



SUBMISSION FOR OPEN DATA STANDARD

Architecture, Standards and Planning Branch

Office of the CIO ● Province of BC

People ● Collaboration ● Innovation

Version: [Draft] V 0.5

December, 2010



Copyright notice

This document is a Provincial Architecture/Standard proposal and is copyright-protected by the province. Except as permitted under the applicable laws of the user's country, neither this draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured. Requests for permission to reproduce should be addressed to the Office of the CIO.

Document Overview

Submission For...	The first section of the document is an internal document intended to provide extra clarification and material to help the audience reviewing the standards. This porting is confidential and contains a copyright statement. This will not be published with the standard but will be kept as a reference document for the standard.
Standard	The Annex A or the last two pages of this document is the standard.



Table of Contents

1	Foreword.....	iv
2	Introduction	1
2.1	Standard Creation Process.....	1
3	Scope	1
4	Normative references.....	2
5	Informational References.....	2
6	Terms and Definitions	2
7	Requirements.....	2
8	Characteristics	3
8.1	Principles.....	3
8.1.1	Accessible	3
8.1.2	Machine Processable.....	3
8.1.3	Non-Discriminatory	3
8.1.4	Non-Proprietary Formats	4
8.1.5	License-Free	5
8.1.6	Primary	Error! Bookmark not defined.
8.1.7	Integrity	5
8.2	Guidelines	6
8.2.1	Straightforward.....	6
8.2.2	Timely.....	6
8.2.3	Independence of Outcomes.....	6
8.2.4	Completeness.....	6
8.2.5	Reviewable	7
8.3	Security Control Objectives	7
8.4	Information characteristics.....	7
9	Evaluation criteria.....	7
9.1	Enquiry Scope	7
9.2	Analysis/Acceptance	7
9.3	Response	7
9.4	Packaging.....	7
Annex A	8



1 Foreword

Quote by Te Whaingā Aronui, Governor-General of New Zealand:

“Now more than ever is there a very present need to bring information the Government holds on behalf of its people into the public domain so that it may be used in ways that stimulate innovation, generate cultural creativity, social interaction and dialogue, while also kick starting economic growth.”

2 Introduction

Open Data is a philosophy and practice requiring that certain data are freely available to the public, without restrictions from copyright, patents or other mechanisms of control. It has a similar ethos to a number of other "Open" movements and communities such as open source and open access. With this new philosophy come many challenges and to move forward and address those challenges we can start by defining and creating some foundational building blocks. These building blocks or foundational standards will not only address some current needs but will also start to position the Government well for the Open Data movement.

The Open data Standard is one of a larger set of documents that support the Open Data Initiative. Open Data is a big initiative in government and a solid definition is needed. This standard provides a common understanding and helps organizations get a handle on what Open Data is. There will be other documents created to support the Open Data movement covering topics such as: implementation, best practices, content safe dataset and security measures to name a few.

2.1 Standard Creation Process

The characteristics outlined within this document are based on the publications provided by the wiki.opengovdata.org with further research from work done by Australia, New Zealand, UK and USA. Refer to section 5 Informational References.

The rationale and examples behind the characteristics were developed with help from the internal Open Data Initiative group as a means to provide clarification.

3 Scope

This document applies to Ministries exposing Open Data. It:

- Specifies characteristics that define Open Data

This document does not apply to:

- Implementations
- Schemas
- Privacy
- Products and or Services

4 Normative references

Standards Manual

Information Management Standards

Though not within the scope of this standard the following related material was considered.

5 Informational References

- <http://wiki.opengovdata.org/index.php?title=OpenDataPrinciples>
- <http://perens.com/OpenStandards/Definition.html>
- The [American Library Association's "Key Principles of Government Information"](#) and the [Open Knowledge Foundation's Open Knowledge Definition](#) outline sets of principles similar to those contained in this document.
- Australia is in the process of developing and/or adopting standards in this area:
<http://www.csiro.au/solutions/ps1u1.html>
- New Zealand Government is promoting open data and is also using the same principles:
<http://data.govt.nz>
- Draft New Zealand Government Open Access and Licensing framework (NZGOAL):
http://epsiplatform.com/layout/set/print/content/download/28721/381845/version/1/file/New+Zealand+open+licensing+Draft_NZGOAL%5B1%5D.pdf

6 Terms and Definitions

For the purposes of this document, the following acronyms apply:

<u>Term</u>	<u>Definition</u>

7 Requirements

For provincial data to be considered “Open” the content must possess ALL characteristics identified as principles.

8 Characteristics

There are a number of characteristics that define Open Data. These characteristics have been divided into two groups, principles and guidelines. The principles are strong measurable characteristics that are essential to creating an open dataset, while the guidelines add to the value and usability of the dataset.

8.1 Principles

8.1.1 Accessible

Open Data must be clearly and easily visible through open protocols to the widest range of users for the widest range of purposes. There must be no requirement for registration. Full anonymity and the absence of authentication offers the greatest opportunity for engagement.

Rationale: Open Data should be accessible without restriction or pre-approvals, should not be hard to find or difficult to download. If any of these apply, then it might as well be considered inaccessible.

Example: The inability to find and access content gives the impression that there is an attempt to hide it. Data sources should be directly accessible via URL or URI.

8.1.2 Machine Processable

Open Data must be reasonably structured to allow automated intake and processing.

Rationale: Although the benefit of information is to people, machines play a fundamental role in the ability to add value. When looking to capitalize on information, we must consider ways we can take advantage of contemporary or open tools to make the information more valuable.

The computer is also a tool on the consumption side of information. Its uses can be summarized as searching, sorting and transforming information. Simple examples are a web search to find information, sorting a table to understand expenses, and creating charts and graphs as visual representations of numerical information.

8.1.3 Example: Poorly tabulated columns taken from reports meant to be read by people – they make it very hard for tools like spreadsheets to parse the data usefully. **Non-Discriminatory**

Open Data must be available to everyone with no requirement for registration or restriction to certain classes of users. Open data must not be restricted from being used in a specific field of endeavor (e.g., in business or military research).

Rationale: Open Data should not be open only to a select or approved few, but to everyone.

Example: The value of Open data is multiplied when there is multidisciplinary access. A health services expert could provide new and innovative insights by accessing environmental or business information.

8.1.4 Non-Proprietary Formats

Open data must be published in formats that are Open Standards(Open Dataset Physical Format Tier 1¹). Open Standards: are available, maximize end-user choice, contain no royalty, do not discriminate, provide extension or subset use and may employ license terms that protect against subversion of the standard.

Rationale: Reliance on a single vendor's tools and formats increases the likelihood of obsolescence which may limit longer term access to the Data. Though all formats are at risk, formats held by private interests are more susceptible than others.

Open standards enable consumer choice in the marketplace, encourage competition (which drives quality up and prices down), and eliminate cost barriers for those with limited financial means.

“The rationale for promoting open standards is technical, economic, and political. The technical rationale is to promote maximum interoperability to enable the universal and efficient exchange of information among technologies, regardless of manufacturer or geographical region. The economic rationale is to foster an environment of competition among products based on the standard and provide fertile ground for national entrepreneurship and innovation. By an economic definition, open standards inherently prevent anticompetitive and monopolistic practices, promote rather than restrict global trade, and provide opportunities for emerging markets to enter global information technology markets. The political rationale for open standards is to contribute to efficient and

¹ Open Data Physical Format Standard; internal standard developed by the OCIO

accountable government functions ranging from issues of political representation, disaster response, and the archiving of public documents”²

Example: A vendor may change or remove its product from the market, which may make data in that vendor’s format effectively inaccessible.

8.1.5 License-Free

Due to the nature of their purpose Open Data must not require users to seek a special license, approvals or pay fees for access or use. “Acknowledgement of Source” is optional.

Rationale: Open Data encumbered by special licensing, royalty, and patent restrictions presents problems and limits usability. Additionally if archivists and the public must pay licensing fees to access public data, the data is no longer public in any practical sense.

8.1.6 Example: Special licenses or fees would restrict access and usability by less privileged users.**Primary**

The publishing of open data must be approved by the custodian as defined within the Roles and Responsibility Guideline³.

Rationale: This characteristic ensures that no organization other than the one responsible for the content’s development can authorize open publishing of Open Data.

Example: Ministry of Health Services cannot publish Ministry of Agriculture Open Data as they are not the authoritative entity.

8.1.7 Integrity

Organizations publishing Open Data must ensure adequate security controls exist within the environment to prevent a breach of the integrity of the information.

Rationale: The public must be able to trust that the information is true and has not been falsified.

Example: An organization posts a dataset on mineral deposits which becomes corrupted resulting in massive losses by mining companies exploring in the wrong areas.

² http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1629833

³ Roles and Responsibility Guideline; developed by the Data Administrator Advisory Committee

8.2 Guidelines

8.2.1 Straightforward

Material should not be intentionally named or configured in any way that would influence any analysis arising from their use.

Rationale: The public should not be misled or drawn to conclusions by naming or by location.

Example: Open Data file “Why Dams Cause Fish Disease.CSV” would not follow the straightforward characteristic. Nor would any Data stored in a directory called “Negative Health Impact Information”.

8.2.2 Timely

New or updated Open Data should be made available as quickly as necessary to preserve the value and relevancy of the data.

Rationale: The primary purpose of this characteristic is to ensure that Open Data are not exposed only when it is convenient or delayed in such a way that reduces the potential outcomes. Publishing periods do not have to be fixed.

Example: Content related to a recent flood is more publicly useful if it was published close to the event.

8.2.3 Independence of Outcomes

Data should be created and published independent of potential outcomes.

Rationale: The potential to generate negative (to the province) findings based on Open Data is always a potential risk. This risk should not pose barriers to publishing data.

Example: A dataset or a publicly generated mash up of multiple data is published that exposes a failure or oversight within government that would not have otherwise been detected.

8.2.4 Completeness

Datasets should contain a full set of information.

Rationale: Data with missing rows reduces the value for analysis or re-use because it only provides a partial view of the subject.

Example: If your data contains attendance from all years except 2001 and 2005 because it was particularly bad in those years and the school doesn't want anyone to see that.

8.2.5 Reviewable

Each organization should have a point of contact to respond to Public questions or issues on their Open Data.

Rationale: The Public needs to have the ability to resolve any suggestions or issues they may have regarding access and usage to individual datasets.

8.3 Security Control Objectives

8.4 Information characteristics

Not in scope

9 Evaluation criteria

9.1 Enquiry Scope

The material contained in this document was open to members of the ASRB, DAAC and OpenDataBC Google Group for comment.

9.2 Analysis/Acceptance

[who did initial acceptance of material]

9.3 Response

[how was acceptance accepted]

9.4 Packaging

[Other document structure information].

Annex A.

IM/IT Architecture & Standards Manual STANDARD Office of the Chief Information Officer Province of British Columbia	Effective Date: YYYY-MM-DD Scheduled Review: Annual Last Updated: YYYY-MM-DD Last Reviewed: YYYY-MM-DD
	Type: Information
X.0 Open Data	
X.Y.0 Open Data Standard	
Keywords: Open, data, Open Government	

Description of Standard:

This standard describes the characteristics that define Open Data. Open Data is machine processable information used to expose government data thereby engaging the public to stimulate innovation, generate cultural creativity, social interaction and dialogue.

Requirements: For provincial data to be considered “Open” the content must possess ALL characteristics identified as principles.

Characteristics:

There are a number of characteristics that define Open Data. These Characteristics have been divided into two groups, principles and guidelines. The principles are strong measurable characteristics that are essential to creating an open dataset, while the guidelines add to the value and usability of the dataset.

Principles	Accessible	Open Data must be clearly and easily visible through open protocols to the widest range of users for the widest range of purposes. There must be no requirement for registration. Full anonymity and the absence of authentication offers the greatest opportunity for engagement.
	Machine Processable	Open Data must be reasonably structured to allow automated intake and processing.
	Non-Discriminatory	Open Data must be available to everyone with no requirement for registration or restriction to certain classes of users. Open data must not be restricted from being used in a specific field of endeavor (e.g., in business or military research).

	Non-Proprietary Format	Open Data must be published in formats that are Open Standards (Open Dataset Physical Format Tier 1 ⁴). Open Standards: are available, maximize end-user choice, contain no royalty, do not discriminate, provide extension or subset use and may employ license terms that protect against subversion of the standard.
	License-Free	Due to the nature of their purpose Open Data must not require users to seek a special license, approvals or pay fees for access or use. “Acknowledgement of Source” is optional.
	Primary	The publishing of Open Data must be approved by the custodian as defined within the Roles and Responsibility guidelines ⁵ .
	Integrity	Organizations publishing Open Data must ensure adequate security controls exist within the environment to prevent a breach of the integrity of the information.
Guidelines	Straightforward	Material should not be intentionally named or configured in any way that would influence any analysis arising from their use.
	Timely	New or updated Open Data should be made available as quickly as necessary to preserve the value and relevancy of the data.
	Independence of Outcomes	Datasets should be created and published independent of potential outcomes.
	Completeness	Datasets should contain a full set of information.
	Reviewable	Each organization should have a point of contact to respond to Public questions or issues on their Open Data.

Where to Apply This Standard

For machine processable Open Data exposed to the public.

Authority and Exemptions

There are no exemptions for this standard either the data meets all requirements and is Open Data or it does not meet all the requirements and is not Open Data.

Metrics and Enforcement

Ministry contact designated as per the Reviewable characteristic.

References

- Standards Manual:

⁴ Open Data Physical Format Standards; internal standard developed by the OCIO

⁵ Roles and Responsibility Guideline; developed by the Data Administrator Advisory Committee



- Information Management Standards: Though not within the scope of this standard the following related material was considered.
- Informative References: <http://wiki.opengovdata.org/index.php?title=OpenDataPrinciples>
The American Library Association's "Key Principles of Government Information" and the Open Knowledge Foundation's Open Knowledge Definition outline sets of principles similar to those contained in this document.

Australia is in the process of developing and/or adopting standards in this area:

<http://www.csiro.au/solutions/ps1u1.html>

New Zealand Government is promoting open data and is also using the same principles:

<http://data.govt.nz>

Draft New Zealand Government Open Access and licensing framework (NZGOAL):

http://epsiplatform.com/layout/set/print/content/download/28721/381845/version/1/file/New+Zealand+open+licensing+Draft_NZGOAL%5B1%5D.pdf

Definition of Open Standards: <http://perens.com/OpenStandards/Definition.html>

Additional Information

Contact

Architecture and Standards Branch

Office of the CIO