"Open" Data & Libraries (in Canada)

November 1, 2013

Open Data as Part of Government Information Day

Berenica Vejvoda

Data Librarian

Map & Data Library

University of Toronto Libraries

Agenda

- What
- Why
- Prior to "open"
- Opening up – DLI / CAPDU
- Open 'public data'
- Open research data

What is Open Data?

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form

- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets

- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed

Interoperability

- Interoperability - Denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets

- This ability to componentize and to 'plug together' components is essential to building large, complex systems

Principles of Open Data (Gov't Canada)

1. Data Must Be Complete

All public data are made available. Datasets released by the Government of Canada should be as complete as possible, reflecting the entirety of what is recorded about a particular subject. All raw information from a dataset should be released to the public, except to the extent necessary to comply with federal Access to Information and Privacy legislation. Metadata that defines and explains the raw data should be included as well, along with any available explanations for how derived data was calculated.

2. Data Must Be Primary

Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.

3. Data Must Be Timely

Data are made available as quickly as necessary to preserve the value of the data.

Principles of Open Data

4. Data Must Be Accessible

Data are available to the widest range of users for the widest range of purposes.

5. Data Must Be Machine Processable

Machines can handle certain kinds of inputs much better than others. Datasets released by the Government of Canada should be stored in widely-used file formats that easily lend themselves to machine processing. These files should be accompanied by documentation related to the format and how to use it in relation to the data.

6. Access Must Be Non-Discriminatory

Data are available to anyone, with no requirement of registration or membership. Anyone can access, anytime w/out need to identify or justify.

Principles of Open Data

7. Data Formats Must Be Non-Proprietary

Data are available in a format over which no entity has exclusive control.

8. Data Must Be License-free

Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.
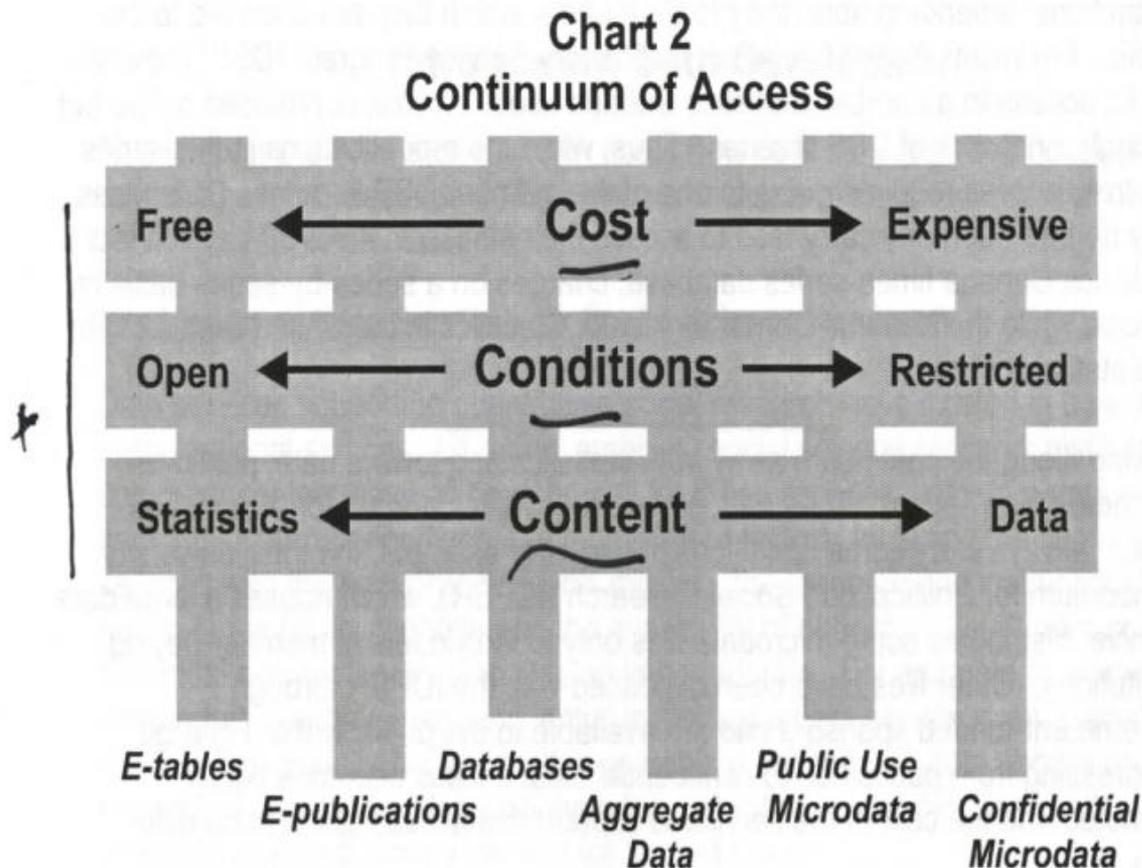
9. Permanence (version-tracking & archiving)

10. Usage Costs

The Government of Canada releases the data on the Open Government site free of charge.

Continuum of Access

Chart 2
Continuum of Access

Free ←	**Cost** →	Expensive
Open ←	**Conditions** →	Restricted
Statistics ←	**Content** →	Data

E-tables Databases Public Use
E-publications Aggregate Microdata Confidential
Data Microdata

UNIVERSITY OF TORONTO LIBRARIES

Why Open Data?

- Transparency and democratic control
- Participation
- Self-empowerment
- Improved or new private products and services
- Innovation
- Improved efficiency of government services
- Improved effectiveness of government services
- Impact measurement of policies
- New knowledge from combined data sources and patterns in large data volumes

Prior to "Open"

- Pricing on individual-user basis (since 1970s)
- Cost to cover duplication and shipping of magnetic tapes
- Too expensive for students
- Exponential hike due to budget in 1980s
- Only few very well-funded researchers could afford
- Lack of infrastructure (few data centers and no national program of support)
- CANSIM2 University Base – few time series, little social data and no microdata
- US (ICPSR) + European Data Archives
- Gave up!

Prior to "Open"

- CAPDU (Canadian Association of Public Data Users) lobbied gov for STC data
- CARL (Canadian Association of Research Libraries)
  - Consortial purchase 1986 Census of Population
  - 5 General Social Survey Files (UofT as disseminator)
- Dilemma: lack of tech & prof support and complexity and size of files
- CARL- Adhoc budget commitments on ongoing basis

Data Liberation Initiative

- Convincing Gov – leakage and loss of revenue
- ULs less reluctant, save lack of skilled librarians
- Proposal:
  - STC and academic partnership to enhance use of STC data
  - All "publicly available" data files (microdata, aggregate data and geography files) would be available to members
  - Funded by STC and universities (large $12K; and others $3K)
  - License agreement which designates a contact for support and to liaise with STC
  - **Academic purposes** (teaching, research and academic planning)
  - Not to be redistributed or used commercially

Data Liberation Initiative

- TBC approved the DLI as a 5-year pilot – January 1996
- Then on-going
- Infrastructure – Dissemination Division (Later LIC and MAD)
- External Advisory Cmte (EAC)
  - Data specialists, UL, researcher, STC staff; Chair external researcher
  - 2x/yr

Data Liberation Initiative

- Infrastructure for universities
  - UofT leader in serving as a repository for smaller institutions
- 75 universities and colleges
- 92 surveys
- 1300 survey files
- 50,000 files (census, tables and dbases)
- 16 years to date
- 70% U; 30% STC funding
- Subscription cost has not changed since 1996
- Focus on training

Open Data in Canada

- Canada's Freedom of Information legislation- which includes the Access to Information Act, the Privacy Act and the Federal Accountability Act, require the government to proactively publish budget documents and audit reports in both official languages

- Although not required to by the this legislation, the government also publishes administrative data sets and has launched an online Open Data Portal, which made more than 260,000 datasets from ten participating departments available to all Canadians

- Most administrative data is released through Ministry/Agency websites (vs. central portal)

- However, while 53% of OECD countries have requirements to publish information in open data formats, Canada does not which could hinder the re-use of information by other parties

Open Data (Federal Gov)

- Canada's President of the Treasury Board is the political lead on Canadian open government, including open data.
- Public information in Canada is the subject of the Information Act and Proactive Disclosure
- On March 17, 2011, Minister Stockwell Day, then president of the TBC announced the launch of the 12-month pilot for Canada's National Open Data site.
- On April 11, 2012, Minister Tony Clement, announced Canada's Action Plan on Open Government, a 3-year plan for open data including:
  - expansion of the number of datasets made available, gathering requirements for the next generation platform, in the first year; and
  - design and implementation of the new data.gc.ca portal, improving the level of standardization of data, in the second and third years
- On June 18, 2013 an updated version of the portal was launched, along with an updated Open Government License

Crown Copyright News (~2010) – Change in permissions

Crown Copyright and Licensing (CCL) is pleased to announce that permission to reproduce Government of Canada works is no longer required, in part or in whole, and by any means, for personal or public non-commercial purposes, or for cost-recovery purposes, unless otherwise specified in the material you wish to reproduce."

Open Data (Federal – Science)

- **National Research Council Canada**
  - Gateway to Scientific Data
  - Datasets
  - Management and Curation

Open Data (Ontario)

- As of November 8, 2012, the Province of Ontario has an open data portal. It launched with 63 files.

- October 21, 2013 – Ontario to release "much more" through the open data portal

- There are several citizen advocacy groups throughout Canada that work with governments and institutions to adopt open data policies:
  - Guelph: OpenGuelph
  - Halton: OpenHalton
  - Hamilton: OpenHamilton
  - London: Open Data London
  - Mississauga: Mississauga Data
  - Ottawa: Open Data Ottawa
  - Waterloo Region: Open Data Waterloo Region
  - Windsor: Open Data Windsor

UNIVERSITY OF TORONTO LIBRARIES

STC Open Data Declaration – February 2, 2012

- Nov 24, 2011 Embassy Magazine released story that all online STC data will be made free and released under the Government of Canada's Open Data License Agreement that allows for commercial re-use

- February 1, 2012 – Press release released by STC to this effect

- FAQs for Open License Agreement

- "This license agreement allows you to use Statistics Canada information without restrictions on sharing and redistribution, for commercial and non-commercial purposes. You must always acknowledge Statistics Canada as the source of the information and adhere to the conditions of the Statistics Canada Open License."

STC Open Data Declaration – February 2, 2012 – "The Details"

- Great news for the general public – CANSIM, PUMFs

- Postal code products (Canada Post copyrighting postal code data)

- RDC confidential master-files

- Still pay for custom tabulations but once purchased they are free

- Retro custom purchases also open

Open Research Data

- Many research funders (now) advocate 'Open Access' to research outputs and have Open Access policies
- This **may or may not** include research data
- **1990: SSHRC Research Data Archiving Policy**
  - "All research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time."
  - Costs associated with preparing research data for deposit are considered eligible expenses in SSHRC research grant programs
  - http://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/edata-donnees_electroniques-eng.aspx

Open Research Data

- 2000: the Social Sciences and Humanities Research Council and the National Archivist of Canada established a **Working Group** of research and archival experts and asked them to assess the need for a national research data access, preservation and management system

- The Working Group found large gaps in the infrastructure for collecting and preserving research data.

- **2004: SSHRC Open Access Policy**
  - Adopted, in principle, a policy of open access that would guide the development of its research support programs."

- **2004: Canada was a signatory of the OECD Declaration on Access to Research Data From Public Funding**
  - But, unlike some other national signatories, Canada has not responded with concrete actions to give its agreement to the MOU substance.

Open Research Data

- **2005: National Consultation on Access to Scientific and Research Data** - "road map" for the implementation of a national plan for open access to publicly funded scientific research data.

- **2008: CIHR – policy on access to research outputs**

  - improve access to research outputs funded by CIHR

- **2013: Consultation on the draft Tri-Agency Open Access** Policy – no new data mandates

Where are we?

- "In a world where data increasingly matters, it is increasingly going to become political. Very political." – David Eaves
- Opening it up does not depoliticize it
- Current gov't has imposed huge cuts to STC
  - "The federal budget chopped $33.9-million from the agency's budget by the 2014-15 fiscal year." – The Globe and Mail, April 23, 2012
  - "These cuts will likely affect the agency's ability to introduce new surveys and update existing surveys to reflect current population trends. Statscan now produces about 350 surveys"
- Quantity and quality of data is under threat
- No new data sharing mandates for research data
- Still no national data archive

UNIVERSITY OF TORONTO LIBRARIES