



COINAtlantic Data Accessibility Self-Assessment Tool

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 five 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; the G8 Charter for Open Data⁴ and the Guidelines on FAIR Data Management in Horizon 2020 of the European Commission Directorate-General for Research and Innovation⁵. 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.

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.

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>

⁵ Guidelines on FAIR Data Management
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

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.

5

4

3 - Ad hoc evaluation is in place.

3

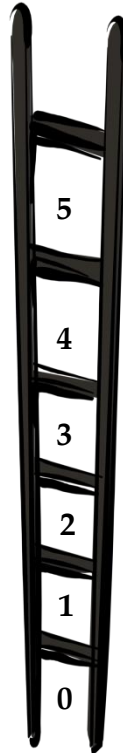
1 - Crisis based evaluation is conducted.

2

1

0 - No evaluation.

0



Data Accessibility Principles:

1. Open Data by Default
2. Completeness
3. Standard identification mechanism
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.

The following organizations have provided comments to improve the CDAST:

1 October 2015: Northeast Coastal and Ocean Data Partnership, "Achieving More Effective and Easy Access to Coastal and Ocean Data: A Self-Assessment Tool for Data Providing Organizations". Watch Video: <https://www.youtube.com/watch?v=TkfpHDvZ1VY>

8 October 2014: NRCAN Adaptation Platform Webinar Series, Managing Climate Change Information for Adaptation. 13 January 2015: NOAA Data Management Integration Team, "Self-Assessment and Data Management Tools for Data Providing Organizations"

April 2015: ICAN 7 workshop / [CoastGIS 2015 Presentation](#)

18 January 2016: CODATA Webinar (International Council for Science: Committee on Data for Science and Technology) - <https://zenodo.org/record/44966#>

EIUI Research group, School of Information Management, Dalhousie University

COINAtlantic Data Accessibility Self-Assessment Tool

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 apply to only some of the data. See evaluation scale for acceptable limitations below) .		Open data policies are recognized as important and are under development.	There is no evidence of an open data policy or development 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 evidence of 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 and understood.		Policies, etc. are hard to find on the website or must be asked for by e-mail or are not easily understood.		Policies exist but on-line access to policies, etc. is recognized as important and under development.	There is no evidence that policies exist.	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.

	All datasets are as complete as possible and as detailed as the original data set.		Some detailed data sets have limited access.		Incomplete or less than fully detailed data sets are released.		Only generalized or low resolution summaries, or small 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 accessible metadata using an acceptable standard for every data set.		Some data sets have complete and accessible metadata and some don't or metadata is hard to find. In-house or mixed standards are used.		Minimal accessible metadata with no standard.		No metadata accessible.	n/a
	5	4	3	2	1	0		
✓								

3. Standard identification mechanism

The data providing organization makes data identifiable and locatable using a standard identification mechanism.(e.g. Digital Object Identifier (DOI) <https://www.doi.org/index.html>)

	A unique standard identification is generated for each data set and is easily locatable on the internet using the identifier.		A unique standard identification is generated for some data sets and is easily locatable on the internet using the identifier.		A unique standard identification is generated for few data sets or is not is easily locatable on the internet using the identifier.	No standard identification is generated.	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 in terms of time if applicable, 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		Only some datasets are not accessible		Only some datasets are	Datasets are not accessible on the	n/a
--	--------------------------------	--	--	--	-------------------------------	------------------------------------	-----

	on the internet.		on the on the internet.		accessible on the internet.	internet; they need to be asked for.	
	5	4	3	2	1	0	
✓							

b. Standard Formats

The data providing organization releases datasets in open, freely available formats that conform to widely accepted standards and easily useable in widely used software.

	Datasets are distributed in internationally recognized formats for the discipline involved.		Data sets are distributed in recognized standard formats.		Datasets are distributed in a mixture of standard and widely used proprietary formats.	Datasets are distributed 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 be widely accepted. Higher scores are reserved for international accepted standards especially if discipline specific which enable easier interoperability. The term widely accepted may be different for different types of data e.g. ocean observation data, satellite imagery, etc.

c. Standard Services

The data providing organization supports web-based services that conform to widely accepted standards.

	Web-based services conform to standard open internet services, and provide access to data that can be styled or processed analytically e.g., OGC WFS for spatial data.		Web-based services conform to widely-used proprietary standards for internet services, and provide access to data that can be styled or processed analytically e.g., ARCGIS REST for spatial data.		Web-based services provide access to visualizations / map images of data e.g. OGC WMS	Web-based services do not comply with standard or widely used proprietary standards.	n/a
	5	4	3	2	1	0	
✓							

d. Interoperability

The data providing organization provides policies, processes and services that maximize the likelihood of the reuse of data by other organizations / researchers and the integration of the data with other data for analytical purposes.

	Internationally approved discipline specific content standards and vocabularies are used including information on appropriate uses of the data and compliance with available (open) software applications.		Content standards and controlled vocabularies are used and well described including information on appropriate uses of the data.		The importance of content standards and controlled vocabularies is recognized and under development.	No content standards or controlled vocabularies are used.	n/a
	5	4	3	2	1	0	
✓							

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 or Canada’s Open Government License.*		An open license with some restrictions beyond attribution.		Open licensing policy is recognized as important and 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.

Canada’s Open Government License

The Information Provider grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the Information, including for commercial purposes, subject to the terms below.

You are free to:

Copy, modify, publish, translate, adapt, distribute or otherwise use the Information in any medium, mode or format for any lawful purpose.

You must, where you do any of the above:

Acknowledge the source of the Information by including any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence.

If the Information Provider does not provide a specific attribution statement, or if you are using Information from several information providers and multiple attributions are not practical for your product or application, you must use the following attribution statement:

“Contains information licensed under the Open Government Licence – Canada”.

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 is recognized as important and is under development.	No infrastructure planning in place.	n/a
	5	4	3	2	1	0	
✓							

b. Data management policies and plans.

	Data management and quality assurance policies and plans in place are followed and meet accepted standards and meeting requirements for certified repositories.		Data management and quality assurance policies and plans in place but not sufficient to meet the requirement for certified repositories.		Data management and quality assurance policies and plans are recognized to be important and are being developed.	There are no data management or quality assurance policies and plans in place or planned.	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 or minimal cost.

	Datasets are distributed free of charge.		Datasets are distributed at the cost of dissemination. (minimal)		Datasets are distributed at a cost that includes cost recovery greater than the cost of dissemination.		Datasets are distributed 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 or comments.

	Policies, programs and easily found contacts have been developed to assist users and to receive and evaluate comments or complaints.		A specific point of contact to answer questions and receive comments or 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		Ad hoc evaluation		Crisis based evaluation		No	n/a
--	----------------------	--	-------------------	--	-------------------------	--	----	-----

Data Accessibility Benchmark Version 2 08 May 2017

	evaluation is in place.		is in place.		is conducted.	evaluation.	
	5	4	3	2	1	0	
✓							

Glossary:

- Open Data** Open Data is the idea that data should be freely available for everyone to access, use and republish as they wish, published without restrictions from copyright, patents or other mechanisms of control. Open Government Partnership
<http://www.opengovguide.com/topics/open-government-data/>
- Completeness** A complete data sets has all parts or elements to permit reuse for research and / or analytical purposes.
- Metadata** the information we create, store, and share to describe things, allows us to interact with these things to obtain the knowledge we need. The classic definition is literal, based on the etymology of the word itself—metadata is “data about data.” National Information Standards Organization (NISO) UNDERSTANDING METADATA
- Standard Identification** A unique identifier for an entity is provided by an identifier registry. An identifier registry is a compilation of unique identifiers, with some information on each item so identified, registered through an organization which maintains it. The registry follows a syntax specification for the identifiers (typically a formal standard), and the agency provides a means of registering identifiers. e.g. Digital Object Identifier
<https://www.doi.org/index.html>
- Open formats** Machine readable formats that have a specification that is publically available enabling any application developer to read or write data into and from that format. Standardized open formats are those open formats whose specification has been endorsed by a national or international standards setting body.
- Interoperability** the ability of two or more systems or components to exchange information and to **use** the information that has been exchanged IEEE Glossary 1991
- Open license** an open license is one which grants permission to access, re-use and redistribute a work with few or no restrictions
- Certified Repositories** a trusted digital repository has “a mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future”. A certified repository is a trusted digital repository that has undergone a recognized audit and certification process. The Center for Research Libraries 2007.
- Retention protocols** policies and processes that determine how long a data set is retained by the organization. Sometimes the length of time is set by regulation or legislation.

Evaluation Evaluation is the systematic assessment of the design, implementation or results of an initiative for the purposes of learning or decision-making.