

RDM, a view from Vancouver

Eugene Barsky, UBC

October 2015

eugene.barsky@ubc.ca



Data rich

Race teams at the **U.S. Grand Prix** collected more than **243 TB** of data in 2015, a few terabytes more data than there are in the **Library of Congress**.



Image - <https://www.flickr.com/photos/caterhamf1/>

Data rich

- Feed of about 150 key parameters, which is transmitted in real time with a 4G connection
- **5 GB** of raw data **per boat per day**



Image - <https://www.flickr.com/photos/markdoliner/>

Data poor...

- APGAR test



Image - <https://www.flickr.com/photos/hippie/>

Image - <http://goo.gl/Hq6ivo>

BIRTH CERTIFICATE
BRITISH COLUMBIA

BIRTH CERTIFICATE

 
Jack Shewchuk
Chief Executive Officer
Vital Statistics Agency

 Certified extract from
Registration of Births
Filed at Victoria, B.C.,
Canada

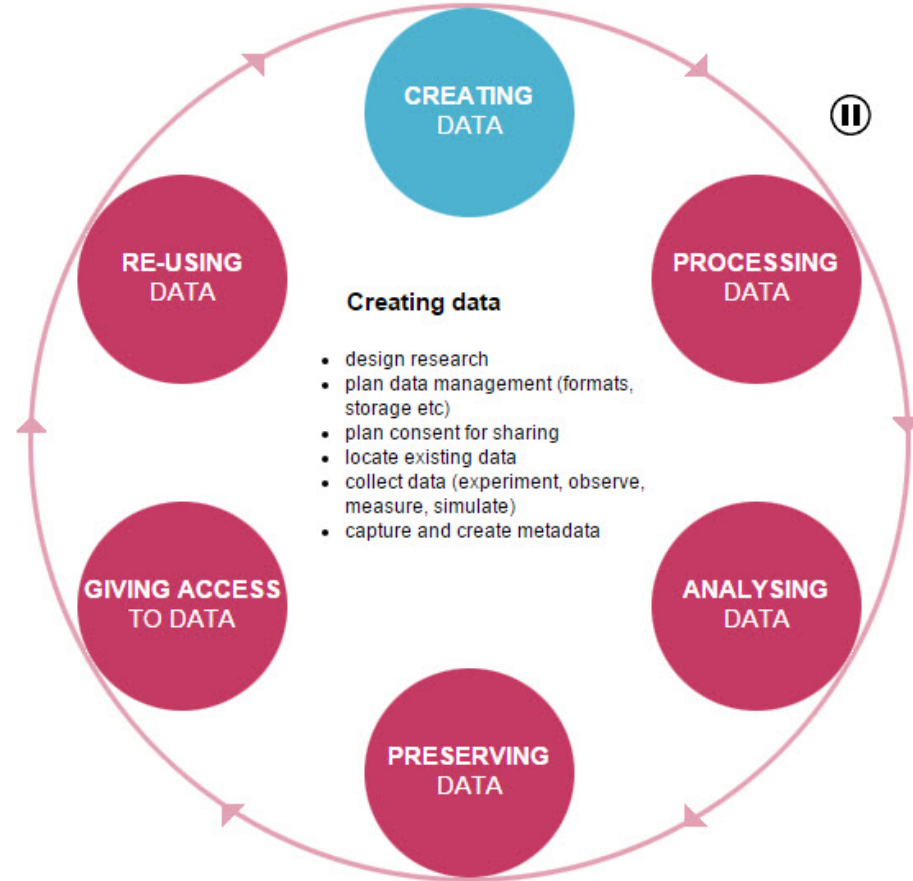
Surname **MASON**
Given Name(s) **SANDRA LEE**
Date of Birth **JUL 19, 1954** Sex **FEMALE**
Place of Birth **VICTORIA**
Registration Number **1954-59-078931**
Date of Registration **JUL 28, 1954** Date of Issue **JAN 02, 2008**
Name of Parent **HOFFER, SARCY**
Birthplace of Parent **ENGLAND**
Name of Parent **MASON, GINO ALLEN**
Birthplace of Parent **BRITISH COLUMBIA, CANADA**

SPECIMEN

 
BC 8 8 7 6 5 4 3 1

Define RDM

- From the excellent Research Data Canada (RDC) glossary - <http://www.rdc-drc.ca/glossary/>
- Research Data Management (RDM) refers to the **storage, access** and **preservation** of data produced during the entire **lifecycle** of the data



Regulations: in US...



In Canada...

Federal mandates - [Draft Tri-Agency Statement of Principles on Digital Data Management](#) - July 2015

Tri-Council open access policy - March 2015 - <http://goo.gl/XTKDII>



Image - <https://www.flickr.com/photos/twosevenoneonerineeightthreesevenatzenosix/>

Tri-Agency expectations for RDM

Researchers:

- Incorporate RDM **best practices** (in their discipline)
- Develop **Data Management Plans** (DMPs)
- Follow institutional **policies** and standards



Tri-Agency expectations for RDM

Funders

Develop **policy** and requirements that
facilitate responsible data management
Provide clear guidance for fulfill RDM
requirements

Promote the importance of excellent RDM
Provide **peer-reviewers** with guidance for
applications assessment



Image - <https://www.flickr.com/photos/sonson/>

Tri-Agency expectations for RDM

Institutions

- Provide researchers access to **repositories** that securely preserve, curate and provide access to research data
- Provide researchers with **guidance** to properly manage their data, including **DMPs**
- **Promote** the importance of research data management to researchers, students and staff

Image -<https://www.flickr.com/photos/hms831/>



How UBC Library work relates...

Dataverse (a data repository)

UBC Abacus [Dataverse](#)

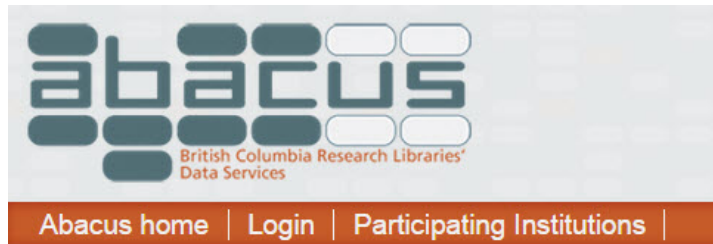
~29,000 datasets (just after Harvard and UNC)

Work closely with Harvard, UofT, Scholars

Portal, UofA, CISTI/NRC

Help UVic, UNBC w/research and licensed data and SFU w/licensed data

Part of **Portage** infrastructure for Canada (UBC, UofT, UofA)



Abacus Dataverse Network

Studies

1,762 Studies, **28,674** Files, **5,986** Downloads

i A **study** is a container for a research data set. It includes cataloging information, data files and complementary files.

Discoverability of UBC data

- Make our data findable
- with **DOIs**
- with [OAI](#)
- feed data to ORCID, Datacite, Google, VIVO, Data Citation Index and more
- **73 percent** of our researchers (our RDM survey) expressed a need in institutional data repository



How UBC Library work relates

DMP Assistant (for data management plans)

- Close collaboration with UofA
- First Data Management Plans (**DMPs**) for Canada (late 2014)
- **Portage** project - National template ready in English/French
- Part of national Portage infrastructure



DMPs:

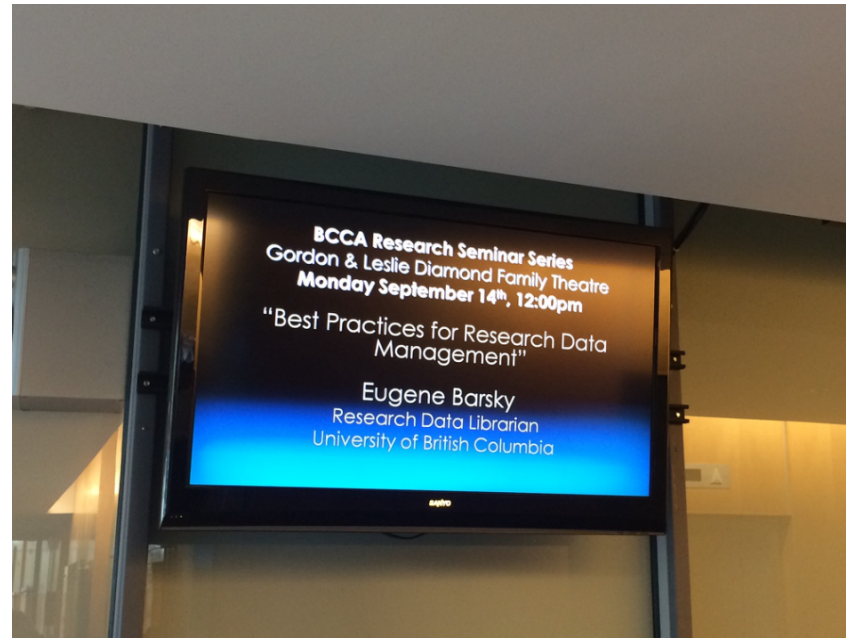
- Support Portage national template
- Help UBC Researchers with **Data Management Plans** as they are introduced by Tri-Agencies
- **77 percent** of our researchers (from our RDM survey) need help with DMPs



How UBC Library work relates

Teaching and Outreach for Research data

- [Research data website](#)
- [DataGuide\(s\)](#)
- [Workshops](#) (Data Cafe and Best Practices)
- **73 percent** of our researchers (in our RDM survey) expressed an interest in RDM workshops



Teaching and Outreach

- Continue being a **hub** for research data outreach with teaching and data cafe
- Build on our [website](#) for UBC best practices with research data
- Continue helping other schools with open access teaching materials



thanks: sue clark, ames hawkins

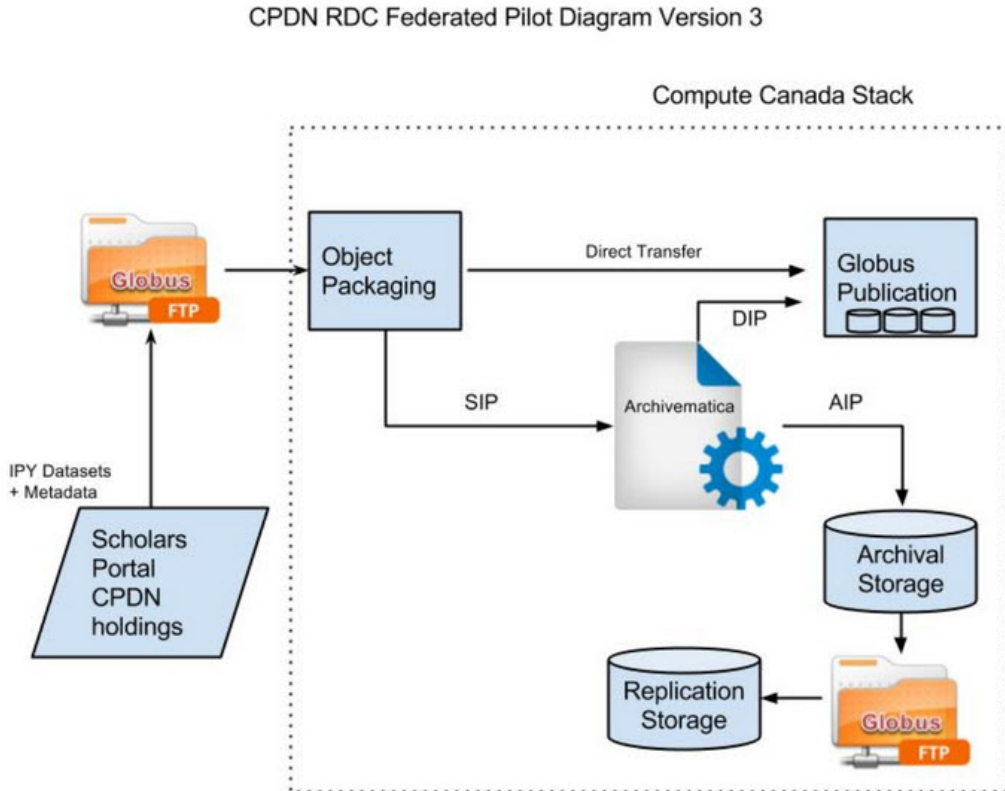
Digital preservation

- Ensure digital **preservation** and **normalization** of selected datasets
- For next generations (with [Archivematica](#))
- Example: [BIRS videos](#) (more than 15TB)



How Canada is different...

- Small country - shared infrastructure and services
- Collaborations between the major schools - UofA, Scholars Portal, SFU and UBC and Compute Canada



Questions?

