Analytical Studies: Methods and References

Data Literacy: What It Is and How to Measure It in the Public Service

by Aneta Bonikowska, Claudia Sanmartin and Marc Frenette

Release date: August 14, 2019

Statistics Canada   Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service   1-800-263-1136
- National telecommunications device for the hearing impaired   1-800-363-7629
- Fax line   1-514-283-9350

   Depository Services Program

   - Inquiries line   1-800-635-7943
   - Fax line   1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Data Literacy: What It Is and How to Measure It in the Public Service

by

Aneta Bonikowska, Claudia Sanmartin and Marc Frenette

Analytical Studies Branch
Statistics Canada

August 2019

Analytical Studies: Methods and References

Papers in this series provide background discussions of the methods used to develop data for economic, health, and social analytical studies at Statistics Canada. They are intended to provide readers with information on the statistical methods, standards and definitions used to develop databases for research purposes. All papers in this series have undergone peer and institutional review to ensure that they conform to Statistics Canada's mandate and adhere to generally accepted standards of good professional practice.

The papers can be downloaded free at www.statcan.gc.ca.

Table of contents

Abstract

This report provides an overview of the definitions and competency frameworks of data literacy, as well as the assessment tools used to measure it. These are based on the existing literature and current practices around the world. Data literacy, or the ability to derive meaningful information from data, is a relatively new concept. However, it is gaining increasing recognition as a vital skillset in the information age. Existing approaches to measuring data literacy—from self-assessment tools to objective measures, and from individual to organizational assessments—are discussed in this report to inform the development of an assessment tool for data literacy in the Canadian public service.

1   Introduction

More and more organizations are discovering the power of data in improving the delivery of products and services to clients, better managing their workforce, and deriving meaningful insights from their data resources. To unleash that power, employees require the skills to derive meaningful insights from data. Data literacy is often described as a critical skill for the 21st century (Bryla 2018). By 2020, Gartner estimates that 80% of organizations will start to roll out data literacy initiatives to upskill their workforce (Gartner 2018). This reality also touches the public service.

Recognizing the power of data in driving better decision making, designing better programs and delivering more effective services to citizens, the Government of Canada released the Data Strategy Roadmap for the Federal Public Service (Government of Canada 2018)**.** One of the potential barriers to realizing this vision is the level of data literacy skills among public servants. To better understand the current gap, an assessment of the current state of data literacy is required.

As a first step toward this goal, this report presents an environmental scan of the concept of data literacy and approaches to measuring it. The concept of data literacy is still fairly new. Much of the work to systematically define the competencies relevant to data literacy has been conducted in the past five years.

2   Methodology

The search was conducted online using Google, Google Scholar and publication databases available through the Statistics Canada library. The search terms included "data literacy", "data literacy public sector", "data literacy govern*", "data literacy survey", "data literacy measure*", "data literacy assessment", "data literacy assessment tool", "alphabétisation de données", "littératie de données". The search covered papers published in academic journals, conference proceedings, white papers and grey literature, and included material from various for-profit and not-for-profit organization websites, which were supplemented in some cases with email or telephone conversations with representatives of the organizations. The search covered publicly available information presented mostly in English.

3   What is data literacy?

Literacy has historically been associated with the ability to read and write. A more comprehensive definition highlighting the many dimensions of the concept was developed by UNESCO in 2003:

Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society. (Robinson 2005, p. 13)

More recently, the term "literacy" is also being used more broadly to describe competency in a particular area, such as statistical literacy, computer literacy and financial literacy. Similarly, most available definitions of data literacy refer to competency in the many dimensions of interacting with data. Most concisely, data literacy is the "ability to derive meaningful information from data" (Sperry 2018). Some authors view data literacy in terms of its overlap with other types of literacy, such as statistical literacy, information literacy and digital literacy (Wolff et al. 2016; Prado and Marzal 2013; Schield 2004; Bhargava et al. 2015). However, existing definitions may become inadequate over time as the types of data available change to become bigger and more complex (Wolff et al. 2016), and emerging technologies such as artificial intelligence change how we think of and use data (Bhargava et al. 2015). For this reason, Bhargava et al. (2015) argue that instead of promoting data literacy and other subtypes of literacy, the focus should be on promoting "literacy in the age of data".

After reviewing the available literature, Wolff et al. (2016) defined data literacy as follows:

Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyse, visualise, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process. (p. 23)

A separate literature review conducted by an interdisciplinary team of researchers at Dalhousie University in Canada (Ridsdale et al. 2015) led to a more concise definition: "Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner" (p. 2). The researchers note that the definition should be allowed to change and evolve with input from stakeholders.

Working from the review by Ridsdale et al., Data to the People put forth an even more concise definition of data literacy, where data literacy is "our ability to read, write and comprehend data, just as literacy is our ability to read, write and comprehend our native language" (Data to the People 2018).

There is a sense in some of the literature that "data literate" should not be a label reserved for data scientists or specialists. Data literacy should be thought of as "the ability of non-specialists to make use of data" (Frank et al. 2016, abstract) and measure "a person's ability to read, work with, analyze and argue with data" (Qlik 2018, p. 3), presumably using simple statistics such as means and percentages.

To summarize, a data literate individual would, at minimum, be able to understand information extracted from data and summarized into simple statistics, make further calculations using those statistics, and use the statistics to inform decisions. However, this definition is context-dependent, which will be illustrated below.

4   Competency frameworks

Measurement and training require a roadmap of the competencies relevant to the concept of data literacy. Efforts to compile a list or model of competencies have been undertaken in academia, particularly in the following areas: library science for training librarians; education, mathematics, computer science and business for teaching students and teachers; and the non-governmental organization (NGO) and private sectors to promote data-informed decision making.

The competency frameworks currently available differ along several dimensions. Some attempt to identify competency areas, giving examples only of the competencies that could be included. Others specify the competencies in different degrees of detail. Some attempt to specify skills that would be needed at different levels of proficiency in a given competency, while others do not. Below is a brief description of some competency frameworks. The competencies covered by each framework are summarized in Table 1.

One of the more detailed lists of competencies and skills was compiled by Ridsdale et al. (2015) following an extensive environmental scan on data literacy, with the goal of informing data literacy education. Although no levels of proficiency are specified, the researchers group the competencies into conceptual competencies, core competencies and advanced competencies. The full matrix of knowledge and tasks associated with the competencies are presented in Appendix A.

Using the Ridsdale et al. framework, Data to the People developed a proficiency scale to accompany the competencies, with the goal of creating a data literacy assessment tool. The scale consists of up to six levels, with some competencies being assessed only at higher levels of proficiency. The description of the skills demonstrated at each level is comparable across competencies. At the lowest level (after not possessing any skills whatsoever in a given competency), a person can perform certain tasks with guidance. At the highest level, a person can assist others with and teach such tasks. The resulting competency framework, called Databilities, includes 15 competencies (a subset of the competencies compiled by Ridsdale et al. [2015]), with a total of 25 subcategories.

Wolff et al. (2016) built a data literacy competency framework around the problem, plan, data, analysis and conclusion (PPDAC) inquiry process, based on work done on both data literacy and statistical literacy (see Appendix B). The authors do not specify which skills might be expected at different proficiency levels, but they do identify four types of citizens who would require different levels of skill complexity given their expected interaction with data: reader, communicator, maker and scientist.

Grillenberger and Romeike (2018) built their competency framework based on work on data management and data science. Their objective was to inform computer science education in schools, and they describe how this objective led them to include or exclude factors that were considered in other papers on data literacy. The interaction between the resulting four content areas and four process areas creates a four-by-four matrix of competencies. The authors list examples of skills that might be considered in each cell of the matrix (see Appendix C). Proficiency levels are not specified in this model—they are left to future work.

Sternkopf (2017) (also summarized by Sternkopf and Mueller 2018) developed a maturity model of data literacy at the organizational level, with a focus on NGOs. It describes the stages that an organization might go through in its journey from not using data in its operations to using data at advanced levels. As such, this model not only identifies relevant data literacy competencies, but also includes four levels of proficiency for each (see Appendix D).

Despite the differing approaches, there is a high degree of overlap in terms of competencies included in the different frameworks, as summarized in Table 1. They all include, either at an individual or at an organizational level, skills necessary to access data, manipulate them, evaluate their quality, conduct analysis, interpret the results, and (in most frameworks) use data ethically. Together, the frameworks clearly show that the concept of data literacy is complex and involves a variety of competencies with a continuum of proficiency.

Table 1
Comparison of selected data literacy competency frameworks

	Ridsdale et al. 2015	Databilities by Data To The People, 2018[1]	Wolff et al. 2016	Sternkopf and Mueller 2018	Grillenberger and Romeike 2018
Plan, implement and monitor courses of action	not applicable	not applicable	Yes	not applicable	not applicable
Undertake data inquiry process	not applicable	not applicable	Yes	not applicable	not applicable
Knowledge and understanding of data, its uses and applications	Yes	not applicable	Yes	not applicable	not applicable
Critical thinking	Yes	not applicable	not applicable	not applicable	Yes ††
Data culture	Yes	not applicable	not applicable	Yes	not applicable
Data ethics (e.g. security, privacy issues)	Yes	not applicable	Yes	Yes	Yes
Data tools	Yes	not applicable	Yes	not applicable	Yes
Data discovery (ability to find and access data)	Yes †	Yes	Yes	Yes	Yes
Data collection	Yes †	Yes	Yes	not applicable	Yes
Data management and organization	Yes	Yes	not applicable	not applicable	Yes
Data manipulation	Yes	Yes	Yes	Yes	Yes
Evaluating and ensuring quality of data and sources	Yes	Yes	Yes	Yes	Yes
Data citation	Yes	not applicable	not applicable	not applicable	not applicable
Basic data analysis	Yes	Yes	Yes	Yes	Yes
Data visualization	Yes	Yes	not applicable	Yes	Yes
Presenting data verbally	Yes	Yes	not applicable	Yes	not applicable
Data interpretation (understanding data)	Yes	Yes	Yes	Yes	Yes
Identifying problems using data	Yes	Yes	Yes	Yes	not applicable
Data driven decision making (DDDM)	Yes	Yes	Yes	not applicable	not applicable
Evaluating decisions/conclusions based on data	Yes	Yes	not applicable	not applicable	not applicable
Metadata creation and use	Yes	Yes	not applicable	not applicable	not applicable
Data duration and re-use	Yes	Yes	not applicable	not applicable	Yes
Data sharing	Yes	not applicable	not applicable	not applicable	Yes
Data preservation	Yes	not applicable	not applicable	not applicable	Yes
Data conversion (from format to format)	Yes	Yes	Yes	Yes	not applicable
Develop hypotheses	not applicable	not applicable	Yes	not applicable	not applicable
Work with large data sets	not applicable	not applicable	Yes	not applicable	not applicable

† Data Discovery and Collection are listed as a single competency in the Ridsdale et al. framework.

†† Grillenberger and Romeike do not list critical thinking as a competency, but it is implicit in their framework.

1. Data literacy surveys conducted by Data to the People include additional competencies tailored to the client organization. This table only lists competencies included in the publicly available Databilities framework.

Note: Adapted by Statistics Canada for accessibility purposes.

Source: Stastistics Canada, authors' compilation.

5   Measuring data literacy—available tools

Publicly available (or advertised) data literacy assessment tools are offered and used mostly by private sector companies and not-for-profit organizations that also offer training in data literacy. These tend to be self-assessments of individual data literacy skills, although some surveys include questions about the extent to which data are used in the decision-making process of an organization as a whole. More objective data literacy assessments have been used in the U.S. education sector and involve scenario-based tests. These are more in line with the way that literacy and numeracy skills are assessed by the Organisation for Economic Co-operation and Development (OECD), such as in the Programme for International Student Assessment and the Programme for the International Assessment of Adult Competencies.

5.1   Assessing individuals' data literacy skills—self-assessment

Individuals can self-assess their levels of proficiency in different data literacy competencies through surveys. The following are some available self-assessment tools.

5.1.1   Databilities

Perhaps the most comprehensive assessment tool of individual data literacy was developed by the Australia-based company Data to the People. It is based on the competency mapping produced by Ridsdale et al. (2015). The tool, called myDatabilities, is an online individual self-assessment survey. Note that the assessment tool that is publicly available on the company's website does not measure data ethics, a competency included in most data literacy competency models.

Each question asks respondents to choose the answer that best describes them. The answers are presented in order from least proficient in a given competency to most proficient and follow the same format for each competency. The questions and answers take the following format:

Which of these statements best describe you?

    a. With guidance, I can describe the analysis I need to perform.

    b. I can develop a simple plan for analysis using frameworks provided to me.

    c. I can develop a simple plan for my analysis.

    d. I can develop a plan for analysis to better understand a range of problems.

    e. I can assist others to develop a plan for analysis.

    f. I can teach and assist others to develop a plan for analysis to better understand a range of problems.

    g. None of these describe me.

The company also conducts organizational assessments that, in addition to the baseline survey described above, include additional questions tailored to the organization, department and job description of the person completing the survey. Such an assessment would identify areas within an organization with high and low concentrations of data literacy skills, as well as identify groups that would benefit from data literacy skills training. The company offers suggestions for the type of training that would benefit groups of individuals within the organization being assessed. Data to the People has been retained to conduct data literacy surveys in various companies worldwide, including individual government departments.

5.1.2 Qlik

The software company [Qlik](#) (as well as the [Data Literacy Project](#) it is part of) offers on its website a publicly available 10-question [survey](#) of individual data literacy skills. There are four answer categories to each question. On the surface, the categories do not appear to represent a smooth skill progression, and in some cases they measure the respondent's feelings toward data rather than skills directly. The scores are aggregated into four levels of proficiency, with titles such as data dreamer and data knight, and provide a short description of the level of proficiency and areas that can be improved upon. Links to suggested training are provided. Here is an example of a question and its answer categories:

How often do you feel overwhelmed with data?

a. Quite often with all the requests that come in.

b. I don't really let it bother me. It's just numbers and stats.

c. Monthly, when the bills come in.

d. Never… it's useful, why would it be overwhelming?

The company has commissioned a data literacy study among employees and decision makers in companies in several countries around the world.

5.1.3 Office of the Maricopa County School Superintendent

The Office of the Maricopa County School Superintendent has posted a [survey](#) on its website entitled "What's Your Data Literacy IQ?" that is aimed at helping teachers reflect on their level of data literacy. It includes seven statements about various data competencies and asks respondents to rate each statement on a five-point scale, where 0 means never true and 4 means always true. For example, one of the statements reads as follows:

Act: I make decisions and take action based on data.

5.1.4 The Open Data Institute

The Open Data Institute (ODI) is a non-profit company based in the United Kingdom that offers consultancy and training in the development and use of open data. It recently conducted a survey of government employees across the world, focusing on skills employees say they want to learn (as opposed to trying to measure individual capabilities in the area of data literacy), on the way they would prefer to pursue training (e.g., face-to-face, online), and on factors affecting the use of open data in the respondent's job. The survey questions can be found on the ODI [website](#).

The ODI also developed what it refers to as a [skills framework](#), which is related to data literacy and is meant to allow individuals to assess their progress in becoming data literate. Each skill box is linked to appropriate training offered by the ODI. The skills are grouped into four categories to describe how advanced the person is: explorer, strategist, practitioner and pioneer. Individual skills are represented by icons, but there is no clear description of individual competencies. The skills framework and the survey are not linked.

5.2 Assessing individuals' data literacy skills—objective measures

A practical assessment is an alternative approach to measuring individual data literacy skills.

5.2.1 U.S. Department of Education

A study of data literacy among teachers in the United States, commissioned by the U.S. Department of Education, assessed a sample of teachers through in-person interviews on their data literacy skills and the way they think about data (Means et al. 2011). The teachers, either individually or in small groups, were presented with hypothetical scenarios about their students and data contained in either a table or a graph. The questions required them to locate appropriate information in the data and perform simple calculations. Their entire responses, including how they arrived at their answers, were recorded and analyzed. The report lists the scenarios used, along with the skills tested in each scenario. The researchers suggest that scenarios like these could be used as part of data literacy training for teachers.

5.2.2 WestEd

A not-for-profit organization in the United States called WestEd has developed four scenario-based assessments to measure the data literacy skills of school teachers.

5.3 Assessing organizational data literacy—self-assessment

In addition to individual assessments, organizational assessments can also be conducted. These measure the degree to which data are used for daily operations in the organization as a whole, i.e., the culture around data use in addition to the data literacy of the workforce. At the organizational level, it is not clear who is responding to the survey and how they synthesize the workforce's literacy skills.

5.3.1 Data literacy maturity model assessment tool

The data literacy maturity model developed by Sternkopf (2017) and summarized by Sternkopf and Mueller (2018), as described in the previous section, is accompanied by an assessment tool available on the website of the Germany-based organization Datenschule [School of Data]. The tool includes questions about nine competencies. There are four answers to each question, describing four levels of proficiency, and the respondent is asked to mark one. An example of a question and its answer categories follows:

How do you evaluate your ability to formulate questions in order to find meaningful answers in data?

   a. Ability to ask questions that can be answered by simple data queries.

   b. No feeling for what questions can be answered by data.

   c. Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve.

   d. Entire projects are based on multidimensional questions that need complex data queries and answer all main questions and sub-questions. No single person can handle these inquiries.

This assessment tool has been used by Datenschule in its work with NGOs and local governments.

5.3.2 PoliVisu Data Literacy Survey

Funded by the European Union's Horizon 2020 research and innovation program, this survey collected evidence on the knowledge and use of big data by public authorities across the European Union. It was aimed primarily at organizations dealing with mobility and transportation. The survey questions asked whether big data were used or produced in the organization, what specific purpose they were used for, and who within the organization was using them. The survey included a question on the degree of proficiency, on a scale from 1 to 8, regarding competencies such as analyzing or interpreting data. Note that links to the online survey are no longer available.

5.3.3 Gartner

The consulting firm Gartner advertises its "Toolkit: Enabling Data Literacy and Information as a Second Language," which can be purchased from the company. It appears to include an assessment tool of data literacy in an organization.

5.3.4 Data Literacy Project

The Data Literacy Project is a collaboration between several consulting companies, including Data to the People and Qlik. Qlik, which develops business intelligence and data visualization software, commissioned a measurement system called the Corporate Data Literacy (CDL) score. This index was developed in collaboration with academics from the Wharton School at the University of Pennsylvania. Companies are assessed along three dimensions: employee data literacy skills, data-driven decision making and data skill dispersion. The total score ranges from 0 to 100. Qlik has gathered data from private corporations in a variety of industry sectors worldwide, and has shown in subsequent analysis that higher CDL scores were associated with higher enterprise value (Qlik 2018). However, a data literate workforce was not always accompanied by efficient use of data within the company (Qlik 2018). The Data Literacy Project aims to create an assessment tool for companies to measure their data literacy against the CDL score, and to offer the tool on the Data Literacy Project website in 2019.

6 Measuring data literacy in the public sector

The OECD lists data literacy as one of six skills essential for innovation in the public sector (OECD 2017). Its report clearly distinguishes between data specialists and non-specialists, and highlights the need for data non-specialists to become data literate. It also outlines a progression scale of data literacy in four areas: (a) data being used in decision making, (b) data being used to manage public services, (c) the ability of data non-specialists to engage data specialists, and (d) the ability of data specialists to communicate effectively with non-specialists about data and results of analysis.

Internationally, data literacy in the public service is still a novel concept, as far as can be gleaned from publicly available documents. The Australian public service has a well-articulated data literacy strategy (Australian Government 2016), but does not appear to have conducted an assessment of the data literacy levels among its workforce. The U.K. government also appears to be interested in data literacy, and its Government Digital Service has developed a data literacy program for civil servants (Cattell 2016; Duhaney 2018). The Office for National Statistics (ONS) was also working to develop a data literacy scale similar to the digital literacy scale used in the government. An ONS blog entry in April 2018 included a draft competency framework (Knight 2018). However, there does not appear to be any existing model for measuring data literacy across a public service.

7   Discussion

Data literacy is a complex concept. Based on some of the competency frameworks described in this report, it may seem like a person needs to become proficient in all of the competencies involved in extracting and using information contained in data to be considered data literate. That may be a desired outcome for some occupations, graduates of certain courses or programs, or teams. However, within an organization—or, more specifically for this report, within the entire public service—it is more reasonable to assume that specialized data handling and analysis tools will remain the domain of a relatively small group of jobs, if for no other reason than because of the training, experience and periodic skills upgrading required to perform such tasks competently. The goal of increasing data literacy within the public service might instead look more like the guidelines outlined by the OECD (2017), i.e., building a culture of data-informed decision making and service provision, and developing an effective working relationship between data specialists and non-specialists. At a minimum, this would require non-specialists to appreciate the value of data in decision making, as well as have an intuitive understanding of how to interpret data. Of course, it would also require advanced data literacy skills among specialists.

In terms of measuring data literacy in the public service, this implies that a comprehensive assessment of all facets of the data literacy required by specialists and non-specialists would be ideal. Most assessment tools being used are subjective—respondents assess their own data literacy skills or attitudes toward data. While it might be reasonable to assume that a self-assessment of attitudes toward data yields reliable results, this may not be the case with self-assessed data literacy skills. Although less expensive than objective assessments, self-assessments may produce a distorted picture of the actual skill distribution. Indeed, previous studies have demonstrated that low performers substantially overestimate their performance on intellectual tasks (Ehrlinger et al. 2008). Computer proficiency reporting also differs by gender. Hargittai and Shafer (2006) showed that women tend to undervalue their abilities in finding information online relative to men. However, in an objective assessment, the demonstrated skills of men and women were not statistically distinguishable from one another.

Some of the required skills for data literacy may not be easily tested with an objective assessment. Tasks such as reading and converting data with software or producing dynamic data visualization tools may simply be too costly to assess objectively, and perhaps too onerous for respondents (except in the context of job competitions). In that case, a subjective assessment may be the most realistic option. Of course, the biases described above should be kept in mind when measuring self-assessed skills at a point in time. If, on the other hand, the goal is to track the evolution of these skills in the public service, it is possible (but not necessarily the case) that the biases remain consistent over time. In that case, self-assessment may be a practical approach to tracking trends in data literacy over time.

In contrast, assessing the understanding of data and statistics may lend itself to an objective assessment. The assessment does not need to be lengthy—the Canadian Financial Capability Survey assesses the financial literacy of Canadians through a 14-question instrument. In principle, a similar tool could be developed to test the basic statistical literacy of public servants. This would be particularly useful in assessing the skills of data consumers—public servants who rely on reading statistical reports or research studies related to their program. These workers may simply need to know how to interpret the findings, rather than how to produce them.

8   Appendix

Appendix A
Data literacy competency framework developed by Ridsdale et al. (2015)

Key ability/knowledge area	Competency	Knowledge/tasks		
Conceptual framework				
	Introduction to data	Knowledge and understanding of data	Knowledge and understanding of the uses and applications of data	
Data collection				
	Data discovery and collection	Perform data exploration	Identifies useful data	Collects data
	Evaluation and ensuring quality of data and sources	Critically assesses sources of data for trustworthiness	Critically evaluates quality of datasets for errors or problems	
Data management				
	Data organization	Knowledge of basic data organization methods and tools	Assesses data organization requirements	Organizes data
	Data manipulation	Assesses methods to clean data	Identifies outliers and anomalies	Cleans data
	Data conversion (from format to format)	Knowledge of different data types and conversion methods	Converts data from one format or file type to another	
	Metadata creation and use	Creates metadata descriptors	Assigns appropriate metadata descriptors to original data sets	
	Data curation, security and re-use	Assesses data curation requirements (e.g. retention schedule, storage, accessibility, sharing requirements, etc.)	Assess data security requirements (e.g. restricted access, protected drives, etc.)	Curates data
	Data preservation	Assesses requirements for preservation	Assesses methods and tools for data preservation	Preserves data

Note: Adapted by Statistics Canada for accessibility purposes.

Source: Ridsdale et al. (2015), p. 38. Appendix 1 - Data Literacy Competencies Matrix.

LEGEND:	Conceptual competencies	Core competencies	Advanced competencies

Appendix A
Data literacy competency framework developed by Ridsdale et al. (2015) (continued)

Key ability/knowledge area	Competency	Knowledge/tasks				
Data evaluation						
	Data tools	Knowledge of data analysis tools and techniques	Selects appropriate data analysis tool or technique	Applies data analysis tools and techniques		
	Basic data analysis	Develops analysis plans	Applies analysis methods and tools	Conducts exploratory analysis	Evaluates results of analysis	Compares results of analysis with other findings
	Data interpretation (understanding data)	Reads and understands charts, tables and graphs	Identifies key take-away points, and integrates this with other important information	Identifies discrepancies within the data		
	Identifying problems using data	Uses data to identify problems in practical situations (e.g. workplace efficiency)	Uses data to identify higher level problems (e.g. policy, environment, scientific experimentation, marketing, economics, etc.)			
	Data visualization	Creates meaningful tables to organize and visually present data	Creates meaningful graphical representations of data	Evaluates effectiveness of graphical representations	Critically assesses graphical representations for accuracy and misrepresentation of data	
	Presenting data (verbally)	Assess the desired outcome(s) for presenting the data	Assesses audience needs and familiarity with subject(s)	Plans the appropriate meeting or presentation type	Utilizes meaningful tables and visualizations to communicate data	Presents arguments and/or outcomes clearly and coherently
	Data driven decisions making (DDDM) (Making decisions based on data)	Prioritizes information garnered from data	Converts data into actionable information	Weighs the merit and impacts of possible solutions/decisions	Implements decisions/solutions	
Data application						
	Critical thinking	Aware of high level issues and challenges associated with data	Thinks critically when working with data			
	Data culture	Recognizes the importance of data	Supports an environment that fosters critical use of data for learning, research, and decision-making			
	Data ethics	Aware of legal and ethical issues associated with data	Applies and works with data in an ethical manner			
	Data citation	Knowledge of widely-accepted data citation methods	Creates correct citations for secondary data sets			
	Data sharing	Assesses methods and platforms for sharing data	Shares data legally, and ethically			
	Evaluating decisions based on data	Collects follow-up data to assess effectiveness of decisions or solutions based upon data	Conducts analysis of follow-up data	Compares results of analysis with other findings	Evaluates decisions or solutions based on data	Retains original conclusions or decisions, or implements new decisions/solutions

Note: Adapted by Statistics Canada for accessibility purposes.
Source: Ridsdale et al. (2015), p. 38. Appendix 1 - Data Literacy Competencies Matrix.

LEGEND: | Conceptual competencies | Core competencies | Advanced competencies |

Appendix B
Data literacy competency framework compiled by Wolff et al. (2016)

Competence	Foundational competence	PPDAC
Inquiry process		
Plan, implement and monitor courses of action	Not applicable	Not applicable
Undertake data inquiry process	Not applicable	Not applicable
Foundational knowledge		
Understand the ethics of using data	Ethics	Not applicable
Use data to solve (real) problems Understand the role and impact of data in society in different contexts	Real-world problem-solving context	Not applicable
Identify problems or questions that can be solved with data	Ask questions from data	Problem
Develop hypotheses, Identify data	Develop hypotheses and identify potential sources of data	Plan
Collect or acquire data, *Critique data*	Collect or acquire data	Data
Transform data into information and ultimately actionable knowledge, Create explanations from data, *Access data, Analyse data, Understand data types, Convert data, Prepare data for analysis, Combine quantitative and qualitative data, Use appropriate tools, Work with large data sets, Summarise data*	Analyse and create explanations from data	Analysis
Interpret information derived from datasets, Critique presented interpretations of data	Evaluate the validity of explanations based on data and formulate new questions	Conclusion

Note: PPDAC signifies Problem, Plan, Data, Analysis and Conclusion. Adapted by Statistics Canada for accessibility purposes.
Source: Wolff et al. (2016), pp. 13-14. Table 1. Categorization of data literacy skills across multiple definitions.

Appendix C

Data literacy competency framework and examples of skills corresponding to the competencies developed by Grillenberger and Romeike (2018)

	P1 Gathering, modeling and cleansing	P2 Implementing and optimizing	P3 Analyzing, visualizing and interpreting	P4 Sharing, archiving and erasing
C1 Data information	- Choose suitable sensors for gathering the desired information as data - Structure the gathered data in a suitable way for later analysis - Evaluate if the captured data represents the original information correctly	- Implement algorithms for gathering the desired data - Implement simple algorithms to download data from web APIs - Discuss optimizations and limits of data gathering	- Combine data to gain new information - Emphasize the desired information in visualization - Interpret data and analysis results to get new information	- Decide whether to share original data - Decide which of the original data to store to keep the required information - Decide on an appropriate way to delete specific data
C2 Data storage and access	- Select a suitable data model - Structure data in a suitable way for storage - Visualize data models in a suitable way	- Decide on a suitable data storage and store the data - Use possibilities for enabling efficient access to data - Increase storage efficiency using compression	- Access the data in a suitable way for analysis - Use suitable data formats for the data to analyze - Store their analysis results appropriately	- Decide whom to give access to the stored data - Determine access rights for the data - Discuss issues related to data validity when erasing data
C3 Data analysis	- Decide whether specific data influences results of analysis - Structure data appropriately for analysis - Connect data from different sources for analysis purposes	- Implement simple analysis algorithms - Determine adjustment screws for analysis - Optimize data analyses in order to gain higher quality results	- Decide for appropriate analysis methods - Visualize data and analysis results - Interpret the results of analysis	- Decide which analysis results to share with whom - Reason whether storing the original data is necessary after analyzing them - Decide whether it is reasonable to share information about the analysis process
C4 Data ethics and protection	- Reflect ethical issues when gathering information - Decide whether combining different data sources is reasonable in specific contexts - Discuss impacts on privacy when continuously capturing data	- Discuss how to anonymize or pseudonymize data appropriately - Exclude data from permanent storage based on ethical considerations - Choose access rights to data based on privacy issues	- Discuss the ethical impacts of the conducted data analysis and their results - Decide whether analysis results are sufficiently anonymized - Reflect whether analyzing specific data raises privacy issues	- Reason whether storing data for further uses should be allowed from an ethical perspective - Decide on appropriate ways to securely erase original data and analysis results - Find ways for appropriately removing attributes that lead to privacy issues

Note: Adapted by Statistics Canada for accessibility purposes.
Source: Grillenberger and Romeike (2018), 8th page. Table 1: Matrix of exemplary competencies for the different combinations of process (P1–P4) and content areas (C1–C4).

Appendix D
Data literacy competency framework developed by Sternkopf (2017)

	1 Uncertainty	2 Enlightenment	3 Certainty	4 Data fluency
	Organizations are unaware of the need for data literacy skills and have no or very vague understanding of what is required. Individuals might have a certain interest in data and work digitally, but are unsure about the different steps that exist when working with data.	Organizations are experimenting with the application of data related topics. Describes a state where a lot about data has already been understood theoretically but cannot be applied in many cases and has to be trained further.	Organizations perform data handling steps with confidence and have data-driven activities built into their routine processes wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, and benefits are understood at all levels of the organization.	Organizations have established a data-informed culture throughout all levels. Data is actively used to improve processes and create workflows.
Data culture	Data is perceived as an ambiguous term that causes insecurities.	Data is perceived as an interesting concept, and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.	Data is not perceived as a source of insecurity, but rather understood as an enabler for progress and support for existing and planned activities. Higher management and leaders support data initiatives.	Psychological barriers of data have been brought down (e.g., insecurities, fear, resignation), and comfort around data is promoted. Higher-level management and project managers understand and support importance of dedicated resources (time, budget, human resources) for data handling and conversion.
Data ethics and security	No awareness for guidelines that ensure confidentiality, integrity, and availability of data.	Rising awareness and uncoordinated attempts to promote the importance of the responsible use of data. No defined guidelines.	Awareness of the impacts of data use. Guidelines for responsible data handling are defined and incorporated internally to activities.	Processes are in place to ensure confidentiality, integrity, and availability of data. Only data that is necessary is collected/used. Consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated.
Ask questions/define	Lacking ability to formulate questions to find meaningful answers in data. Indifference about which questions can be answered by data.	Questions can be asked to data in a limited number of situations, and answers are provided through simple queries.	Questions to data are formulated precisely and target-oriented to find meaningful answers in most of the cases.	Entire projects are based on multidimensional questions. Answers to informational needs can be found consistently in data, because of the high awareness of what questions can be answered by data (no overinterpretation).
Find	Limited understanding of possible data sources. Use of basic search engines to find data. No experience for identifying and selecting most relevant data sources.	Knowledge limited to only a few data sources. Advanced use of search engines. Use of internal data sources and data requests at public institutions are common practice.	Broad understanding of different data sources, and most relevant ones can be chosen from a selection of data sources. Awareness and use of data portals for specific topics.	Profound understanding of the various possible types of data sources. Assessment criteria for selecting the ones most relevant to an informational need are formulated. Ability to detect when a given problem or need cannot be solved with the existing data, and knowledge about research techniques to obtain new data (e.g., complex queries).
Get	Data is derived from full text and used as base for further processing.	Use of downloads and data formats such as .csv. Frequent use of internal programs to access data (e.g., CRM).	Data can be accessed using more complex data formats (e.g., JSON, XML). Use of APIs to get data.	Access to data through sophisticated methods (e.g., automated data scrapers/scripts). Ability to convert input format into a form that can be used for further processing and analysis.

Note: Adapted by Statistics Canada for accessibility purposes.
Source: Sternkopf and Mueller (2018), pp. 5051-5052. Table 3. Data literacy maturity grid.

Appendix D
Data literacy competency framework developed by Sternkopf (2017) (continued)

	1 Uncertainty	2 Enlightenment	3 Certainty	4 Data fluency
	Organizations are unaware of the need for data literacy skills and have no or very vague understanding of what is required. Individuals might have a certain interest in data and work digitally, but are unsure about the different steps that exist when working with data.	Organizations are experimenting with the application of data related topics. Describes a state where a lot about data has already been understood theoretically but cannot be applied in many cases and has to be trained further.	Organizations perform data handling steps with confidence and have data-driven activities built into their routine processes wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, and benefits are understood at all levels of the organization.	Organizations have established a data-informed culture throughout all levels. Data is actively used to improve processes and create workflows.
Verify	Critical evaluation of data does not exist. Data is taken at face value. Data evaluation criteria cannot be described.	Critical check of simple data quality measures.	Multiple layers of data checking are implemented in standard procedures across the organization.	Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability, and quality are precisely defined.
Clean	No awareness that given data might have to be checked, cleaned, or normalized. Data is further processed as is.	Awareness that given data most often is not perfect. Awareness of some data quality criteria (e.g., empty fields, duplicates) and manual fixing of errors.	Invalid records can be detected and are removed using programs that support data cleaning (e.g., OpenRefine). High awareness of data quality criteria (e.g., machine processable, empty fields, duplicate detection).	Independent ability to remove invalid records and translating all the columns to use a set of values through an automated script. Ability to combine different datasets into a single table, remove duplicate entries, or apply any number of other normalizations.
Analyze	Bar and pie charts, simple use of data tables, and basic summaries of data.	Ability to work with basic descriptive statistics. Pivot tables for aggregating information, histograms, and box plots.	Ability to work with advanced statistics (e.g., inferential view of data, linear regression, decision trees).	Full suite of machine learning tools (e.g., clustering, forecasting, boosting, ensemble learning).
Visualize	No awareness of the multiplicity of how data can be presented. No understanding of when standard visualizations are chosen; decisions based on what looks best (trial and error).	Ability to find specific outputs in accordance with information that needs to be represented (e.g., in Excel).	Creation of interactive charts/dashboards. Uncertainties are always visualized along with the data.	High awareness of the various forms in which data can be presented (written, numerical or graphic). Sophisticated visualizations are programmed, linked, and dynamic dashboards that anticipate user requests are designed.
Communicate	Insights from data are not communicated or put into a broader context.	Limited ability to find specific outputs. Simple narrative support static visualizations/key numbers (e.g., reporting to funding partners, newsletters).	Own projects are supported by interactive visualizations and more sophisticated narrative in a broader context (e.g., data storytelling, conferences, talks, monthly updates, blog posts).	Ability to synthesize and communicate in ways suited to the nature of the data, their purpose, and the audience (e.g., data storytelling, data-driven campaigning, workshops, conferences, monthly updates, blog posts, reproducible research).
Assess and interpret	Data outputs are used at face value without questioning their correctness and message.	Growing awareness for critically assessing data outputs and interpreting the results. Insecurities regarding what exactly to pay attention to.	Data outputs and results are interpreted confidently and critically. Evaluation criteria are internalized.	Data outputs and results are consistently questioned and challenged, interpretation extends the obvious, and information are successfully translated into actionable knowledge.

Note: Adapted by Statistics Canada for accessibility purposes.
Source: Sternkopf and Mueller (2018), pp. 5051-5052. Table 3. Data literacy maturity grid.

References

Australian Government. 2016. "Data skills and capability in the Australian public service." *Department of the Prime Minister and Cabinet*. Available at: https://www.pmc.gov.au/resource-centre/public-data/data-skills-and-capability-australian-public-service (accessed November 15, 2018).

Bhargava, R., E. Deahl, E. Letouzé, A. Noonan, D. Sangokoya, and N. Shoup. 2015. "Beyond data literacy: Reinventing community engagement and empowerment in the age of data." *Data-Pop Alliance*. Available at: https://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/ (accessed November 14, 2018).

Bryla M. 2018. "Data literacy: A critical skill for the 21st century." *Tableau Blog*. Available at: https://www.tableau.com/about/blog/2018/9/data-literacy-critical-skill-21st-century-94221 (accessed January 16, 2019).

Cattell, J.A. 2016. "Planning a data literacy programme in government." *Public Sector Stuff*. January 29. Available at: https://jacattell.wordpress.com/2016/01/29/planning-a-data-literacy-programme-in-government/ (accessed January 15, 2019).

Data to the People. 2018. *Databilities: A Data Literacy Competency Framework*. Available at: https://www.datatothepeople.org/databilities (accessed May 23, 2019).

Duhaney, D. 2018. "Data literacy—improving conversations about data." *Data in Government*. February 21. Available at: https://dataingovernment.blog.gov.uk/ (accessed January 15, 2019).

Ehrlinger, J., K. Johnson, M. Banner, D. Dunning, and J. Kruger. 2008. "Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent." *Organizational Behavior and Human Decision Processes* 105: 98–121.

Frank, M., J. Walker, J. Attard, and A. Tygel. 2016. "Data literacy: What is it and how can we make it happen?" *The Journal of Community Informatics* 12 (3): 4–8.

Gartner. 2018. "Gartner keynote: Do you speak data?" *Smarter with Gartner*. Available at: https://www.gartner.com/smarterwithgartner/gartner-keynote-do-you-speak-data/ (accessed January 16, 2019).

Government of Canada. 2018. *Report to the Clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service*. Ottawa: Government of Canada. Available at: https://www.canada.ca/content/dam/pco-bcp/documents/clk/Data_Strategy_Roadmap_ENG.pdf (accessed January 16, 2019).

Grillenberger, A., and R. Romeike. 2018. "Developing a theoretically founded data literacy competency model." In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, Potsdam, Germany, October 4 to 6, 2018. New York: ACM.

Hargittai, E., and S. Shafer. 2006. "Differences in actual and perceived online skills: The role of gender." *Social Science Quarterly* 87 (2): 432–448.

Knight, M. 2018. "A data literacy scale?" *ONS Digital: Blog for the @ONSdigital Team*. April 5. Available at: https://digitalblog.ons.gov.uk/2018/04/05/a-data-literacy-scale/ (accessed January 3, 2019).

Means, B., E. Chen, A. DeBarger, and C. Padilla. 2011. *Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports.* Washington, D.C.: Office of Planning, Evaluation and Policy Development, U.S. Department of Education.

OECD (Organisation for Economic Co-operation and Development). 2017. "Core skills for public sector innovation." In *Skills for a High Performing Civil Service.* Paris: OECD Publishing.

Prado, J.C., and M.A. Marzal. 2013. "Incorporating data literacy into information literacy programs: Core competencies and contents." *Libri* 63 (2): 123–134.

Qlik. 2018. *The Data Literacy Index.* Available at: https://thedataliteracyproject.org/ (accessed January 3, 2019).

Ridsdale, C., J. Rothwell, M. Smit, H. Ali-Hassan, M. Bliemel, D. Irvine, D. Kelley, S. Matwin, and B. Wuetherick. 2015. *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.* Available at: https://dalspace.library.dal.ca/xmlui/handle/10222/64578 (accessed November 15, 2018).

Robinson, C. 2005. *Aspects of Literacy Assessment: Topics and Issues from the UNESCO Expert Meeting.* Paris: UNESCO.

Schield, M. 2004. "Information literacy, statistical literacy and data literacy." *IASSIST Quarterly* Summer/Fall: 6–11.

Sperry, J. 2018. "Data literacy: Exploring economic data." *2018 Economic Programs Webinar Series.* U.S. Census Bureau.

Sternkopf, H. 2017. *Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-Governmental Organizations.* Master's thesis, Berlin School of Economics and Law.

Sternkopf, H., and R.M. Mueller. 2018. "Doing good with data: Development of a maturity model for data literacy in non-governmental organizations." In *Proceedings of the 51st Hawaii International Conference on System Sciences.* Available at: http://hdl.handle.net/10125/50519 (accessed January 3, 2019).

Wolff, A., D. Gooch, J.J. Cavero Montaner, U. Rashid, and G. Kortuem. 2016. "Creating an understanding of data literacy for a data-driven society." *The Journal of Community Informatics* 12 (3): 9–26.