Open Data

What it is, and where to find it

Martin Chandler
Data Services Librarian
martin.chandler@mcgill.ca

Download this presentation

- https://tinyurl.com/OpenDataMcGill

Agenda

- A little about Data
- What is Open Data?
- Kinds of Open Data
- Government Data
- Research Data
- Big Data
- Questions?

What is data?

- Data is facts, in different forms, about a particular subject, used to answer a question or to make an argument.
- This can include numbers (numerical data), information about locations (geospatial data), or words about something (descriptive or otherwise)
- Eg: The number of couples introduced through the pineapple export business is a point of data (being 1, as far as I know)
- The number of dogs living in a particular household is another data point

Some terminology

- Data: facts, or factual information, used for reasoning or analysis
- Data point: one fact or unit of study
- Dataset: A collection of facts, often gathered together to be manipulated

Dataset

Home ID	Postal Code	Dog	Breed	Temperament
001	M6R 1N6	0	N/A	N/A
002	H4C 3C5	1	Beagle	Friendly
003	B2R 1S2	2	Dachshund; Terrier	Cute; Derpy
004	L2S 2A7	0	N/A	N/A

Data point

What is Open Data?

- Open data is digital data, either private or public, produced and disseminated with an open license the guarantees free access, and the allowance of reuse without technical, legal, or financial restrictions.
- What about older data, collected by hand?
- Yes, that can be made open, but a large part of the idea of "Open" is a threshold of accessibility. A ledger in an archive may be freely accessible, but it is not easily accessible to anyone for use and reuse, unless digitized.

Source: Monino, J.-L., & Sedkaoui, S. (2016). *Big data, open data and data development* (Ser. Smart innovation set, volume 3). pp 23-24. ISTE, Ltd. https://onlinelibrary-wiley-com.proxy3.library.mcgill.ca/doi/pdf/10.1002/9781119285199.ch2

What is Open Data?

- Often separated into 3 types:
  1. Open Government Data – information funded, created, collected, processed, maintained, and disseminated by a government or public institution.
  2. Open Research Data – any data used for research purposes where access is opened beyond producers.
  3. "Big Data" – aggregations of data from many sources, including public and private

Source: Wessels, B., Sveinsdottir, T., Finn, R., & Wadhwa, K. (2017). *Open Data and the Knowledge Society*. Amsterdam: Amsterdam University Press. Retrieved October 20, 2020, from https://openresearchlibrary.org

What is Open Data?

- In short:
- It is "Accessible, understandable, and open to reuse." (Wessels et al, 2014)
- Can be divided into
  1. public data or information from the public sector
  2. data from research institutions, especially publicly funded
  3. private sector data

Where to find Open Data

Where to find Open Data

- “It depends.” – Martin Chandler
- Questions to consider:
  - Are you looking for governmental data, already-produced research data, or miscellaneous data?
  - What is the geographic region you’re looking at (if any)?
  - What is the group you’re researching?
  - What time period are you researching?
  - Who else might be asking your question?

Where to find Open Data

- Are you looking for governmental data, already-produced research data, or miscellaneous data?
  - Each has different places/modes for searching
- What is the geographic region you're looking at (if any)?
  - Geography will determine where to search, especially for government data
- What is the group you're researching?
  - Partly will determine whether this will exist as open data or not – data, especially about people, is often protected

- What time period are you researching?
  - The older the data, the less available it will be (eg data pre-1950 will be more difficult to find, unless it's a heavily-researched subject)
- Who else might be asking your question?
  - Depending on the people researching, they may be more or less committed to open data, and/or knowledgeable about it

Where to find: Government Data

- Geography is the big question here!
- The more localized your data need, the more options for searching (and more places you need to check).
- Federal, Provincial, and many Municipal governments have or are developing open data repositories.
- As well, inter-governmental organizations often have open data repositories (eg UN, World Bank, etc).
- Non-governmental organizations often partner with governments, as well, and have their own repositories.

Where to find: Government Data

- Eg. "I'm searching for data on snails in Stanley Bridge, Prince Edward Island."

- In this case, you can check the PEI Open Data and the Canada Open Data site (Stanley Bridge is too small to have an Open Data repository need – any data would go to the PEI portal).
- The "snails" might also send us to the World Wildlife Fund or similar.

(Some) Government Open Data

- http://bartlettr.github.io/opendatawebsites/
- https://library.carleton.ca/find/gis/geospatial-data/open-data-repositories
- https://www.data.gov/
- https://data.gov.uk/
- https://data.europa.eu/euodp/en/data/
- https://dataportal.opendataforafrica.org/
- You can search for "[Place X] Open Data"
- Remember counties/regions (eg. Niagara Open Data)!

Where to find: Research data

- Partly depends on geography, and partly on where the researcher who produced that data was…
- Research data is something we're all trying to figure out, at the moment.
- Researchers often don't make their data open when working with it, so they can get the publication out there.
- At the end of the research lifecycle, placing data in repository for use is easy to miss.
- But please do it (if there aren't restrictions on the data)! It will help everyone.

Where to find: Research data

- Federated Research Data Repository (FRDR): https://www.frdr-dfdr.ca/repo/
- To help searching the above, Geodisy: https://geo.frdr-dfdr.ca/
- <Odesi>: https://search1.odesi.ca/#/
- Many institutions have their own research data repositories (McGill's: https://dataverse.scholarsportal.info/dataverse/mcgill)
- To find the above, it's good to know any publications from the data – you can then use the institutional affiliation of the researcher to look in that institution's repository.

Where to find: Big data

- Google
  - Various (research?) https://datasetsearch.research.google.com/
  - Public https://www.google.com/publicdata/directory
- Esri's Living Atlas (mainly geospatial data).
- Various companies (eg BIXI).
- Creative Commons.
- Often searching "[Company/Subject X] Open Data" will bring results (maybe, and of varying degrees of trustworthiness).

Where to find: Big data

- The hype continues.
- Often not open, though companies may have a branch that makes SOME data open.
- Social media? Fuzzy lines (eg Twitter scraping).
- All told, this is the least clear realm of “open data”.

So how do I find…

- Great question! Some links here, or send me an email. If it exists, I'll do my best to find it!
- If you're creating data, or re-using data in different ways, make it open! How, you ask? Contact rdm.library@mcgill.ca, and our Research Data Management Specialist can help you out!
- Think about your subject, geography, and timeframe. Who is creating this data, and who would collect data on it?

Questions?

martin.chandler@mcgill.ca

Sources

- Chandler, M. (2020). “It depends.” [Commentary developed for presentation].
- Monino, J.-L., & Sedkaoui, S. (2016). *Big data, open data and data development* (Ser. Smart innovation set, volume 3). ISTE, Ltd: London, UK. https://onlinelibrary-wiley-com.proxy3.library.mcgill.ca/doi/pdf/10.1002/9781119285199.ch2
- Smulan77. Prince Edward Island - 2003 (11). Flickr (photography social media). Retrieved October 20, 2020 from https://www.flickr.com/photos/smulan77/3358139277/in/photostream/
- Wessels, B., Finn, R., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood et al. (2014) ‘Issues in the Development of Open Access to Research Data’, Prometheus Critical Studies in Innovation, 32:2, 49–66.
- Wessels, B., Sveinsdottir, T., Finn, R., & Wadhwa, K. (2017). *Open Data and the Knowledge Society*. Amsterdam: Amsterdam University Press. Retrieved October 20, 2020, from https://openresearchlibrary.org