

Discovery of research data

Eugene Barsky

eugene.barsky@ubc.ca

<http://researchdata.library.ubc.ca/>

Fall 2018

Full presentation available - <https://dx.doi.org/10.14288/1.0372083>

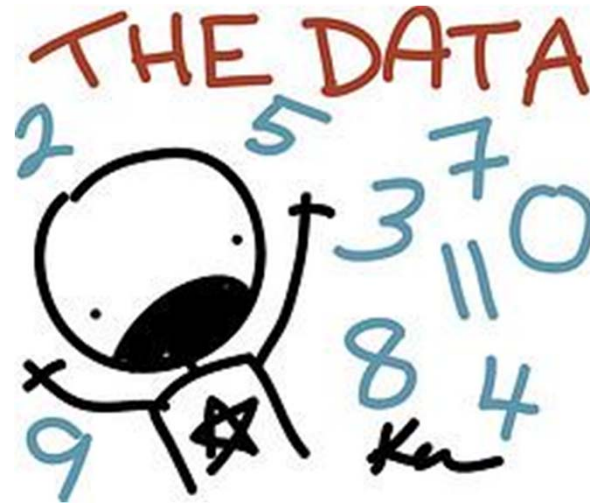


Image - <https://www.flickr.com/photos/kenfagerdotcom/>

Outline

- Background:
 - Definitions
 - Tri-Agencies direction for data discovery
- How to make research data findable and discoverable
 - Principles and best practice
- We do it too:
 - Abacus Dataverse
 - Federated Research Data Repository (FRDR)



Image by <http://epicgraphic.com/metaphors/>

Data rich

Soccer clubs, like Arsenal, record on average 10 data points per second for every player on the field, or about 1.4 million data points per game.

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

Source - <https://www.forbes.com/sites/bernardmarr/2015/03/25/big-data-the-winning-formula-in-sports/#2a9791e234de>



Define research data

Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results.



Source - CASRAI Glossary - http://dictionary.casrai.org/Research_data

* Image - <https://www.flickr.com/photos/34547181@N00/>

Timeline

- Tri-Council introduced Draft RDM policy in June 2018 - http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html
- Public consultation for a period of two-three months.
- Six months after the policy has been publically available, institutions will be expected to enact RDM policies.
- Realistic timeline - Fall 2019 for compliance.



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

DRAFT TRI-AGENCY DATA MANAGEMENT POLICY

- For consultation
- Feedback will inform final policy
- Proposed policy includes 3 possible requirements:
 1. Institutions: Institutional Strategy
 2. Researchers: Data Management Plans
 3. Researchers: Data Deposit
- Implementation: Phased, incremental

Requirement #2. Data Repositories and Discovery

“Grant recipients are required to deposit into a recognized digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications, pre-prints, and other research outputs that arise from agency-supported research. The repository will ensure safe storage, preservation, and curation of the data. The agencies encourage researchers to provide access to the data where ethical, legal, and commercial requirements allow, and in accordance with the standards of their disciplines”



Focus on Data Deposit for Discovery

Set of Principles:

- ☐ Common metadata
- ☐ Persistent identification
- ☐ Open access
- ☐ Common licensing
- ☐ Collaboration (coexistence in the scholarly ecosystem)

White papers released in 2016/17:

- Barsky, E., Brosz, J., & Leahey, A. (2016, July 31). Research Data Discovery and the Scholarly Ecosystem in Canada : A White Paper. doi:<http://dx.doi.org/10.14288/1.0307548>
- Leggott, Mark, Shearer, Kathleen, Ridsdale, Chantel, Barsky, Eugene, & Baker, David. (2016, September 9). Unique Identifiers: Current Landscape and Future Trends. Zenodo. <http://doi.org/10.5281/zenodo.557106>
- Fenner, M., Crosas, M., Grethe, J., Kennedy, D., Hermjakob, H., Rocca-Serra, P., ... & Clark, T. (2017). A data citation roadmap for scholarly data repositories. bioRxiv, 097196.

Practical principles for Discovery: Metadata



- Use common and established metadata schemas - Dublin Core, DDI, Datacite...
 - For instance Google new Data Search - <https://toolbox.google.com/datasetsearch> is using Schema.org metadata standard
- Dataset landing page -
 - Metadata need to be embedded into the dataset landing page so that the indexers/harvesters can find them
- Search engine is only as good as the metadata that go into it!!

Practical principles for Discovery: Persistent Identifiers

- All datasets intended for discovery should have a globally unique persistent identifier that can be expressed as unambiguous URL (e.g. DOI, ARK or Handle)
- It should be embedded in the landing page in machine-readable format
- This persistent identifier expressed as URL must resolve to a landing page specific for that dataset
- Persistent identifiers for datasets should support multiple levels of granularity, where appropriate (e.g. DOIs for individual files in a study dataset)

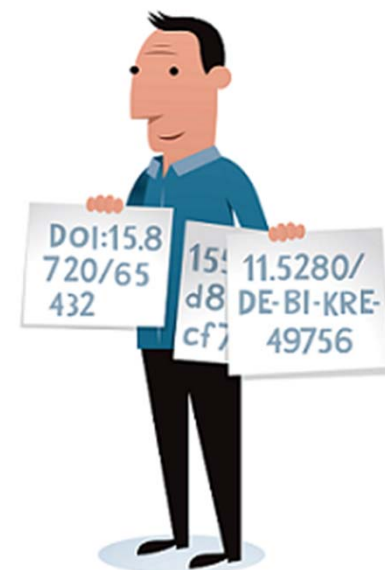


Illustration by Jørgen Stamp CC BY 2.5 Denmark

Practical principles for Discovery: Open Access and APIs

- A repository should provide an API or at least work with OAI-PMH protocol
- OAI-PMH protocol provides consistent, structured, and interoperable formats for metadata exchange
- Caveat: Harvesting metadata doesn't address issues or concerns about metadata quality, completeness, or a common metadata across repository systems

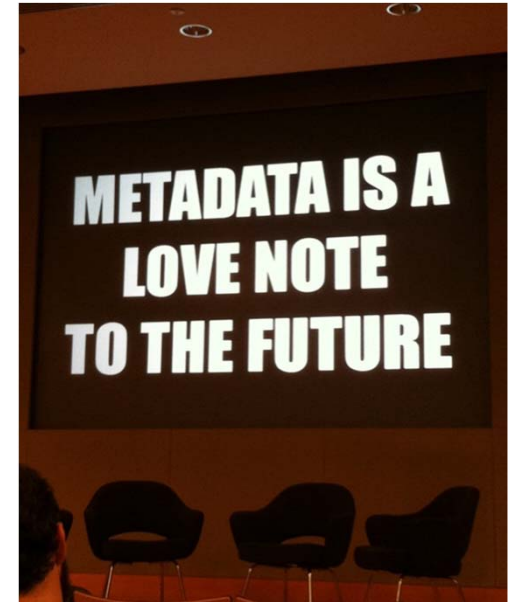


Image - <https://www.flickr.com/photos/centralasian/>

Practical principles for Discovery: Licensing

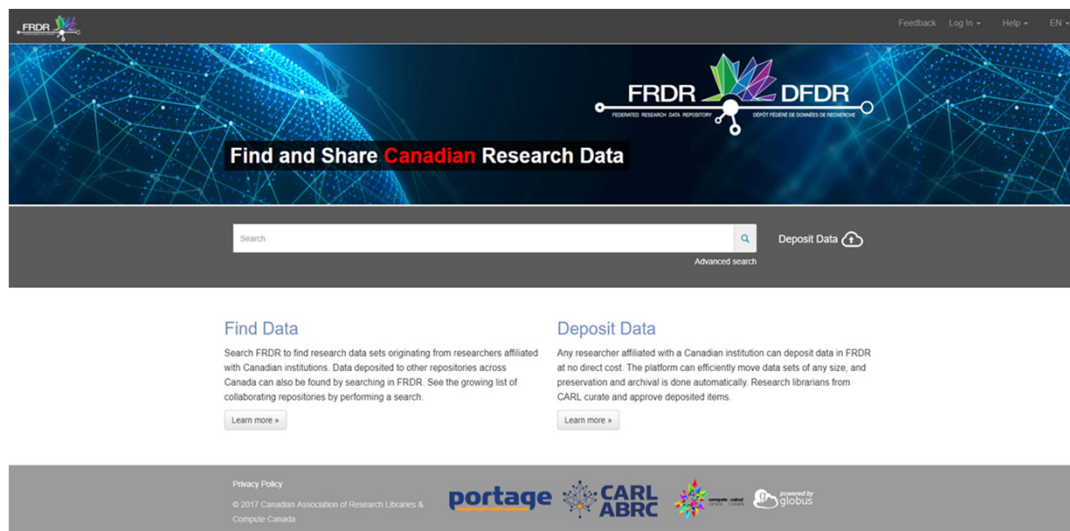
- We believe that nobody yet has solved all the complexities of making data openly available and reusable.
- We prefer applying **CC-0** license to open data (same as Dryad, Biomed Central, Europeana and others). See more:
 - Einhorn, David, et al. "Post-Publication Sharing of Data and Tools." Nature, vol. 461, no. 7261, 2009, pp. 171-173.



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

Data Repositories and Discovery - FRDR

- We have worked to create the national research data discovery layer with Federated Research Data Repository (FRDR) - national discovery layer for research data
- <https://www.frdr.ca/>



Dataverse Repositories

- UBC Abacus Dataverse - <http://dvn.library.ubc.ca/dvn/> - has more than 38,000 data files under management
- We mint DOIs and expose research metadata into:
 - Summon
 - Google
 - Bing
 - Datacite
 - Google Scholar and more...

[CITATION] Minding the gaps: Comparing engineering research output and library holdings at four large universities

E Barsky, SJ Dooley, T Mawhinney, A Saundry... - 2015 - open.library.ubc.ca

... Koerner Library; Law Library; Music, Art and Architecture Library; Rare Books and Special Collections; University Archives; Woodward Library; Xwi7xwa Library. UBC Vancouver Off-Campus; Biomedical Branch Library; UBC Okanagan Campus; Innovation Library; Okanagan Library. Use The Library: Borrowing Services; My Library Account; How to Get a Library Card; See More... Computers & Technology; Print, Copy, Scan; Public Computers & Software; See More... Guides for Library Users; Undergraduate Students; Faculty & ...

☆ 99 88

Abacus Dataverse Network

MINDING THE GAPS: COMPARING ENGINEERING RESEARCH OUTPUT AND LIBRARY HOLDINGS AT FOUR LARGE UNIVERSITIES

hdl:11272/10106

Version: 15 - Released: Thu Jan 12 14:29:24 PST 2017

Cataloging Information

DATA & ANALYSIS

Comments (0)

Versions

Minding the gaps: Comparing engineering research output and library holdings at four large universities

Eugene Barsky, Sarah Jane Dooley, Tara Mawhinney, Amber Saundry & Michelle Spence

Work published 2015 via The University of British Columbia

This is supplementary data for the ASEE 2015 conference paper titled - "Minding the gaps: Comparing research output and library holdings in four large engineering schools"

https://doi.org/10.14288/1.0314343 Cite

Minding the gaps: Comparing research output and library holdings in ...

dvn.library.ubc.ca/dvn/dv/UBC_RD/faces/study/StudyPage.xhtml?...?studyId...2

Jan 6, 2015 - Minding the gaps: Comparing research output and library holdings in four large engineering schools ... Tara (McGill University); Saundry, Amber (University of British Columbia); Spence, Michelle (University of Toronto).

Questions?



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