British Columbia
Library Association

"What's the Dealio with Data Curation?"

By Danielle W. Westbrook

Alex Garnett and Carla Graebner of Simon Fraser University (SFU) presented "What's the Dealio with Data Curation?" on day two of the 2013 BC Library Conference. Garnett and Graebner talked about what data curation is, how it came about, why it is important, and what they have learnt so far from the SFU Data Curation project.

The session started with a discussion on open data. Garnett and Graebner pointed to the rise in open-data initiatives, such as the Government of Canada's Open Data Portal, and how the public benefits from data being made machine readable and freely available; open data improves accountability, drives innovation and supports research.

From their summary of open-data initiatives, Garnett and Graebner shifted to the focus of their session: research data and the importance of data stewardship and validating research findings. At SFU, research data is not limited to one subject area or type. It has been broadly defined to include all disciplines and can consist of many data types, including numeric, image-based, observational and survey data.

Garnett and Graebner highlighted the benefits of organizing and looking after research data, and emphasized that data curation is not just about stewardship. Data is living, and when it is managed, maintained and made available, it reduces duplication of research and has the potential to increase research visibility, impact, and opportunities for collaboration.

Garnett and Graebner stressed that while several research councils have had mandates, rules and regulations on retaining research data for some time, the onus has always been on the researcher. And with the emergence of data citation guides and the growing importance of data citation for tenure review, the need for data curation is ever growing. Garnett and Graebner noted that the provision of data curation services within universities seems to be a natural fit for libraries.

With the support of a University Priorities Fund grant, SFU Library has undertaken a proof of concept project to build a data repository. The focus of the project is on data integrity, preservation, managing the life cycle of data and tracking data outputs. Graebner, as project manager, and Garnett, as the data curation and digital preservation specialist, are looking at policies, procedures and costs associated with creating and maintaining a data repository. The SFU Research Data Repository is currently closed to the public, but they are creating an infrastructure that lets faculty elect to make their data open.

The project has sought data from a number of disciplines and has surveyed the SFU community to determine what kind of data is being created. Although there are commonalities across data types, SFU is aware of disciplinary considerations; from privacy concerns for clinical and health data to licensing prohibitions for text mining, not all data is made equal. Garnett and Graebner report that faculty have been responsive to the project and are excited about the emergence of big data as a research trend.

Since the start of the project, SFU resolved to make a proof of concept that is scalable and benefit-oriented. In the repository, Garnett and Graebner want to provide a system for all SFU researchers to make storing, organizing and accessing data easier. They have identified a need for normalization of terminology and standardization to improve interoperability for data storage, data use and participation in larger national and international data curation initiatives.

Garnett and Graebner concluded their session with a helpful overview of "skills to pay the bills" for libraries interested in data curation. In addition to recommending lots of discussion with community users to identify discipline specific needs, they pointed to the Inter-University Consortium for Political and Social Research (ICPSR), the Data Curation Profiles Toolkit, the Digital Curation Centre, and the Data Management for Librarians Mendeley group.

To find out more about the SFU Research Data Repository, check out their website and blog, Data Curation @SFU.

Danielle W. Westbrook is currently an E-Resources Licensing Specialist at the University of British Columbia.