University of Alberta
Campus Data Summit
March 22-23 2012

"Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly."
***Nature**, September 9, 2009*
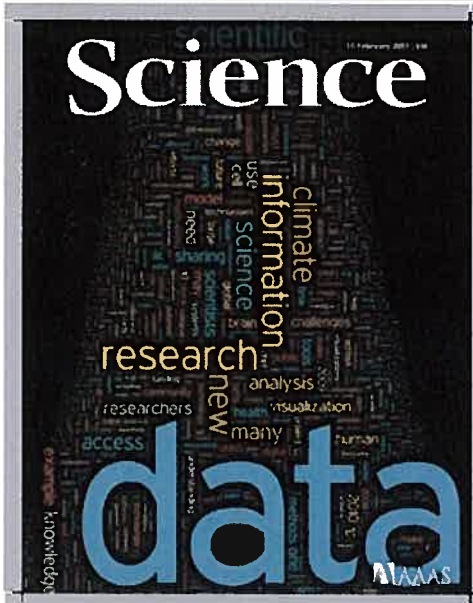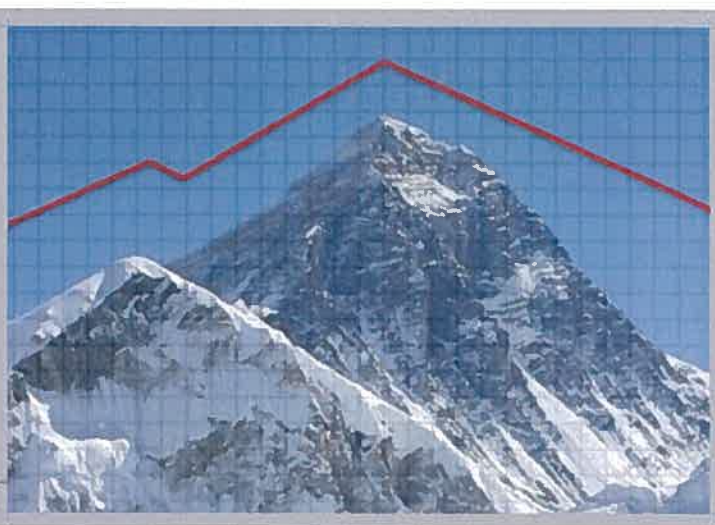
Research Data Management and Stewardship

Data are the lifeblood of research. Whether the product of observation, experimentation or simulation, data are the evidence upon which research is conducted. Given the long-standing importance of data to research, it has only been in the past few years that attention has been expressed about the treatment of data. For one thing, data are increasingly being viewed and managed as an asset of research, especially because of the substantial public investments now being made in the production of high-quality data. Rarely are data now looked upon as something to be discarded at the end of a project. As an asset, however, data need some form of stewardship to ensure their long-term preservation and usability.

The time is right to bring together significant stakeholders from across our campus to engage in a dialogue about establishing the future of data stewardship at the University of Alberta. We propose hosting a summit dedicated to the issues of producing, managing, sharing and preserving research data.

See the recommended reading list for the first Data Summit.

Why a Data Summit?

Whether identified as part of the "data deluge," data-intensive science (also known as e–Science) or simply "big data," the importance of research data has become a topic of increasing significance as field after field of scholarly study has adapted digital instrumentation and methods, replacing earlier analogue technology. In conjunction with this underlying digital shift:

- ✓ Research norms are changing with respect to the sharing of data. This is manifested in the Open Data movement and in calls for greater return on the investments made in data produced through public funding.

- ✓ New developments in scholarly communication are enabling the linkage of research findings with the evidence upon which they are based, by packaging new knowledge with the data on which it is based.

- ✓ The Web itself is generating demands for greater access to data as the semantic Web takes shape. The ability to 'mash-up' and repurpose information is becoming common practice. This requires greater attention to data preservation and dissemination capacities.

- ✓ Methods of citizen science or crowd sourcing in the collection of data (e.g., bird censuses) or its analysis (e.g., SETI, the search for extraterrestrial intelligence) have increased societal awareness of the value of data.

Simply put, data have become a mainstream topic in and outside our University. We need to move beyond this mere recognition, however, to embrace practices that protect data as an asset on this campus. To achieve this collectively, new forms of data management and greater institutional stewardship will be required.

Recommended Reading List

- Bull, J.R. *Research with Aboriginal Peoples: authentic relationships as a precursor to ethical research.* **Journal of Empirical Research on Human Research Ethics** 2010, 5, 13-22.
- Parsons, Mark A.; de Bruin, Taco; Tomlinson, Scott; Campbell, Helen; Godoy, Øystein and LeClert, Julie. *The State of Polar Data -- the IPY Experience*. In **Understanding the Earth's Polar Challenges: International Polar Year 2007-2008**. Editors, I Krupnik, I Allison, R Bell, P Cutler, DS Hik, J Lopez-Martinez, V Rachold, E Sarukhanian and C Summerhayes. 2011. CCI Press: Edmonton. pp 457-476.
- Parsons, Mark A.; Godoy, Øystein; LeDrew, Ellsworth; et al. *A conceptual framework for managing very diverse data for complex, interdisciplinary science*. **Journal of Information**

Whether identified as part of the "data deluge," data-intensive science (also known as e–Science) or simply "big data," the importance of research data has become a topic of increasing significance as field after field of scholarly study has adapted digital instrumentation and methods, replacing earlier analogue technology. In conjunction with this underlying digital shift:

- ✓ Research norms are changing with respect to the sharing of data. This is manifested in the Open Data movement and in calls for greater return on the investments made in data produced through public funding.

- ✓ New developments in scholarly communication are enabling the linkage of research findings with the evidence upon which they are based, by packaging new knowledge with the data on which it is based.

- ✓ The Web itself is generating demands for greater access to data as the semantic Web takes shape. The ability to 'mash-up' and repurpose information is becoming common practice. This requires greater attention to data preservation and dissemination capacities.

- ✓ Methods of citizen science or crowd sourcing in the collection of data (e.g., bird censuses) or its analysis (e.g., SETI, the search for extraterrestrial intelligence) have increased societal awareness of the
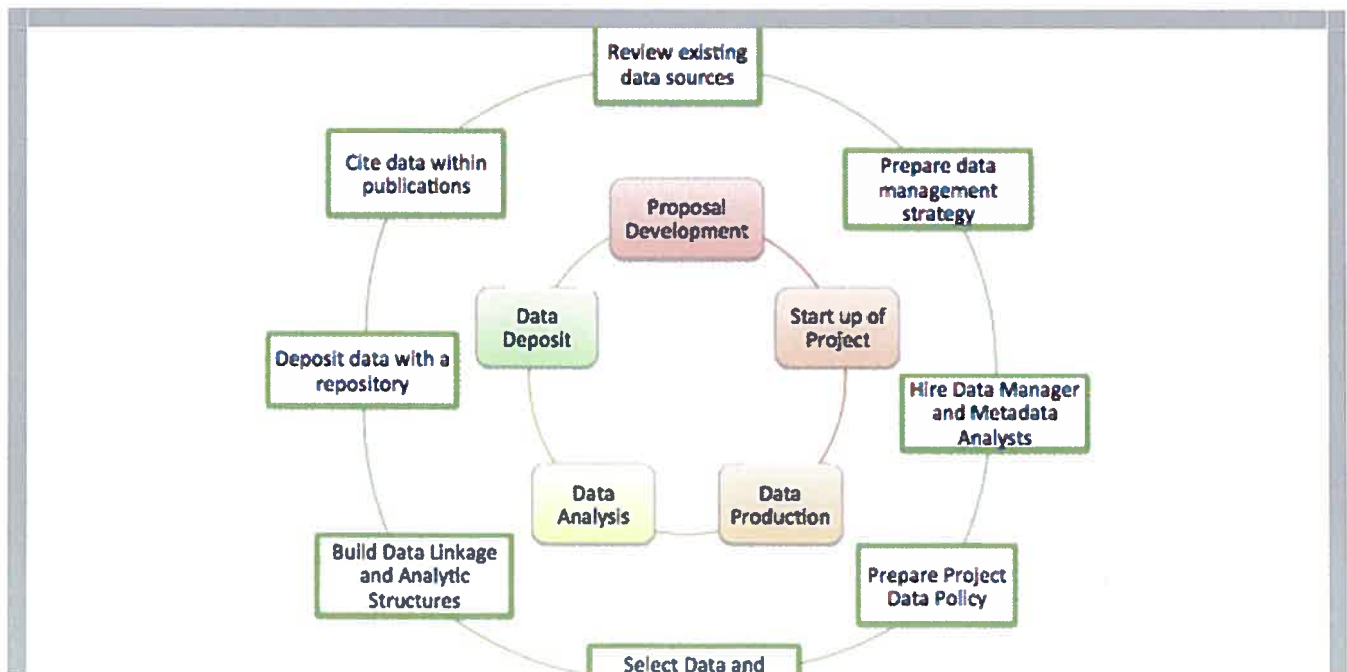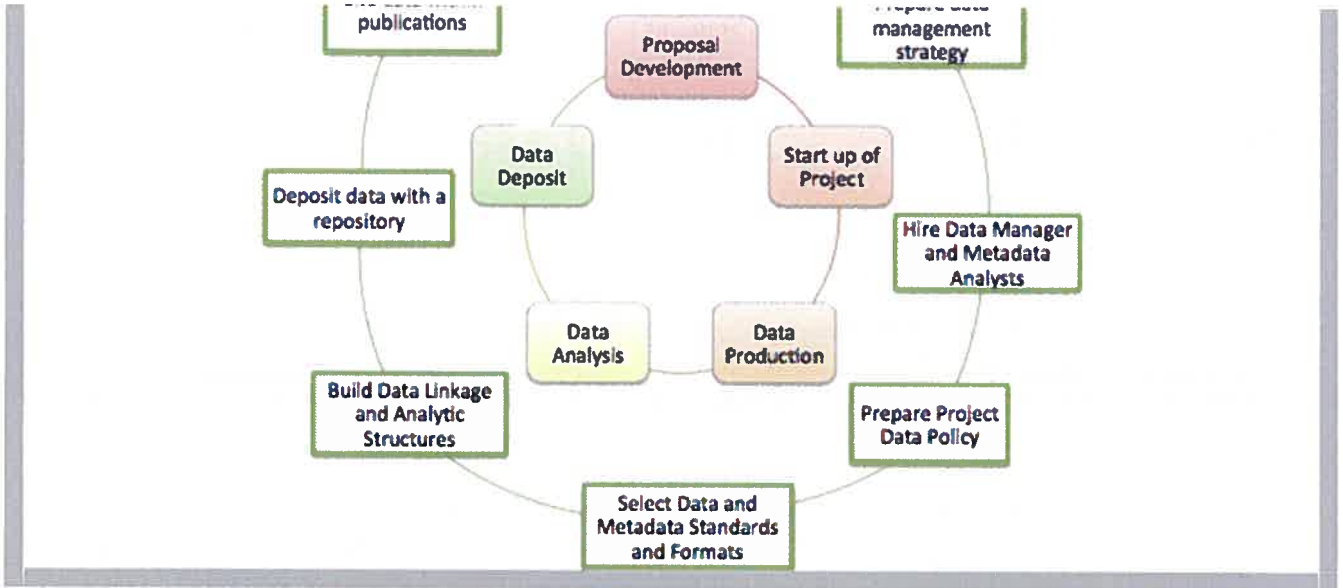
Research Lifecycle and Data Curation

Research Lifecycle

A lifecycle management approach observes each stage in a process to understand the overall process better. The research lifecycle begins with the conceptualization of a research project, continues through stages dedicated to methodology and design, data collection, analysis and the publishing of research outputs.

Each stage in a lifecycle consists of a set of related activities focused on accomplishing a primary task, the outcome of which is passed to the next stage. The granularity at which activities are described presents different views of a project. Similarly, stages in the research lifecycle can be aggregated or disaggregated into larger or smaller groupings. Nevertheless, there is a level at which a primary task will be accomplished and its outcome passed to another stage.

The research lifecycle is fundamental to understanding data curation for it is the management of data across the research lifecycle that determines the activities of data curation. The management of research data consists of many activities related to design, production, manipulation, analysis and preservation of the data and its supporting metadata. The stewardship of research data ensures that responsibilities for all data and metadata activities across the lifecycle are assigned, understood and carried out. The combination of the activities of **research data management** and the responsibilities of **data stewardship** over the research lifecycle embodies data curation.

Proposal Development

Prepare data management strategy

publications

Data Deposit

Start up of Project

Deposit data with a repository

Hire Data Manager and Metadata Analysts

Data Analysis

Data Production

Build Data Linkage and Analytic Structures

Prepare Project Data Policy

Select Data and Metadata Standards and Formats
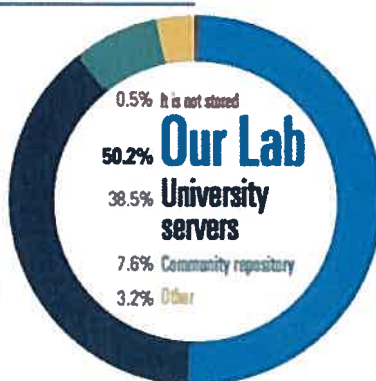
Data Sharing and Management

In conjunction with the Data Summit Series, a survey is being conducted of University of Alberta researchers to identify current data sharing and management practices. We would like to know more about the issues and challenges that individual researchers on campus face regarding the data used in their research. One aim of the Survey is to identify gaps in service that exist across the pillars of research funding, high-performance computing, long-term data preservation and access to research data collections.

If you have any questions regarding this Survey, please contact Lindsay Johnston at lindsay.johnston@ualberta.ca or call 780-492-5946.

Below is a question that the editors of **Science** asked its peer reviewers in a 2010 survey and that was subsequently reported in the February 11, 2011 issue of **Science**. This is an example of just one question also asked in our Campus Data Summit survey in which you can participate now.

Where do you archive most of the data generated in your lab or for your research?

" Even within a single institution **there are no standards for storing data**, so each lab, or often each fellow, uses ad hoc approaches. "

0.5% It is not stored
50.2% **Our Lab**
38.5% **University servers**
7.6% Community repository
3.2% Other

CREDIT: M. TWOMBLY/SCIENCE; SOURCE: SCIENCE ONLINE SURVEY

Sponsors

The University of Alberta Libraries

The University of Alberta Libraries is committed to developing research data collections and to providing access to and preserving data. The Libraries already provide access to valuable research data and have established a Data Assembly Centre supporting long-term access to research data from the recent Canadian International Polar Year. Through the Libraries digital preservation mandate, data are an integral part of digital resources being protected for long-term access. The Institutional Repository, ERA, plays a role in this infrastructure, as does the emerging national collaborative research data infrastructure of which the Libraries are a lead in Canada.

The Canadian Circumpolar Institute

The CCI is an interdisciplinary unit of the Vice President Research supporting Northern and Polar research. Data are vital to the research activities of the CCI. Consequently, the management and stewardship of the data produced by CCI researchers or required for new Northern and Polar studies are a high priority.

The Canadian Mountain Studies Initiative

The CMSI is a new group that was created to conduct, facilitate, promote, and publish research in all aspects of mountain studies. It will bring together scholars from across the campus and stakeholders in the community — environmentalists, mountain and mountaineering professionals, mountain residents, mountain businesses, mountain Indigenous peoples — to further dialogue and understanding about mountain places, activities and cultures in Canada and around the world. In this context, data management and sharing are critical.