
Eugene Barsky, UBC
June 2016
eugene.barsky@ubc.ca

Image - https://www.flickr.com/photos/jannekestaaks/
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/xavier33300/
Define RDM

- From the excellent Research Data Canada (RDC) glossary - [http://www.rdc-drc.ca/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

Image - [http://www.data-archive.ac.uk/create-manage/life-cycle](http://www.data-archive.ac.uk/create-manage/life-cycle)
In Canada...


Image - [https://www.flickr.com/photos/morbokat/](https://www.flickr.com/photos/morbokat/)
Tri-Agency expectations for RDM

Researchers:

- Incorporate RDM best practices (in their discipline)

- Develop Data Management Plans (DMPs)

- Follow institutional policies and standards

Image - https://www.flickr.com/photos/jdhancock/
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/](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
How UBC Library work relates...

Dataverse (a data repository)

- UBC Abacus Dataverse
- ~30,000 data files (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)
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

Image - [https://www.flickr.com/photos/jannekestaaks/](https://www.flickr.com/photos/jannekestaaks/)
Dataverse features

Data access:
- Granular, for study or file level
- Allows to open metadata only
- Can open data to a specific group / UBC / or anyone

Version Control:
- Automatic (but not Git of course!)
- Shows differences between versions

Image - https://www.flickr.com/photos/jkohen/
Dataverse features

Data analysis:
- In your browser for tabular data
- Universal Numeric Fingerprints (UNFs)
- Or download data to any statistical analysis software for analysis (R, SPSS, etc)

Image - [https://www.flickr.com/photos/lendingmemo/](https://www.flickr.com/photos/lendingmemo/)
Digital preservation

- Ensure digital **preservation** and **normalization** of selected datasets

- For next generations (with **Archivematica**)  

- With Dataverse 4.X

Image - [https://www.flickr.com/photos/98216330@N00/](https://www.flickr.com/photos/98216330@N00/)
Code translated to French

- By our colleagues @ UMontreal
- Also to Spanish, Portuguese and Mandarin

Image - https://www.flickr.com/photos/jimforest/