Open government data and environmental science: a federal Canadian perspective

Dominique G. Roche[a]*, Monica Granados[b], Claire C. Austin[b], Scott Wilson[c], Gregory M. Mitchell[c], Paul A. Smith[c], Steven J. Cooke[a], and Joseph R. Bennett[a]

[a]Canadian Centre for Evidence-Based Conservation, Department of Biology and Institute of Environmental and Interdisciplinary Sciences, Carleton University, Ottawa, ON K1S 5B6, Canada; [b]Science and Technology Strategies Directorate, Environment and Climate Change Canada, Gatineau, QC K1A 0H3, Canada; [c]Canadian Wildlife Service, Environment and Climate Change Canada, Gatineau, QC K1A 0H3, Canada

*dominique.roche@mail.mcgill.ca

Abstract

Governments worldwide are releasing data into the public domain via open government data initiatives. Many such data sets are directly relevant to environmental science and complement data collected by academic researchers to address complex and challenging environmental problems. The Government of Canada is a leader in open data among Organisation for Economic Co-operation and Development countries, generating and releasing troves of valuable research data. However, achieving comprehensive and FAIR (findable, accessible, interoperable, reusable) open government data is not without its challenges. For example, identifying and understanding Canada's international commitments, policies, and guidelines on open data can be daunting. Similarly, open data sets within the Government of Canada are spread across a diversity of repositories and portals, which may hinder their discoverability. We describe Canada's federal initiatives promoting open government data, and outline where data sets of relevance to environmental science can be found. We summarize research data management challenges identified by the Government of Canada, plans to modernize the approach to open data for environmental science and best practices for data discoverability, access, and reuse.

Key words: conservation, data sharing, ecology, FAIR data, Government of Canada, public data archiving, science policy

Introduction

Environmental science increasingly relies on combining disparate data from multiple sources to answer questions involving, for example, broad geographical areas, multiple time periods, many species, and complex systems or issues (Barone et al. 2017; Cheruvelil and Soranno 2018; Poisot et al. 2019). The resulting data sets (sometimes leading to Big Data) may generate deeper and more robust insights than spatially and temporally restricted data, and have the potential to better inform environmental management and conservation policy (Huang and Qiao 2011; McClenachan et al. 2012; Wolkovich et al. 2012). Concurrent with the technological advances in data sharing (Hipsley and Sherratt 2019), there is a growing sentiment among scientists and governments that open and FAIR (findable, accessible, interoperable, reusable) data (Appendix 1) increase the speed, scope, quality, and value of scientific research (Molloy 2011; Borgman 2012). Key benefits of open research data include: preserving information to ensure the continuity of scientific discovery, allowing the verification of research results by improving transparency and reproducibility, maximizing returns on public

investments in research, enabling data synthesis and reuse to advance research and innovation, and facilitating education and public outreach and participation (Roche et al. 2014). Consequently, a growing number of funders and publishers are adopting policies that require scientists to share their data and computer code on public platforms upon publishing their results (Roche 2016; Nature Editorial 2018; Renaut et al. 2018). These policies are gaining traction across the natural and social sciences, resulting in greater data deposition rates by academics into publicly accessible repositories (Vines et al. 2013; Science Digital 2019).

Such policies have strong potential to incentivize data sharing by academic researchers. However, governments also conduct research and amass rich environmental data sets and in some cases are less beholden to funding agencies. For example, over 1.1 million scientists and engineers work for the United States government (National Science Foundation 2015). In Canada, the federal government employs approximately 25 000 scientists in natural sciences and engineering alone (Statistics Canada 2012). Several science-based departments and agencies (SBDAs) within the Canadian government generate data of direct relevance to environmental management and biodiversity conservation, often through long-term research and monitoring programs (Table 1). Between 2012 and 2015, federal SBDAs published 17 695 scientific journal articles, representing 6.8% of all peer-reviewed publications by Canadian scientists (Open Science Metrics Working Group 2018). Canada is also fourth among Organisation for Economic Co-operation and Development (OECD) countries in terms of the spending on government research as a fraction of gross domestic product (OECD 2019b).

Government employees, as is the case for academic researchers, must share the data necessary to reproduce their results when publishing in journals requiring open data. However, government employees also regularly publish using a diversity of other media, including reports, specialized publications, and government periodicals that are not subject to academic publishers' open data policies. Many federal SBDAs (e.g., Statistics Canada, the Department of Fisheries and Oceans, and Environment and Climate Change Canada) have been publicly releasing comprehensive data sets for decades (Environment and Climate Change Canada 2019). However, all departments and agencies must now participate in open data in line with Canada's commitments to international agreements and important changes in government policy and directives (Fig. 1) (Open Science Metrics Working Group 2018).

In general, recent progress on open data in the academic and government sectors is encouraging. However, the implementation of new open data policies poses important technical and cultural challenges that have yet to be resolved, both in government and in academic institutions. For example, data management practices vary considerably among disciplines, institutions, and research groups, slowing the adoption of data sharing standards and best practices (Hipsley and Sherratt 2019; Poisot et al. 2019). As a result, researchers often struggle to share their data in formats that are readily understandable and reusable by third parties (Poisot et al. 2019). For example, a 2015 study found that over 50% of open data sets shared by ecologists and evolutionary biologists were unusable (Roche et al. 2015). Additionally, the ever-increasing volume and complexity of data sets in environmental science requires that institutions allocate considerable resources to help scientists curate, store, and share their data effectively. Presently, most research data are shared in a decentralized fashion, across a diversity of repositories and portals (Open Science Metrics Working Group 2018; Privy Council Office Government of Canada 2018), creating challenges for their discovery, integration, and reuse. In addition, restricted data (e.g., personally identifiable information, confidential business information, intellectual property) and sensitive data (e.g., the location of commercially valuable species or species at risk) (e.g., Lennox et al. 2020) require an additional layer of data management to balance access and security.

Along with 34 other OECD countries, Canada has established formal requirements for publicly sharing government data. In 2019, Canada ranked third among OECD members on the basis of

Table 1. Science-based departments and agencies (SBDAs) of the Government of Canada that collect environmental data of relevance to biodiversity conservation and environmental management.

Department or agency	Mandate and types of environmental data produced
Agriculture and Agri-Food Canada agr.gc.ca OGP data sets: 339 (0.4%)	**Mandate:** support the growth and development of the agriculture and agri-food sector. **Data:** inventory and geospatial products related to agricultural land use: biomass and crop inventory, spatial density of major crops, plant hardiness zones, soil and land types; Agroclimate maps (drought indices, precipitation, temperature, soil moisture); National Ecological Framework for Canada. **Open data contact:** aafc.opengovernment-gouvernementouvert.aac@agr.gc.ca
Canadian Environmental Assessment Agency canada.ca/en/environmental-assessment-agency.html OGP data sets: 1 (<0.01%)	**Mandate:** deliver environmental assessments that contribute to informed decision-making in support of sustainable development and reduced environmental impacts. **Data:** The Canadian Environmental Assessment Registry and Canadian Environmental Assessment Archives provide metadata on federal environmental assessments. Copies of records produced, collected or received in the context of an environmental assessment are available upon request but not data (raw or processed). **Open data contact:** N/A
Environment and Climate Change Canada canada.ca/en/environment-climate-change.html OGP data sets: 821 (1.0%)	**Mandate:** minimize threats from pollution to Canadians and their environment; provide information on weather, water, and climate conditions; and conserve and restore the natural environment. **Data:** a broad range of environmental and climate data (e.g., Breeding Bird Survey, Air Pollutant Emission Inventory, historical water quality data, database of tornado occurrences in Canada, data from the Canadian Centre for Climate Modelling and Analysis). **Open data contact:** ec.gestiondonnees-datamanagement.ec@canada.ca
Fisheries and Oceans Canada dfo-mpo.gc.ca OGP data sets: 204 (0.3%)	**Mandate:** protect and manage Canada's aquatic ecosystems (marine and freshwater), fisheries, and waterways. **Data:** data on Canada's aquatic ecosystems and species, fisheries and aquaculture activities (e.g., occurrence, fishing effort, landing and biological data for commercial species; temperature and salinity data for locations across Canada; digital images and video of cetaceans for photo identification). **Open data contact:** dfo.infoserv-servinfo.mpo@dfo-mpo.gc.ca
Health Canada canada.ca/en/health-canada.html OGP data sets: 60 (0.08%)	**Mandate:** help Canadians maintain and improve their health, ensure that high-quality health services are accessible, and work to reduce health risks. **Data:** data on environmental risks **Open data contact:** N/A
Natural Resources Canada nrcan.gc.ca OGP data sets: 69 451 (87.0%)	**Mandate:** enhance the responsible development and use of Canada's natural resources and the competitiveness of Canada's natural resources products. **Data:** data on energy, mining, forestry, natural hazards, and climate change (e.g., hydrocarbon exploration and field production statistics, sediment data from the Geological Survey of Canada, National Permafrost Database, energy consumption and greenhouse gas emissions). **Open data contact:** geoginfo@nrcan-rncan.gc.ca
Parks Canada canada.ca/en/parks-canada.html OGP data sets: 447 (0.6%)	**Mandate:** protect and manage nationally significant examples of Canada's natural and cultural heritage such as protected areas. **Data:** information on and complete list of National Parks, National Marine Conservation Areas and National Historic Sites (e.g., location of dams, buildings, and campsites on Parks Canada sites; occurrence and abundance data for various plant and animal species in different national parks). **Open data contact:** N/A
Statistics Canada statcan.gc.ca OGP data sets: 7134 (8.9%)	**Mandate:** collect, compile, analyse, abstract, and publish statistical information relating to the commercial, industrial, financial, social, economic, and general activities and condition of Canadians. **Data:** a broad range of data on Canada's economy, society and environment (e.g., census, housing, tourism, income, energy) at different geographical and temporal scales. **Open data contact:** statcan.od-do.statcan@canada.ca
Transport Canada tc.gc.ca OGP data sets: 24 (<0.01%)	**Mandate:** develop transportation policies and programs that promote safe, secure, efficient and environmentally responsible transportation. **Data:** data related to air, water, and ground transportation (e.g., location of dangerous goods accidents; location of fire stations across Canada; location and information on rail stations; ballast water management plans for vessels in Canadian waters; road network for Canada, the United States, and Mexico). **Open data contact:** tc.opengovernment-gouvernementouvert.tc@tc.gc.ca

Note: The percent of total data sets does not take into account the size or usability of individual data sets. E-mail addresses for assistance with open data in each SBDA are indicated when available. OGP data sets; absolute number and percent of total federal Open Government Portal data sets on the portal in May 2019.
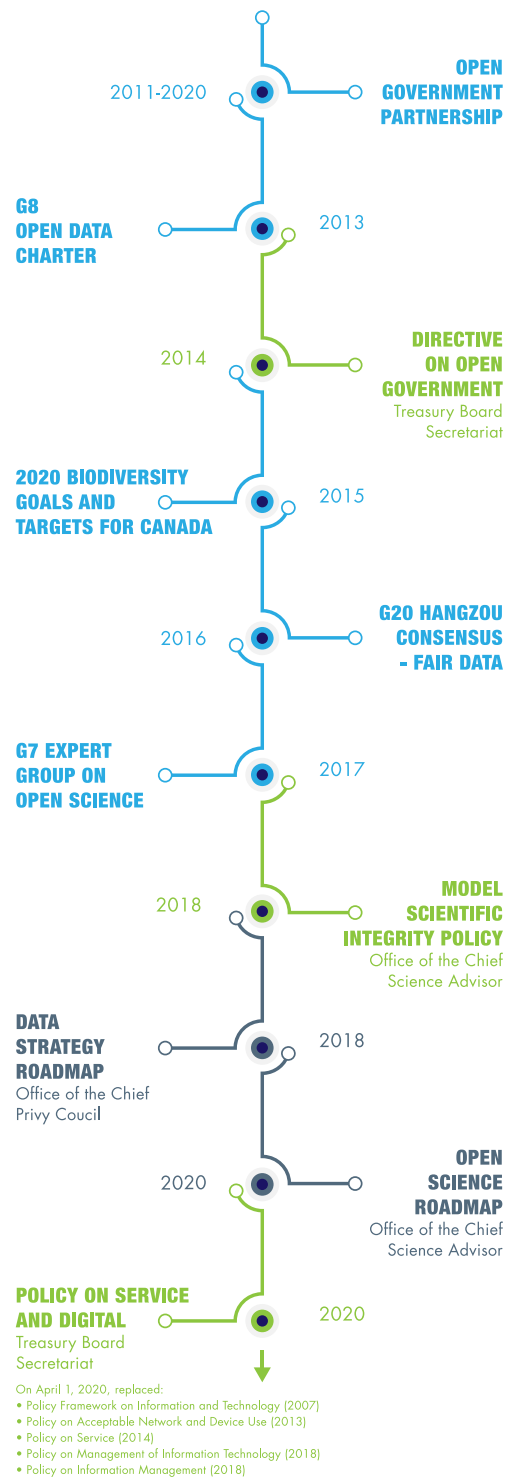
2011-2020 — OPEN GOVERNMENT PARTNERSHIP

G8 OPEN DATA CHARTER — **2013**

2014 — DIRECTIVE ON OPEN GOVERNMENT
Treasury Board Secretariat

2020 BIODIVERSITY GOALS AND TARGETS FOR CANADA — **2015**

2016 — G20 HANGZOU CONSENSUS – FAIR DATA

G7 EXPERT GROUP ON OPEN SCIENCE — **2017**

2018 — MODEL SCIENTIFIC INTEGRITY POLICY
Office of the Chief Science Advisor

DATA STRATEGY ROADMAP
Office of the Chief Privy Coucil — **2018**

2020 — OPEN SCIENCE ROADMAP
Office of the Chief Science Advisor

POLICY ON SERVICE AND DIGITAL
Treasury Board Secretariat — **2020**

On April 1, 2020, replaced:
• Policy Framework on Information and Technology (2007)
• Policy on Acceptable Network and Device Use (2013)
• Policy on Service (2014)
• Policy on Management of Information Technology (2018)
• Policy on Information Management (2018)

Fig. 1. Timeline of international commitments, policies, and guidelines adopted by the Government of Canada in support of open (research) data.

government data being Open, Useful, and Reusable (OURdata Index) (Lafortune and Ubaldi 2018; OECD 2019a). However, standards and procedures for open data vary across and within SBDAs. Disparate open data policies and guidelines across different levels of government and distribution of open government data across multiple repositories and portals hinders discoverability (Privy Council Office Government of Canada 2018). Recognizing these challenges, the Government of Canada is intent on achieving more efficient and cohesive data governance. Here, we synthesize the current status of open research data (Appendix 1) produced by the Government of Canada, focusing on environmental sciences (broadly defined—including natural, social, and physical science perspectives given the inherent interdisciplinary nature of environmental science (Dick et al. 2016)).

This paper targets diverse parties with a stake in open environmental data, including Government of Canada researchers, information technology officers, policy analysts, and executives, as well as academic researchers outside of government. Our aim is to outline: (*i*) the landscape of federal initiatives that promote open data in environmental science, (*ii*) indicate how and where these data are shared (including existing infrastructure and best practices), and (*iii*) discuss planned improvements to open data practices in the Government of Canada.

The Government of Canada's commitments, policies, and directives on open data

In recent years, the Government of Canada has pledged to increase public accessibility of its data by signing international agreements and promulgating policies and directives to this end via its executive and legislative branches (Fig. 1). These policies and directives apply to both administrative agencies and SBDAs. We identified nine SBDAs within the Government of Canada that produce research data relevant to environmental science (Table 1).

International commitments

In 2011, Canada adopted the *International Open Data Charter* (opendatacharter.net) and launched its first Action Plan on Open Government in 2012. Four action plans have now been tabled in the context of the Open Government Partnership (OGP; opengovpartnership.org), an international initiative to improve transparency, public participation, and accountability in governance. Canada joined the OGP in 2011 and co-chaired the organization in 2018 and 2019. National Action Plans on Open Government are tabled biennially and formulate commitments to be implemented over a two-year period. For example, in its third biennial plan to the OGP (2016–2017), Canada formulated 22 commitments to promote transparency in government, including increasing the openness of federal science activities (Commitment 13) and aligning the implementation of open data across Canada (Commitment 14) (Government of Canada 2016). Governments must report on the implementation of their action plan to the OGP via a self-assessment report and OGP's independent review process (Government of Canada 2018).

In 2013, the Government of Canada adopted the *G8 Open Data Charter* (G8 Leaders 2013), stipulating that all government data should be published openly by default and outlining core principles to increase the quality, quantity, and reuse potential of open data. The government then formally endorsed a set of data principles in 2016 via the *G20 Hangzou Concensus* (G20 Leaders 2016), a communiqué in which the G20 leaders support efforts to "promote open science and facilitate appropriate access to publicly funded research results based on Findable, Accessible, Interoperable and Reusable (FAIR) data principles" (Wilkinson et al. 2016).

Canada is also a member of the *G7 Expert Group on Open Science* tasked with making recommendations on incentives and infrastructure to increase the speed and coherence of national transitions towards open science (G7 Science and Technology Ministers 2016). In 2017, the working group

recommended that G7 nations promote research data management practices that adhere to FAIR data principles and develop infrastructure allowing researchers to deposit, access, and analyze data across disciplines on a global scale (G7 Open Science Working Group 2017).

Finally, international commitments specific to the environment demonstrate Canada's recognition of the importance of open data. For example, Canada's commitments under the Convention on Biological Diversity are articulated at a national level in the "2020 Biodiversity Goals and Targets for Canada". Target 14 specifically addresses the need for a "science base [that is] better integrated and more accessible", with indicators of performance under this target specifically related to the amount of environmental data being contributed to online repositories and portals (Biodivcanada 2015).

Policies and directives

Policies are developed by the Treasury Board of Canada Secretariat (TBS) to provide direction to departments and agencies on how to orient their activities toward the achievement of results, enact good stewardship of public resources, and assist ministers in their accountability to Parliament (Treasury Board of Canada Secretariat 2008). Policies provide formal directions that impose specific responsibilities on departments and expectations for heads of departments and agencies and their officials. Directives detail how policy objectives must be met by providing formal instruction obliging departments to take or avoid specific action. Several policies and directives formulate specific requirements regarding transparency and data sharing within the Government of Canada. An in-depth analysis of the hierarchy of policy instruments demonstrated that the guidelines and authorities that support open science can be traced back to the values and ethics code for the public service (Austin 2020). All TBS policies and directives are mandatory across federal departments and agencies.

The Policy on Service and Digital

This policy came into effect on 1 April 2020 (Treasury Board of Canada Secretariat 2020) and replaced several existing policies such as the *Policy on Information Management* and the *Policy Framework for Information and Technology* (Fig. 1). This new policy provides an integrated set of rules for how Government of Canada organizations and employees must manage data, information, and cyber security, among other things. Key aspects of the policy relating to research data management include official languages, privacy, and accessibility requirements, as well as commitments to Government of Canada Digital Standards such as good data stewardship, working in the open by default, using open standards and tools, and addressing data security and privacy risks.

Scientific integrity policies

On 30 July 2018, The Office of the Chief Science Advisor released a *Model Policy on Scientific Integrity* (Office of the Chief Science Advisor 2018). This model serves as a framework for the adoption of scientific integrity policies by 21 SBDAs to proactively strengthen the reliability and credibility of research and scientific activities undertaken by the federal government (Table S1) (Treasury Board of Canada Secretariat 2018d). A core principle of the Model Policy is that research data and information be made openly accessible in a timely and transparent manner unless there are clear and compelling reasons preventing full disclosure.

Directive on Open Government

Issued in 2014 (Treasury Board of Canada Secretariat 2014), this government-wide directive outlines six requirements that must be met by each SBDA's Information Management Senior Official or Chief Information Officer to bolster information management practices enabling the proactive and ongoing release of government information in accessible and reusable formats. Its objective is to maximize the

release of government information and data provided there are no privacy, confidentiality, or security issues, in line with the principle that data should be "as open as possible but as closed as necessary".

Roadmaps

Two roadmaps have recently been published containing guidelines and recommendations intended to clarify a path forward for the Government of Canada to continue making progress on open data and open science more generally (Treasury Board of Canada Secretariat 2008).

Data Strategy Roadmap

Co-authored by Statistics Canada, the Treasury Board Secretariat, and the Office of the Privy Council in 2018 (Privy Council Office Government of Canada 2018), this document outlines the current state of data management practices within the federal government, ongoing challenges, and a comprehensive set of 21 recommendations to improve how the government manages, uses, and shares data both internally and externally.

Roadmap for Open Science

This document (Office of the Chief Science Advisor 2020) provides overarching principles and recommendations to guide Open Science activities in Canada, in particular for science and research funded by federal government departments and agencies.

Implementation of open environmental data by federal science-based departments and agencies

Government scientists and SBDAs seeking to make environmental data publicly available have a range of options of where and how to share them (Culina et al. 2018; Open Science Metrics Working Group 2018). Data repositories are servers that host data, whereas data portals and registries (Appendix 1) contain links and information about data sets stored elsewhere. Portals and repositories are searchable via their own search engines. Additionally, general and discipline-specific repositories can be searched via data aggregators (e.g., DataOne, DataCite, Google Dataset Search) (Culina et al. 2018; Canino 2019). For example, data sets on Canada's Open Government Portal (see below) are discoverable via Google Dataset Search, resulting in approximately 5% of the overall traffic to data set pages on the portal (Treasury Board of Canada Secretariat, personal communication, 2019).

Selecting a suitable data repository is key for ensuring data preservation and discoverability. Similarly, the way in which data are shared (i.e., format, resolution, and metadata) greatly influences their reuse potential. The government's recent commitments and policies towards greater transparency have led to improvements in digital infrastructure and guidelines for sharing research data. Some of these improvements, and the remaining challenges, are described below.

Infrastructure

The Government of Canada currently maintains dozens of SBDA-led data repositories and portals, which can be readily browsed and (or) searched by potential reusers seeking discipline- or agency-specific data sets (Table S2). All SBDAs producing data of relevance to environmental science release data for public consumption, except for the Impact Assessment Agency of Canada (IAAC). Environment and Climate Change Canada, for example, hosts over 40 different data portals providing access to data sets such as species occurrences, climate data, and data on air and water quality (Environment and Climate Change Canada 2019). Fisheries and Oceans Canada shares data on aquaculture, fisheries, and import and export of fisheries products, most of which date back to the 1980s and 1990s. IAAC makes records of environmental assessments available upon request, but does not own or curate data collected by third parties responsible for project assessments. However, the 2019

Impact Assessment Act will facilitate public release of documents and data through the IAAC's online public registry and the Government of Canada's proposed new Cumulative Effects Open Science and Data Platform (Government of Canada 2019). This platform is intended to provide public access to information and data that support decision-making on major projects such as pipelines and mines.

The Government of Canada launched the Open Government Portal (open.canada.ca) in March 2011 (previously named the Open Data Portal) to centralize cataloguing of federal data collected by different SBDAs. The Portal acts as a federated registry of government open data sets, supporting national policies and meeting international commitments to foster open data. A federated registry is a system that integrates several independent repositories or platforms in a single user interface: when the registry is queried, the system finds which of its component repositories or platforms contains the requested data and passes the request to it. Canada's Federated Research Data Repository is an example of a federated system outside the Government of Canada (frdr.ca). At the time of writing, the Open Government Portal was supported by a team of 5–6 technical staff in the Treasury Board Secretariat: it can host data sets and link to data sets archived elsewhere, but it is not intended to replace existing repositories. Rather, it is meant to consolidate open data metadata from across government into a readily searchable, centralized catalogue or registry that links to data found on federal or external platforms. The Open Government Portal consists of three components: (*i*) financial, human resources and government operations information (the Proactive Disclosure Portal), (*ii*) publications and digital information resources (the Open Information Portal), and (*iii*) data sets (the Open Data Portal). As of May 2019, 79 824 federal data sets were listed on the Open Data Portal, nearly all of them (98.3%) contributed by SBDAs producing data of relevance to environmental science (Table 1). Metadata for these data sets are updated daily by Canada's Treasury Board Secretariat (Treasury Board of Canada Secretariat 2019). Anyone can search the Open Data Portal by key words.

The Open Data Portal is also intended as a means of cataloguing provincial and municipal government open data. At present, nine provinces and 66 municipalities in Canada have registered open data portals (open.canada.ca/en/maps/open-data-canada). However, differences in data hosting technology and the absence of standards for file and metadata formats complicates the integration of these data sets in a national portal. In 2018, a pilot project by the Treasury Board Secretariat added over 2000 open data sets by the Government of Alberta to the Open Data Portal, an initiative soon expected to extend to other provinces and municipalities. Some provinces (e.g., Quebec) already federate open municipal data, facilitating the integration of open data at a national scale.

Best practices in research data management and sharing

General guidelines and best practices in data stewardship (Fig. 2) are applicable across domains, to government and nongovernment scientists alike. Open data should comply with the FAIR data principles (Wilkinson et al. 2016), which are described on the GO-FAIR website (go-fair.org/). Briefly, data are findable and accessible when they are archived and freely available from an indexed and searchable public data repository. Data are interoperable and reusable when humans and computers can readily understand them (e.g., via metadata) and explore and reuse them using a range of nonproprietary software. Interoperability and reusability hinge largely on the ease of use of data and file formats (White et al. 2013; Hart et al. 2016), a metadata language (Wilkinson et al. 2018), and a usage license (e.g., creativecommons.org/licenses). Metrics have recently been developed to assess compliance with the FAIR data principles (Wilkinson et al. 2018).

Scientists do not necessarily have a background or training in research data management (Popkin 2019). However, best practice guidelines (Box 1) and open education resources are a good source of free training in research data management (Box 2). Good data stewardship begins with a Data Management Plan (DMP) outlining essential aspects of research data management a priori (Austin

1- DEVELOP A DATA MANAGEMENT PLAN (DMP)

plan steps 2-7
before collecting the data

2- COLLECT, STORE AND BACKUP DATA

ensure that all team
members are familiar
with the DMP

**3- PROVIDE INFORMATIVE METADATA
AND USE INTEROPERABLE FORMATS**

store metadata (information about
the data) alongside the data and
use meta(data) formats allowing
exploration/re-use across a variety
of platforms

4- ADDRESS DATA SENSITIVITY ISSUES

determine security
and privacy concerns
for releasing the data
and implement measures
to minimize risks

5- CHECK DATA QUALITY AND STANDARDIZATION

comply with government-
or discipline-specific standards

6- SELECT AN APPROPRIATE DATA REPOSITORY

use a recognized,
general or discipline-
specific data repository

7- PUBLISH DATA IN THE PUBLIC DOMAIN

OPEN DATA

select a data reuse license
and obtain a permanent identifier
(e.g., DOI) allowing the data to be citable

Fig. 2. Workflow outlining key steps in research data management and stewardship to achieve FAIR (findable, accessible, interoperable and reusable) open data.

FACETS

Box 1. Open access resources providing guidance for managing and publishing open data.

Data management

Some simple guidelines for effective data management (Borer et al. 2009).

Ten simple rules for creating a good data management plan (Michener 2015).

Primer on data management: what you always wanted to know (Strasser et al. 2012).

Data formatting and documentation (metadata)

Guidelines for the deposit and preservation of research data in Canada (Austin et al. 2015).

Ten Simple Rules for Digital Data Storage (Hart et al. 2016).

Tidy Data (Wickham 2014).

Nine simple ways to make it easier to (re) use your data (White et al. 2013).

The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016).

Sensitive data protection

Guide to best practices for generalising sensitive species-occurrence data (Chapman and Grafton 2008).

Registered access: authorizing data access (Dyke et al. 2018).

Troubleshooting public data archiving: suggestions to increase participation (Roche et al. 2014).

Balancing open science and data privacy in the water sciences (Zipper et al. 2019).

---

et al. 2015; Michener 2015). Key elements include determining adequate file formats, data formats, documentation needed to describe the data (i.e., metadata), storage and backup locations, and compliance with ethical considerations in the case of research on humans and other vertebrates (step 1 in **Fig. 2**) (White et al. 2013). The online Portage Network DMP Assistant (assistant.portagenetwork.ca; see **Box 1**) and the DMP Common Standard (RDA 2020) may be useful resources.

Once data collection begins, data must be stored and backed-up safely (step 2 in **Fig. 2**). Adequate formats must be used and metadata provided allowing the data generator(s) and third parties to easily interpret and explore and reuse the data (step 3 in **Fig. 2**; **Box 1**). Next, data quality and standardization must be checked to identify erroneous entries and comply with any government- or discipline-specific standards (steps 4 and 5 in **Fig. 2**). Resources such as Statistics Canada's Data Quality Toolkit (statcan.gc.ca/eng/data-quality-toolkit) are broadly applicable to assist with this step. In government, researchers can consult with their respective SBDA's information management team or data management officer(s) for assistance.

Adequate data protection measures must then be implemented if necessary, and a suitable public data repository identified (step 6, in **Fig. 2**; **Box 3**) (Culina et al. 2018). Sensitive data might include personal identifiers and the location of valuable infrastructure or species, which might pose a security, privacy, or conservation risk if disclosed publicly (Lindenmayer and Scheele 2017). Various means of protecting data exist to allow their release and align with international commitments (e.g., by the G7, G8, and G20) to make government data "as open as possible and as closed as necessary". These

Box 2. Open access online training resources for data management.

Research Data Management Training Resources (Portage)

- portagenetwork.ca/portage-training-resources

- A range of training materials (online modules, webinars, brief guides, primers, guidance on repositories) offered by Portage Canada.

Data Management Education (DataONE)

- dataone.org/Education

- Education modules in Powerpoint, webinars and tutorials by DataONE (Data Observation Network of Earth).

Managing and Sharing Research Data (FOSTER)

- fosteropenscience.eu/node/2328

- Course offered by FOSTER Open Science (EU).

Research Data Management (RDM) Open Training Materials (Zenodo)

- zenodo.org/communities/dcc-rdm-training-materials/?page=1&size=20

- Training materials collection curated by Laura Molloy on the repository Zenodo.

Research Data Management Training (EUDAT)

- eudat.eu/training

- Training programme offered by EUDAT (European e-infrastructure of integrated data services and resources).

MANTRA Research Data Management Training (EDiNA)

- mantra.edina.ac.uk

- Course offered by the University of Edinburgh via its centre for digital expertise.

Research Data Management Toolkit (Jisc)

- rdmtoolkit.jisc.ac.uk

- Training materials specific to researchers, research support staff or IT specialists by Jisc (a membership organisation providing digital solutions for UK education and research).

Research Data Management and Sharing (Coursera)

- classcentral.com/course/coursera-research-data-management-and-sharing-5758

- MOOC (Massive open online course) offered by the University of North Carolina at Chapel Hill and the University of Edinburgh.

Research Data Alliance (RDA) Training and Webinars

- rd-alliance.org/plenaries-events/webinars/past-webinars

- Webinars related to RDA recommendations and outputs and to facilitate data sharing and reuse.

How to practice Open Science (OpenAIRE)

- openaire.eu/support

- Reference material on open science topics for researchers, administrators and funders, including webinars, factsheets, FAQs, and handbooks (e.g., a primer on managing your research data).

Box 3. Resources for publishing data on the Open Government Portal.

Directive on Open Government

tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108

A policy document outlining the context and objective of the Directive on Open Government as well as the requirements and responsibilities of government organizations.

Open Government Guidebook

open.canada.ca/ckan/en/dataset/9eaa6d0e-4b8c-5241-acf7-c6885294b8c1 (Treasury Board of Canada Secretariat 2018b)

A guidebook outlining the Government of Canada's Digital Principles and approach to the Open Data Charter. Practical guidance for releasing data on the Open Government Portal, including a step-wise process (10 steps) and 11 appendices with specific guidelines on topics such as producing metadata, assessing security and privacy concerns, and using the Digital Asset Manager interface to manage open data sets.

Open Government Data and Information Quality Standards (Draft)

open.canada.ca/ckan/en/dataset/bfb87332-5da3-5780-9546-8722a389c91c (Treasury Board of Canada Secretariat 2018a)

A document that details standards and best practices to guide Government of Canada organizations on releasing open information and data sets (including geospatial data) that are maximally readable and reusable by humans and computers.

Open Government Registry Guide (French)

open.canada.ca/ckan/en/dataset/ecc9bb5f-1123-5fe5-81a1-7b6fc10361c8 (Treasury Board of Canada Secretariat 2018c)

A guide outlining how to register data sets with a designated platform for the mandatory release of government data: either on the Open Data Registry (for all nongeospatial data sets) or the Federal Geospatial Platform (for all geospatial or location-based data sets). The Open Government Registry is an internal metadata catalogue used by Government of Canada organizations to manage and release data and information to open.canada.ca.

Open Government Implementation Plans

Search the Open Information Portal (search.open.canada.ca) using the keyword OGIP.

Department- and agency-specific plans to implement the Directive on Open Government containing detailed information on departmental direction, strategies, and initiatives undertaken to meet the requirements of this directive, including governance structures, decision processes, and roles and responsibilities of key stakeholders The Open Government Implementation Plans (OGIPs) were the first deliverable of the Directive on Open Government. Table S3 for a list of OGIPs submitted by science-based departments and agencies that collect environmental data.

Note: A workflow (ref) also helps staff of the Treasury Board Secretariat ensure that data posted on the Open Data Portal comply with the government's web accessibility policies (e.g., requirements for standardized metadata and data provided in both official languages).

measures include anonymizing data (i.e., removing identifiers such as participants' names), reducing spatial resolution (for geospatial data), applying an embargo on data release (i.e., archiving data but delaying their public release), and implementing access control (i.e., vetting the credentials of potential data reusers) (**Box 1**). Once these considerations are addressed, and a publishing license has been chosen (Dodds 2013), a public identifier (e.g., a digital object identifier—DOI) is assigned to the data, which can then be shared publicly (step 7 in **Fig. 2**).

Modernization of research data management and open data within the Government of Canada

The Open Government Directive holds the promise of making vast amounts of untapped government research data accessible to environmental scientists and other interested parties. However, achieving this directive's full potential presents various challenges, in part because the volume of data collected by SBDAs is growing exponentially, in concert with advances in digital technologies (Privy Council Office Government of Canada 2018). In its 2018 Data Strategy Roadmap, the Government of Canada recognizes these challenges and the need to act rapidly to meet its international commitments (Privy Council Office Government of Canada 2018); the authors of the report noted:

> Data are often collected in ways—based on informal principles and practices—that make it difficult to share with other departments or Canadians. Their use is inconsistent across the government and their value sub-optimized in the decision-making process and in day-to-day operations. [. . .] Managing, using and sharing data will be crucial in the coming years, but the government is not set up to treat data as a strategic asset for policymaking, program design or service delivery, or to create value for the public, private, not-for-profit, and research sectors. Across the government, data are created, used and stored within individual areas and are often limited to use for a single purpose. Access, use and re-use between and within organizations are often difficult, in part because of a lack of awareness that data could be useful to others and a reticence to share information with others. [. . .] While intragovernmental data sharing does occur, it is often via ad hoc technical solutions requiring formal letters of agreement signed at the deputy minister level, adding to administrative burden. [. . .] Compounding these challenges is the absence of a governance structure or senior-level decision-making table charged with providing strategic direction on data issues and driving cultural change (p. 4).

Specifically, the Data Strategy Roadmap identifies four overarching challenges across government: the absence of horizontal governance for strategic direction on data issues; a lack of data literacy and cultural resilience to break silos; a lack of adequate digital infrastructure and complex rules framework; and difficulties associated with acquiring, governing, and managing large volumes of disparate data (Privy Council Office Government of Canada 2018). These issues are echoed in several individual SBDAs' Open Government Implementation Plans (**Table S3**), creating inefficiencies within departmental operations. They also compound more general impediments to data sharing, including researchers' fear of being "scooped", concerns about potential errors in their data and analyses being exposed, and the considerable additional time that would be needed to appropriately manage data and metadata.

To remedy these problems, the Data Strategy Roadmap (Privy Council Office Government of Canada 2018) outlines 21 recommendations aimed at modernizing and improving how SBDAs create, manage, protect, use, and share data. These recommendations center around four themes: (*i*) stronger governance to ensure that data are managed holistically as a strategic asset, (*ii*) capacity-building to improve data literacy and skills, (*iii*) enabling infrastructure and legislation to facilitate data

management and analytics, and (*iv*) a more focused treatment of data as valuable assets such that they are discoverable and available to maximize their social and economic value.

Some recommendations are in the process of being implemented—for example, in September 2019, individual SBDAs were asked to submit their own customized data strategies aligning with the standards and principles in the Data Strategy Roadmap. Forty such strategies had been submitted by January 2020 (Statistics Canada, personal communication, 2020). Other concrete recommendations are expected to be rolled out in the coming years. Two key recommendations of the Data Strategy Roadmap are the launch of a digital academy to develop digital and data skills of existing employees and the appointment of a Government of Canada Chief Data Steward to help coordinate decisions on data management issues spanning the whole of government.

Conclusion

The nature of the activities undertaken by government implies that departments and agencies collect, store, manage, and publish considerable amounts of data, including research data (Altayar 2018). By publicly sharing these data, governments contribute to advancing democracy and development through promoting transparency, accountability, and public participation as well as by supporting research and innovation (Zeleti et al. 2016; Ruijer and Martinius 2017). In a 2019 study, 156 open government data initiatives were identified across 61 countries, illustrating rapid progress on this front at a global scale. Key to the success of these initiatives is the implementation of a national data portal enabling users to readily search for and access data sets (OECD 2019a). Portals setup by international governing bodies (e.g., United Nations, European Union, World Bank) and other national governments (Table S4), serve as a valuable reference for governments in the early stages of open data. Ongoing efforts by the Canadian government to prioritize the release of pre-existing data sets (registry.open.canada.ca) and federate open metadata from provincial and territorial governments (open.canada.ca) could prove to be a fruitful national strategy. Such initiatives, in concert with necessary evaluations of the effectiveness of open government data initiatives (Zuiderwijk et al. 2019), are key to the government's 2019 commitment of "raising the bar on openness, effectiveness and transparency" (Office of the Prime Minister 2019).

The ability of environmental scientists to access and synthesize large amounts of data impacts their capacity to understand the natural world and contribute to solving some of the "grand challenges" of the 21st century ((Reid et al. 2010); grandchallenges.org) and addressing the Sustainable Development Goals ((Reid et al. 2017); sustainabledevelopment.un.org). Much of the environmental data collected in Canada originate from the federal government through longstanding monitoring efforts; internal research programs; and collaborations with industry, academia, and citizen scientists. At present, research data management and sharing occurs in a fragmented fashion within the federal government. However, bold commitments and initiatives are underway to modernize data governance and stewardship across the whole of government. Importantly, policies and directives mandate that research data, including many environmental data sets, become "open by default" (Privy Council Office Government of Canada 2018). Achieving this objective requires a multi-pronged approach— on one hand, changes in data governance and culture will help overcome legacy practices that hinder data sharing, on the other hand, massive investments in digital infrastructure and capacity building will be needed to manage and share data in compliance with internationally recognized standards and best practices. Ultimately, environmental science will benefit from a modern, digital, and data-enabled government that is open to sharing.

Acknowledgements

Funding

Author contributions

DGR, SW, GMM, PAS, SJC, and JRB conceived and designed the study. DGR, MG, and CCA performed the experiments/collected the data. DGR analyzed and interpreted the data. DGR, SJC, and JRB contributed resources. DGR, MG, CCA, SW, GMM, PAS, SJC, and JRB drafted or revised the manuscript.

Competing interests

The authors have declared that no competing interests exist.

Data availability statement

No empirical data were collected for this study.

Supplementary material

The following Supplementary Material is available with the article through the journal website at doi:10.1139/facets-2020-0008.

Supplementary Material 1

References

Altayar MS. 2018. Motivations for open data adoption: an institutional theory perspective. Government Information Quarterly, 35(4): 633–643. DOI: 10.1016/j.giq.2018.09.006

Austin CC. 2020. The open science ecosystem: a systematic framework anchored in values, ethics, and FAIRER data. SSRN preprint [online]: Available from papers.ssrn.com/sol3/papers.cfm?abstract_id=3654298.

Austin CC, Baker D, Best M, Born A, Brown S, Cui X, et al. 2015. Guidelines for the deposit and preservation of research data in Canada. Research Data Canada, Ottawa, Ontario [online]: Available from rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Canada-2015.pdf.

Barone L, Williams J, and Micklos D. 2017. Unmet needs for analyzing biological big data: a survey of 704 NSF principal investigators. PLoS Computational Biology, 13(10): e1005755. PMID: 29049281 DOI: 10.1371/journal.pcbi.1005755

Biodivcanada. 2015. 2020 biodiversity goals and targets for Canada [online]: Available from biodivcanada.chm-cbd.net/2020-biodiversity-goals-and-targets-canada.

Borer ET, Seabloom EW, Jones MB, and Schildhauer M. 2009. Some simple guidelines for effective data management. Bulletin of the Ecological Society of America, 90(2): 205–214. DOI: 10.1890/0012-9623-90.2.205

Borgman CL. 2012. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6): 1059–1078. DOI: 10.1002/asi.22634

Canino A. 2019. Deconstructing Google dataset search. Public Services Quarterly, 15(3): 248–255. DOI: 10.1080/15228959.2019.1621793

Chapman AD, and Grafton O. 2008. Guide to best practices for generalising primary species occurrence-data. Version 1.0. Global Biodiversity Information Facility, Copenhagen, Denmark. 27 p. ISBN: 87-92020-06-2 [online]: Available from gbif.org/document/80512.

Cheruvelil KS, and Soranno PA. 2018. Data-intensive ecological research is catalyzed by open science and team science. BioScience, 68(10): 813–822. DOI: 10.1093/biosci/biy097

CODATA. 2019. The Beijing declaration on research data. DOI: 10.5281/zenodo.3552330

Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen-Windhouwer S, and Manghi P. 2018. Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology & Evolution, 2(3): 420–426. PMID: 29453350 DOI: 10.1038/s41559-017-0458-2

Dick M, Rous AM, Nguyen VM, and Cooke SJ. 2016. Necessary but challenging: multiple disciplinary approaches to solving conservation problems. FACETS, 1(1): 67–82. DOI: 10.1139/facets-2016-0003

Dodds L. 2013. Publisher's guide to open data licensing. The Open Data Institute, London, United Kingdom [online]: Available from theodi.org/article/publishers-guide-to-open-data-licensing/.

Dyke SOM, Linden M, Lappalainen I, De Argila JR, Carey K, Lloyd D, et al. 2018. Registered access: authorizing data access. European Journal of Human Genetics, 26(12): 1721–1731. PMID: 30069064 DOI: 10.1038/s41431-018-0219-y

Environment and Climate Change Canada. 2019. Federal progress in implementing open science: 2019 annual report. DOI: 10.5281/zenodo.3369384.

G7 Open Science Working Group. 2017. Annex 4: G7 Expert Group on Open Science [online]: Available from g7italy.it/sites/default/files/documents/ANNEX%204_WG%20Open%20Science/index.pdf.

G7 Science and Technology Ministers. 2016. Tsukuba Communiqué [online]: Available from g8.utoronto.ca/science/2016-tsukuba-en.pdf.

G8 Leaders. 2013. G8 open data charter and technical annex [online]: Available from gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex.

G20 Leaders. 2016. G20 Leaders' Communique Hangzhou Summit [online]: Available from ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967.

Government of Canada. 2016. Third biennial plan to the Open Government Partnership. open.canada.ca/en/content/third-biennial-plan-open-government-partnership.

Government of Canada. 2018. End-of-term self-assessment report on Canada's third biennial plan to the Open Government Partnership 2016–2018 [online]: Available from open.canada.ca/en/content/end-term-self-assessment-report-canadas-third-biennial-plan-open-government-partnership.

Government of Canada. 2019. Discussion paper on information requirements and time management regulatory proposal: a proposed impact assessment system. 22 p. [online]: Available from canada.ca/en/services/environment/conservation/assessments/environmental-reviews/environmental-assessment-processes/discussion-paper-information-requirements-time-management-regulatory-proposal.html.

Hart EM, Barmby P, Lebauer D, Michonneau F, Mount S, Mulrooney P, et al. 2016. Ten simple rules for digital data storage. PLoS Computational Biology, 12(10): e1005097. PMID: 27764088 DOI: 10.1371/journal.pcbi.1005097

Hipsley CA, and Sherratt E. 2019. Psychology, not technology, is our biggest challenge to open digital morphology data. Scientific Data, 6(1): 41. PMID: 31028285 DOI: 10.1038/s41597-019-0047-0

Huang X, and Qiao G. 2011. Biodiversity databases should gain support from journals. Trends in Ecology & Evolution, 26(8): 377–378. PMID: 21665319 DOI: 10.1016/j.tree.2011.05.006

ISO. 2017. ISO 16363: space data and information transfer systems—audit and certification of trustworthy digital repositories. International Standard, Geneva, Switzerland [online]: Available from crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/iso16363.

Lafortune G, and Ubaldi B. 2018. OECD 2017 OURdata Index: methodology and results. OECD Working Papers on Public Governance No. 30. OECD Publishing, Paris, France. DOI: 10.1787/2807d3c8-en

Lennox RJ, Harcourt R, Bennett JR, Davies A, Ford AT, Frey RM, et al. 2020. A novel framework to protect animal data in a world of ecosurveillance. BioScience, 70(6): 468–476. DOI: 10.1093/biosci/biaa035

Lindenmayer D, and Scheele B. 2017. Do not publish. Science, 356(6340): 800–801. PMID: 28546170 DOI: 10.1126/science.aan1362

McClenachan L, Ferretti F, and Baum JK. 2012. From archives to conservation: why historical data are needed to set baselines for marine animals and ecosystems. Conservation Letters, 5(5): 349–359. DOI: 10.1111/j.1755-263X.2012.00253.x

Michener WK. 2015. Ten simple rules for creating a good data management plan. PLoS Computational Biology, 11(10): e1004525. PMID: 26492633 DOI: 10.1371/journal.pcbi.1004525

Molloy JC. 2011. The open knowledge foundation: open data means better science. PLoS Biology, 9(12): e1001195. PMID: 22162946 DOI: 10.1371/journal.pbio.1001195

National Science Foundation. 2015. National survey of college graduates. National Center for Science and Engineering Statistics, Alexandria, Virginia [online]: Available from nsf.gov/statistics/srvygrads/.

Nature Editorial. 2018. Does your code stand up to scrutiny? Nature, 555: 142. PMID: 29517015 DOI: 10.1038/d41586-018-02741-4

NIST. 2019. Big data interoperability framework: volume 1, definitions. NIST Special Publication 1500-1r2 version 3. National Institute of Standards and Technology, Big Data Public Working Group, Gaithersburg, Maryland. DOI: 10.6028/NIST.SP.1500-1r2

OECD. 2019a. Digital government and open government data. *In* Government at a Glance 2019. OECD Publishing, Paris, France. DOI: 10.1787/8ccf5c38-en

OECD. 2019b. Main science and technology indicators: GERD as a percentage of GDP [online]: Available from stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB.

Office of the Chief Science Advisor. 2018. Model policy on scientific integrity [online]: Available from ic.gc.ca/eic/site/063.nsf/eng/h_97643.html.

Office of the Chief Science Advisor. 2020. Roadmap for open science [online]: Available from ic.gc.ca/eic/site/063.nsf/eng/h_97992.html.

Office of the Prime Minister. 2019. Mandate letters [online]: Available from pm.gc.ca/en/mandate-letters.

Open Science Metrics Working Group. 2018. Monitoring open science implementation in federal science-based departments and agencies: metrics and indicators. Report of the Science-based Departments and Agencies (SBDAs) Open Science Metrics Working Group [online]: Available from ecccdocs.techno-science.ca/documents/ECCC_STB_STSD_OpenScienceMetricsReportADMOvf-accessible.pdf.

Poisot T, Bruneau A, Gonzalez A, Gravel D, and Peres-Neto P. 2019. Ecological data should not be so hard to find and reuse. Trends in Ecology & Evolution, 34(6): 494–496. PMID: 31056219 DOI: 10.1016/j.tree.2019.04.005

Popkin G. 2019. Data sharing and how it can benefit your scientific career. Nature, 569(7756): 445–447. PMID: 31081499 DOI: 10.1038/d41586-019-01506-x

Privy Council Office Government of Canada. 2018. A Data Strategy Roadmap for the federal public service [online]: Available from canada.ca/en/privy-council/corporate/clerk/publications/data-strategy.html.

RDA. 2020. Data management plan (DMP) common standard. Research Data Alliance [online]: Available from github.com/RDA-DMP-Common/RDA-DMP-Common-Standard (original work published 2019).

Reid AJ, Brooks JL, Dolgova L, Laurich B, Sullivan BG, Szekeres P, et al. 2017. Post-2015 Sustainable Development Goals still neglecting their environmental roots in the Anthropocene. Environmental Science & Policy, 77: 179–184. DOI: 10.1016/j.envsci.2017.07.006

Reid WV, Chen D, Goldfarb L, Hackmann H, Lee YT, Mokhele K, et al. 2010. Earth system science for global sustainability: grand challenges. Science, 330(6006): 916–917. PMID: 21071651 DOI: 10.1126/science.1196263

Renaut S, Budden AE, Gravel D, Poisot T, and Peres-Neto P. 2018. Management, archiving, and sharing for biologists and the role of research institutions in the technology-oriented age. BioScience, 68: 400–411. DOI: 10.1093/biosci/biy038

Roche DG. 2016. Open data: policies need policing. Nature, 538(7623): 41. PMID: 27708293 DOI: 10.1038/538041c

Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, et al. 2014. Troubleshooting public data archiving: suggestions to increase participation. PLoS Biology, 12(1): e1001779. PMID: 24492920 DOI: 10.1371/journal.pbio.1001779

Roche DG, Kruuk LE, Lanfear R, and Binning SA. 2015. Public data archiving in ecology and evolution: how well are we doing? PLoS Biology, 13(11): e1002295. PMID: 26556502 DOI: 10.1371/journal.pbio.1002295

Ruijer EH, and Martinius E. 2017. Researching the democratic impact of open government data: a systematic literature review. Information Polity, 22(4): 233–250. DOI: 10.3233/IP-170413

Science Digital. 2019. The State of Open Data Report 2019 [online]: Available from digital-science.com/resources/portfolio-reports/the-state-of-open-data-2019/.

Statistics Canada. 2012. Canada year book. Science and technology [online]: Available from www150.statcan.gc.ca/n1/pub/11-402-x/2012000/chap/science/science-eng.htm.

Strasser C, Cook R, Michener W, and Budden A. 2012. Primer on data management: what you always wanted to know. DataOne [online]: Available from dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf.

Treasury Board of Canada Secretariat. 2008. Foundation framework for treasury board policies [online]: Available from tbs-sct.gc.ca/pol/doc-eng.aspx?id=13616.

Treasury Board of Canada Secretariat. 2014. Directive on Open Government [online]: Available from tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108.

Treasury Board of Canada Secretariat. 2018a. DRAFT open government data and information quality standards. 34 p. [online]: Available from open.canada.ca/ckan/en/dataset/bfb87332-5da3-5780-9546-8722a389c91c.

Treasury Board of Canada Secretariat. 2018b. Open government guidebook: government of Canada's guide for releasing open government resources on open.canada.ca. 70 p. [online]: Available from open.canada.ca/ckan/en/dataset/9eaa6d0e-4b8c-5241-acf7-c6885294b8c1.

Treasury Board of Canada Secretariat. 2018c. Open government registry guide. 33 p. [online]: Available from open.canada.ca/ckan/en/dataset/a77d26c3-7af7-5493-9b90-7f22baa08f46.

Treasury Board of Canada Secretariat. 2018d. Scientific integrity policies [online]: Available from canada.ca/en/treasury-board-secretariat/services/information-notice/scientific-integrity-policies.html.

Treasury Board of Canada Secretariat. 2019. Open data portal catalogue [online]: Available from open.canada.ca/data/en/dataset/c4c5c7f1-bfa6-4ff6-b4a0-c164cb2060f7.

Treasury Board of Canada Secretariat. 2020. Policy on service and digital [online]: Available from tbs-sct.gc.ca/pol/doc-eng.aspx?id=32603.

Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, et al. 2013. Mandated data archiving greatly improves access to research data. The FASEB Journal, 27(4): 1304–1308. PMID: 23288929 DOI: 10.1096/fj.12-218164

White EP, Baldridge E, Brym ZT, Locey KJ, McGlinn DJ, and Supp SR. 2013. Nine simple ways to make it easier to (re)use your data. Ideas in Ecology and Evolution, 6(2): 1–10. DOI: 10.4033/iee.2013.6b.6.f

Wickham H. 2014. Tidy data. Journal of Statistical Software, 59(10): 1–23. DOI: 10.18637/jss.v059.i10

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3: 160018. PMID: 26978244 DOI: 10.1038/sdata.2016.18

Wilkinson MD, Sansone S-A, Schultes E, Doorn P, Da Silva Santos LOB, and Dumontier M. 2018. A design framework and exemplar metrics for FAIRness. Scientific Data, 5: 180118. PMID: 29944145 DOI: 10.1038/sdata.2018.118

Wolkovich EM, Regetz J, and O'Connor MI. 2012. Advances in global change research require open science by individual researchers. Global Change Biology, 18(7): 2102–2110. DOI: 10.1111/j.1365-2486.2012.02693.x

Zeleti FA, Ojo A, and Curry E. 2016. Exploring the economic value of open government data. Government Information Quarterly, 33(3): 535–551. DOI: 10.1016/j.giq.2016.01.008

Zipper SC, Stack Whitney K, Deines JM, Befus KM, Bhatia U, Albers SJ, et al. 2019. Balancing open science and data privacy in the water sciences. Water Resources Research, 55: 5202–5211. DOI: 10.1029/2019WR025080

Zuiderwijk A, Shinde R, and Janssen M. 2019. Investigating the attainment of open government data objectives: is there a mismatch between objectives and results? International Review of Administrative Sciences, 85(4): 645–672. DOI: 10.1177/0020852317739115

Appendix 1. Glossary

Research data: Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and (or) are commonly accepted in the research community as necessary to validate research findings and results. All other digital and nondigital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third-party data, public sector data, monitoring data, processed data, or repurposed data. Source: IRiDiuM International Research Data Management glossary (dictionary.casrai.org/Category:Research_Data_Domain).

Data: The word "data" may be used very broadly to comprise research data and the ecosystem of digital things that relate to data, including metadata, software, and algorithms, as well as physical samples and analogue artefacts, and the digital representations and metadata relating to these things (CODATA 2019). For brevity, we refer to scientific data artefacts (data, computer code, and information) as data in the text.

Big Data: Big Data consist of extensive data sets, primarily in the characteristics of volume, velocity, variety, and (or) variability, that require a scalable architecture for efficient storage, manipulation, and analysis (NIST 2019).

Open data: Data that are publicly accessible under a licence making them free to use, modify, combine, and share. Open data do not necessarily comply with the FAIR data principles.

FAIR data principles: Guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The principles emphasize machine actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention), because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data (go-fair.org/fair-principles).

Metadata: Data providing information about other data such as (e.g., ownership, provenance, reuse license, abbreviations, units of measurement). Some metadata follow strict standards to enable interoperability, such as the Dublin Core™ Metadata Initiative (dublincore.org).

Data portal: A website providing access to data that may be stored elsewhere (e.g., in a data repository that is independent from the portal).

Data registry: A metadata catalog of where other data sets and information resources are stored and managed (Treasury Board of Canada Secretariat 2018c).

Data repository: An online platform used to store digital data. The Trusted Digital Repository (TDR) Checklist is a useful reference for building or assessing a data repository (ISO 2017).