data repository program aimed at little science: Establishing metadata standards, determining a standard communication framework for the exchange of metadata, developing technical tools to support/create metadata and providing the means to preserve metadata. Provides extensive analysis of the role of data in science and the ways in which researchers do or could re-use data. Conclusions: Scientists would ideally like to share

data but lack tools for data analysis; lack metadata standards; and most importantly, lack career incentives to preserve and share their data.

Rowan Brownlee. *Research Data and Repository Metadata: Policy and Technical Issues at the University of Sydney Library*. Vol.47, 2009.

The author describes recent efforts by the library at the University of Sydney to establish metadata standards for its research repository. Currently the repository houses three small, but significant data collections from the departments of geoscience, arts & archeology and biological sciences. Data are typically held in databases such as filemaker, MySQL or Excel. The primary challenge at USL has been to develop metadata standards that are sustainable and support findability, while also retaining the "granularity" of the depositor's supplied metadata. Details the institution's choice to apply Dublin core metadata standards to its data deposits - which have thus far been mapped from the depositor's supplied metadata by the appropriate subject librarian. The author suggests that this activity would ideally be mapped by a cataloguer-subject librarian-researcher team.

G. Sayeed Choudhury. *Case Study in Data Curation at Johns Hopkins University*. Vol.57, 2008.

Discusses institutional repositories (IR) and the gap between their promise and their reality - particularly focusing on the "crisis" in scholarly communication that only seems to exist for librarians. Posits that researchers are generally satisfied with the current scholarly publishing models and therefore have little motivation to support IRs and/or the open access movement. Suggests that IRs may be revitalized as data curation centers but concedes that the scale and complexity of "big science" renders its contributions unsuitable for a local repository. The author suggests IRs might instead choose to partner with like-minded institutions to share storage, preservation and dissemination services and/or limit submissions to select portions of a large project. Finishes by stressing the importance of using open source software platforms.

Tracy Gabridge. *The Last Mile: Liaison Roles in Curating Science and Engineering Research Data*. Association of Research Libraries, 2009.

Focuses on the librarian's role in creating valuable data services; soliciting contributions to library digital repositories and curating data deposits over the long term. Points out the expertise that already exists in terms of social science data services - though to date this has largely been comprised of datasets from external agencies. Argues that librarians could leverage this knowledge to expand services to include faculty-generated data. Identifies key skills librarians have to apply to curation duties: Selection, planning data life cycle; teaching preservation best-practices; collecting and disseminating the data appropriately; help establish/maintain metadata and data standards.

Anna Gold. *Cyberinfrastructure, Data, and Libraries, Part 2*. Vol.13, 2007.

Describes traditional types of data curated by data librarians: Social science, GIS and bioinformatic data. Suggests that librarians can leverage this experience and expertise to extend services to include scientific and technical data. Discusses potential roles that

librarians could play in supporting e-science: Selecting & acquiring datasets; offering long-term data preservation services; and helping develop best-practices to assure data longevity. stresses that librarians' skill-set will require that they limit their reach to small and medium-size research and that subject specialization may be required by librarians to assist researchers most effectively. Cautions that data services cannot expand without long-term commitment to staffing and funding.

Stuart MacDonald and Luis Martinez Uribe. *Libraries in the Converging Worlds of Open Data, E-Research, and Web 2.0*. Vol.32. Information Today Inc, 2008.

Considers the ways in which the ever increasing amounts of data being created will affect both the research lifecycle and the librarians who will be expected to support this e-research. Highlights the benefits of open access to research data and details tools, technologies and organizations which were key to the open data movement at the time this article was published. Discusses possible strategies that librarians might employ to support e-research, such as providing data curation services via institutional repositories.

Michael Witt. *Institutional Repositories and Research Data Curation in a Distributed Environment*. Vol.57, 2008.

Describes the challenges that arise from the "data deluge" generated by e-science: Masses of unorganized data are created, used once, preserved locally if at all and eventually lost or discarded. Discusses traditional roles librarians have played in disseminating research and suggests roles librarians might have in curating research data, such as creating metadata, providing access to datasets, appraisal of submissions, and providing the means to preserve and disseminate submissions. Describes Purdue Libraries' (PL) digital repositories for archives and documents and explains how knowledge developed over the course of establishing those units informed the development of PL's distributed data curation center. Outlines current research/study at Purdue aimed at improving the center's services and ensuring its long-term sustainability.

### Appendix 3 - Data Management Questions for Researchers on Campus

Name of  
Respondent: \_\_\_\_\_

Position: \_\_\_\_\_

Dept/Admin. Unit:  
\_\_\_\_\_

Contact  
information: \_\_\_\_\_

UBC Library has convened a working group to examine the potential role that the library could play in data management and preservation. We are at an exploratory stage of our work and have developed a short questionnaire to determine what is already being done at UBC, what areas researchers are concerned with, and where we might focus our efforts. Your input will inform the development of a wider survey of researchers at UBC and will impact the directions that the Library takes regarding data management.

**Note:** The following questions refer to research data, not administrative data:

1. What data do you store in your (faculty/department/administrative unit)?
2. What format is the data in?
3. Who has access to this data?

4. If the data is shared, by what method is the data shared?

Follow-up exploratory questions:

5. Under what conditions (if any) would you be willing to share your data? (For example, the researcher wants to be notified each time there is a request for the data; the data can only be shared within UBC; conversely, the data can be shared beyond the university, etc.)

6. What assistance/tools/funding would you need to prepare your data for use by other researchers (including preparing metadata)?

7. Any other comments?

## Appendix 4 - UBC Library's Data Management Questions for IT and other Campus Managers

Name of Respondent: \_\_\_\_\_

Position: \_\_\_\_\_

Dept/Admin. Unit: \_\_\_\_\_

Contact information: \_\_\_\_\_

UBC Library has convened a working group to examine the potential role that the Library could play in data management and preservation. We are at an exploratory stage of our work and have developed a short questionnaire to determine what is already being done at UBC, what areas researchers and administrators are concerned about, and where we might focus our efforts. Your input will inform the development of a wider survey of researchers at UBC and will impact the directions that the Library takes regarding data management.

**Note:** The following questions refer to research data, not administrative data:

1. What data do you store in your faculty, department or administrative unit?
2. What format is the data in?
3. Who has access to this data?
4. If the data is shared, what is the method used to share it? Does your method allow for the re-use and re-purposing of data by others?
5. Do you receive requests for data management and support that you cannot meet? If yes, what are the obstacles?

6. How much storage is required for the data in your unit? How is it backed up?
  
7. How much staff time is required to support data infrastructure and management in your unit?
  
8. How is the data currently being described/catalogued and what standards, if any, are being used to do this? Who has access to the descriptive information?
  
9. Are you aware of any data repositories outside of your faculty, department or administrative unit (either on or off campus) that are, or potentially could be, used by researchers in your unit? If yes, which one(s)?
  
10. What data management services would you suggest should be offered centrally?