

Data Accessibility Benchmark

Organizational Self-Assessment Tool

Rationale for Development of the Benchmark

The fall 2010 Report of the Commissioner of the Environment and Sustainable Development, Office of the Auditor General of Canada's stated: "Acquiring reliable environmental data and information is the first step in addressing the most pressing environmental priorities. Solid, objective, and accessible information is essential to identify and respond to the quickening pace and complexity of environmental change, in Canada and globally".

Since 1992, the Atlantic Coastal Zone Information Steering Committee has been encouraging the accessibility to and sharing of data and information to support integrated coastal and ocean management. This data and information is held by a diverse set of data providing organizations in government, universities, industry, non-governmental organizations and communities.

The data accessibility self-assessment tool is for use in these organizations to develop a benchmark to monitor progress in improving an organization's policies and procedures for providing effective and easy access to the data and information held by them.

Benchmark Components: Principals and Evaluation Ladders

The data accessibility principles used in the benchmark are an aggregation of four documents: the Organisation for Economic Co-Operation and Development's Principles and Guidelines for Access to Research Data from Public Funding¹; the Government of Canada's Operating Principles for the Open Government site²; the United States Memorandum on Open Data Policy³; and the G8 Charter for Open Data⁴. See summary list of principles on page 3.

For each principle one or more evaluation ladders have been developed with questions designed to assist the user in positioning their organization's policies and procedures on the ladder.

¹ OECD Principles and Guidelines for Access to Research Data from Public Funding, <http://www.oecd.org/science/sci-tech/38500813.pdf>

² Operating Principles for the Open Government site <http://data.gc.ca/eng/open-data-principles>

³ Open Data Policy-Managing Information as an Asset
<http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

⁴ G8 Open Data Charter and Technical Annex <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

Guide for using the Data Accessibility Benchmark

The benchmark is designed to be used repetitively to measure improvement in the effectiveness and ease of accessing data and information by users both within and outside the organization. Use of the benchmark tool on an annual basis would be appropriate.

The aggregate score obtained from the benchmark ladders is less important than the individual scores for each ladder. It is the relative scoring on each of the principles / evaluation ladders that will be most useful in focussing initiatives to improve the overall score.

The tool could also be used to separately score different business units or data dissemination infrastructures within an organization especially if responsibility for data management and data accessibility is a distributed responsibility.

Organizations may want to weight their scores based upon internal priorities e.g. data management policies and plans may be considered a higher priority than user support. It is important that the weightings be explicit and consistently used in each use of the tool.

Wording to assist users in positioning their organization on the ladders is only provided for scores '5', '3', '1' and '0'. Score '2' or '4' if a description for level 1 or 3 is less than the situation in your organization but 3 or 5 describes a situation which your organization has not yet attained.

Example: Choose your organization's position on the ladder for principle #11 'Evaluation'.

Principle 11: Evaluation

Periodic evaluation by user groups

5 - A regular program of evaluation is in place.

3 - Ad hoc evaluation is in place.

1 - Crisis based evaluation is conducted.

0 - No evaluation.



Data Accessibility Principles:

1. Open Data by Default
2. Completeness
3. Primacy
4. Timeliness
5. Ease of Physical and Electronic Access
6. Non-discrimination
7. Licencing
8. Permanence
9. Usage Costs
10. Supporting Use
11. Evaluation

Acknowledgements

Funding for the development of the Data Accessibility Benchmark was provided by the Adaptation Platform Program of the Climate Change Impacts and Adaptation Division of Natural Resources Canada.

The tool was developed in collaboration with Dr. Bertrum MacDonald, School of Information Management, Dalhousie University.

Tool testing has been assisted by Robert Branton, Emeritus Scientist, Ocean Tracking Network, Dalhousie University.

Data Accessibility Benchmark – Self Assessment

1 Open Data by Default

a. Formal Policies and Limitations

The data providing organization has adopted policies which presume open data to the extent permitted by law and subject to privacy, confidentiality, security, and intellectual property rights. This includes other valid restrictions such as the restriction to access to data on the location of biological and archeological and similar resources for the sake of conservation.

	Formal, approved policies with the presumption that data is open.		Open data policies for only some of the data without presumption data is open.		Open data policies are under development.	There is no evidence of an open data policy.	n/a
	5	4	3	2	1	0	
✓							
	Limitations or restrictions to the open data are clearly defined and supported by legislation or policy.		Limitations or restrictions to the open data are not explicit or vague.		Limitations or restrictions to the open data are under development.	There is no policy on open data limitations or restrictions.	n/a
	5	4	3	2	1	0	
✓							

b. Ease of access to information on policies, etc.

Information is disseminated online regarding the data providing organization’s data access policies, data collection methodology, standards employed, and publishing processes.

	Policies, etc. are prominently displayed on websites and are easily found.		Policies, etc. are hard to find on the website or must be asked for by e-mail.		On-line access to policies, etc. is under development.	There is no on-line access to policies, etc.	n/a
	5	4	3	2	1	0	
✓							

2. Completeness

a. Level of Detail, Data

The data providing organization releases datasets that are as complete as possible at the finest level of granularity available.

	Datasets are complete and detailed.		Limited access to full detailed data sets.		Only partially detailed data sets, or incomplete data sets are released.	Only generalized, summaries, large scale data sets are released.	n/a
	5	4	3	2	1	0	
✓							

b. Metadata

The data providing organization releases metadata that defines and explains the data, explains how derived data was calculated, and states the standards employed for the data, metadata and quality assessment.

	Complete and readily available metadata using an acceptable standard for every data set.		Some data sets have complete metadata and some don't or metadata is hard to find. In-house or mixed standards are used.		Minimal metadata with no standard.	No metadata.	n/a
	5	4	3	2	1	0	
✓							

3. Primacy

The data providing organization releases primary source data; this includes the original raw data collected and metadata on how the data was collected.

	Raw / primary data is released with collection metadata.		Raw / primary data is released but without collection metadata.		Limited raw / primary data is released.	No raw / primary data is released.	n/a
	5	4	3	2	1	0	
✓							

4. Timeliness

The data providing organization releases datasets to the public in a timely fashion to maintain the value of the data, with priority given to data whose utility is time sensitive.

	Policy states, and in practise, datasets are released in a timeframe that maintains the full value of the data.		Policy states, and in practise, release of datasets are delayed for the shortest possible time for clear and documented reasons.		Policy allows for, and in practise, the release of datasets in a delayed timeframe.	No regard is given to the timeframe of release of datasets either in policy or in practise.	n/a
	5	4	3	2	1	0	
✓							

Note: High scores in the timeliness principle ladder do not imply real time access or near real time access is necessary. It is intended to measure the timeframe of usefulness of the data which for only some types of data may require real time or near real time access e.g. measured tidal height for safe navigation. This ladder may need to be evaluated for more than one type of data depending on the principle uses of the data.

5 Ease of Physical and Electronic Access

a. Internet Access

The data providing organization releases datasets using the internet.

	Datasets are easily accessible on the internet.		Only some datasets are not accessible on the on the internet.		Only some datasets are accessible on the internet.	Datasets are not accessible on the internet; they need to be asked for.	n/a
	5	4	3	2	1	0	
✓							

b. Standard Formats

The data providing organization releases datasets in open, freely available formats and web-based services that conform to widely accepted standards.

	Datasets are accessible as standard internet services, e.g., OGC for spatial data.		Datasets are accessible in standard formats.		Datasets are accessible in a mixture of standard and widely used proprietary formats.	Datasets are not in standard or widely used proprietary formats.	n/a
	5	4	3	2	1	0	
✓							

Note: This principle allows for the use of proprietary standards for access but requires that the formats and services be widely accepted. Higher scores are reserved for international accepted standards like OGC standards which enable easier interoperability. The term widely accepted may be different for different types of data e.g. ocean observation data, satellite imagery, etc.

6. Non-discrimination

The data providing organization provides barrier free and non-discriminatory access to data at any time without having to provide identification or justification for access.

	Datasets are accessible to everyone without disclosing reason for use or identity.		Datasets are accessible to everyone but reason for use or identity must be disclosed.		Datasets are accessible to everyone but accreditation (i.e. proof of identity) is needed.	Accessibility to datasets may be declined for some uses.	n/a
	5	4	3	2	1	0	
✓							

7. Licencing

The data providing organization releases datasets under an open licence with minimal restrictions pertaining only to intellectual property, personally-identifiable and sensitive information.

	An open license consistent with constraints permitted by Creative Commons Attribution (CC BY) license.*		An open license with some restrictions beyond attribution.		Open licensing policy is under development.	Restrictive licensing, formal contract required.	n/a
	5	4	3	2	1	0	
✓							

* <https://creativecommons.org/licenses/> This widely used license lets others distribute, remix, tweak, and build upon the data providing organization's work, even commercially, as long as credit is given for the original creation. This is the most accommodating of licenses offered in the suite of **Creative Commons** licenses and is recommended for maximum dissemination and use of licensed materials.

8. Permanence

a. Infrastructure

The data providing organization has established formal administrative responsibility for the long-term sustainability of the infrastructure required for data access and for ensuring data are effectively preserved, managed, archived and made accessible, permanently where long term retention has been determined to be necessary

	Long-term (>10 yrs) infrastructure future plan in place.		Short-term (< 5 yrs) infrastructure plan in place.		Infrastructure planning under development.	No infrastructure planning in place.	n/a
	5	4	3	2	1	0	
✓							

b. Data management policies and plans.

	Data management policies and plans in place are followed and meet accepted standards.		Data management policies and plans in place but non-standard or not followed.		Data management policies and plans are being developed.	There are no data management policies and plans in place.	n/a
	5	4	3	2	1	0	
✓							

c. Retention Protocol and Assessment

The data providing organization uses accepted retention protocols and conducts periodic cost-benefit assessments to develop and refine retention protocols.

	Regular cost benefit analysis is conducted to determine data set retention.		Ad hoc cost benefit analysis is conducted to determine data set retention.		Cost benefit analysis is conducted only on a crisis basis to determine data set retention.	No assessment for data retention is conducted.	n/a
	5	4	3	2	1	0	
✓							

Note: This permanence principle ladder does not require permanently retaining data but does measure the organization's compliance with developing and implementing policies and procedures for establishing retention periods for data. The readiness of data for dissemination may also be different depending on the demand for or age of the data (i.e. accessible on-line vs accessible by request) as long as the method of accessibility has been determined through the application of carefully developed retention policies.

9. Usage Costs

The data providing organization releases datasets free of charge.

	Datasets are provided free of charge.		Datasets are provided at the cost of dissemination. (minimal)		Datasets are provided at a cost that includes cost recovery greater than the cost of dissemination.		Datasets are provided for a cost greater than full cost recovery.	n/a
	5	4	3	2	1		0	
✓								

10. Supporting Use

The data providing organization has developed policies and programs to assist users with data use and provides a point of contact to respond to complaints.

	Policies, programs and easily found contacts have been developed to assist users and to receive and evaluate complaints.		A specific point of contact to answer questions and receive complaints about the use of the data is clearly defined and easily found.		General contacts for the organization are defined and easily found.		Contacts are not given or general contacts are hard to find.	n/a
	5	4	3	2	1		0	
✓								

11. Evaluation

The data providing organization periodically evaluates the performance of data access arrangements by user groups.

	A regular program of evaluation is in place.		Ad hoc evaluation is in place.		Crisis based evaluation is conducted.		No evaluation.	n/a
	5	4	3	2	1		0	
✓								