**Let’s Talk About Open Data**
Handout

**GLOSSARY** Source: [*opendatahandbook.org/glossary/en*](http://opendatahandbook.org/glossary/en/)

**Creative Commons**: A non-profit organisation founded in 2001 that promotes re-usable content by publishing a number of standard licences, some of them open (though others include a non-commercial clause), that can be used to release content for re-use, together with clear explanations of their meaning. [*creativecommons.org*](https://creativecommons.org/)

**Hackathon**: An event, usually over one or two days, where developers, subject experts and others come together to create apps, visualisations, and prototypes that aim to address problems in a particular domain, usually making heavy use of data. The hackathon is a popular format in the open source community.

**Human Readable**: Data in a format that can be conveniently read by a human. Some human-readable formats, such as PDF, are not machine-readable as they are not structured data, i.e. the representation of the data on disk does not represent the actual relationships present in the data.

**Machine-readable**: Data in a data format that can be read and processed by a computer. Format examples include CSV, JSON, XML, etc.

**Mash-up**: A web page or application created by combining data or functionality from different sources.

**Open Access**: The principle that access to the published papers and other results of research, especially publicly-funded research, should be freely available to all. This contrasts with the traditional model where research is published in journals which charge subscription fees to readers. Besides benefits similar to the benefits of open data, proponents suggest that it is immoral to withhold potentially life-saving and valuable research from some readers who may be able to use or build on it. Open-access journals now exist and the interest of research funders is giving them some traction, especially in the sciences.

**Open data**: Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike. Specifically, open data is defined by the [Open Definition](http://opendatahandbook.org/glossary/en/terms/open-definition/) (Open Definition “is the main international standard for open data and open data licences, providing principles and guidance for all things ‘open’”; *opendatahandbook.org/glossary/en/terms/open-definition*) and requires that the data be:

**Legally open**: that is, available under an open (data) license that permits anyone freely to access, reuse and redistribute.

**Technically open**: that is, that the data be available for no more than the cost of reproduction and in machine-readable and bulk form.

**Open government**: Open government, in line with the open movement generally, seeks to make the workings of governments transparent, accountable, and responsive to citizens. It includes the ideals of democracy, due process, citizen participation and open government data. A thorough approach to open government would also seek to enable citizen participation in, for example, the drafting and revising of legislation and budget-setting.

**Open Source**: Software for which the source code is available under an open licence. Not only can the software be used for free, but users with the necessary technical skills can inspect the source code, modify it and run their own versions of the code, helping to fix bugs, develop new features, etc. Some large open source software projects have thousands of volunteer contributors. The [Open Definition](http://opendatahandbook.org/glossary/en/terms/open-definition/) was heavily based on the earlier Open Source Definition, which sets out the conditions under which software can be considered open source.

**Public domain**: Content to which copyright does not apply, for example because it has expired, is free for any kind of use by anyone and is said to be in the public domain. CC0, one of the licences of Creative Commons, is a ‘public domain dedication’ which attempts so far as possible to renounce all rights in the work and place it in the public domain.

**Structured data**: All data has some structure, but ‘structured data’ refers to data where the structural relation between elements is explicit in the way the data is stored on a computer disk. [XML](http://opendatahandbook.org/glossary/en/terms/xml/) and [JSON](http://opendatahandbook.org/glossary/en/terms/json/) are common formats that allow many types of structure to be represented. The internal representation of, for example, word-processing documents or [PDF](http://opendatahandbook.org/glossary/en/terms/pdf/) documents reflects the positioning of entities on the page, not their logical structure, which is correspondingly difficult or impossible to extract automatically.

**CANADA'S OPEN DATA PRINCIPLES**Source: *open.canada.ca/en/open-data-principles*

**1. Completeness**

Datasets 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, unless there are Access to Information or Privacy issues. Metadata that defines and explains the raw data should be included, along with explanations for how the data was calculated.

**2. Primacy**

Datasets should come from a primary source. This includes the original information collected by the Government of Canada and available details on how the data was collected. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

**3. Timeliness**

Datasets released by the Government of Canada should be made available to the public in a timely fashion. Whenever feasible, information collected by the Government of Canada should be released as quickly as it is gathered and collected. Priority should be given to data whose utility is time sensitive.

**4. Ease of Physical and Electronic Access**

Datasets released by the Government of Canada should be as accessible as possible, with accessibility defined as the ease with which information can be obtained. Barriers to electronic access include making data accessible only via submitted forms or systems that require browser-oriented technologies (e.g., Flash, Javascript, cookies or Java applets). By contrast, providing an interface for users to make specific calls for data through an Application Programming Interface (API) make data much more readily accessible.

**5. Machine readability**

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 (e.g. CSV, XML). These files should be accompanied by documentation related to the format and how to use it in relation to the data.

**6. Non-discrimination**

Non-discrimination refers to who can access data and how they must do so. Barriers to use of data can include registration or membership requirements. Datasets released by the Government of Canada should have as few barriers to use as possible. Non-discriminatory access to data should enable any person to access the data at any time without having to identify him/herself or provide any justification for doing so.

**7. Use of Commonly Owned Standards**

Commonly owned standards refer to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of that company's program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. Removing this cost makes the data available to a wider pool of potential users. Datasets released by the Government of Canada should be in freely available file formats as often as possible.

**8. Licensing**

The Government of Canada releases datasets under the [Open Government Licence – Canada](http://open.canada.ca/en/open-government-licence-canada) agreement. The licence is designed to increase openness and minimize restrictions on the use of the data.

**9. Permanence**

The capability of finding information over time is referred to as permanence. For best use by the public, information made available online should remain online, with appropriate version-tracking and archiving over time.

**10. Usage Costs**

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

**USEFUL WEBSITES**

|  |  |
| --- | --- |
| Image result for city of vancouver | City of Vancouver Open Data Catalogue [*data.vancouver.ca*](https://vancouver.ca/your-government/open-data-catalogue.aspx) |
| Image result for government of canada | Government of Canada Open Government Portal [*open.canada.ca/en*](http://open.canada.ca/en) |
| Image result for Canadian Open Data Summit | CODS (Canadian Open Data Summit)[*opendatasummit.ca*](http://opendatasummit.ca/) |
| Image result for data bc | DataBC (Government of BC)[*data.gov.bc.ca*](https://data.gov.bc.ca/) |
| Image result for open data bc | Open Data BC (Open Data Society of BC)[*opendatabc.ca*](https://www.opendatabc.ca/pages/about) |
| European Commission logo | European Data Portal (see eLearning)[*europeandataportal.eu*](https://www.europeandataportal.eu/en/) |
| Image result for open data charter | Open Data Charter [*opendatacharter.net*](http://opendatacharter.net/) |
|  | Open Data Commons *[opendatacommons.org](https://opendatacommons.org)* |
| Image result for open knowledge network international | Open Knowledge Network International *[okfn.org](https://okfn.org/)* They produce the Open Data Handbook:[*opendatahandbook.org*](http://opendatahandbook.org/) |

**VPL RESOURCES**

|  |  |
| --- | --- |
| **VPL Events Guide** | www.vpl.ca/events |
| **Lynda.com** (*Open Data: Unleashing Hidden value*) | www.vpl.ca/lynda |
| **VPL Information Services**  | We are here to help!If you have any further questions, please ask us! Phone: 604-331-3603 / Email: info@vpl.ca |
| **VPL Subject Guides**Search *By Subject* > Technology & Computers | guides.vpl.ca |
| **VPL One to One Training** individual help on topics covered in VPL courses.  | Call 604-331-3603 to book an appointment. |
| **VPL Tech Cafe**A drop-in computer lab on Level 3 for practicing new skills with information staff assistance. | Central Library / 2:00 - 3:00 pm, Tuesdays & Thursdays |