

Unclassified ~~Choose sensitivity level~~

NRC-CMRC

Data Literacy

Report No.: [Click here to enter text.](#)

Date: 21 April 2022

Author: Stephen Downes

Digital Technologies Research Centre



National Research
Council Canada

Conseil national de
recherches Canada

Canada

Executive Summary

Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner. It includes the skills necessary to discover and access data, manipulate data, evaluate data quality, conduct analysis using data, interpret results of analyses, and understand the ethics of using data.

These are core skills required to support key competencies in intelligence and trend analysis, mission-driven metric reporting, health and human response to stress and injury, training and development functions, deployment, supply management and logistics, and information warfare, to name a few.

Competency Model or Framework

Data literacy includes the skills necessary to discover and access data, manipulate data, evaluate data quality, conduct analysis using data, interpret results of analyses, and understand the ethics of using data, where by *data* we mean the representation of facts in media. Data can be structured or unstructured, can be analyzed as a batch or in a stream, and can be presented in media as varied as text, numbers, graphics, images, sound, or video.

Data literacy derives from, and can be seen as including, skills and competencies related to data management, data-based analysis and decision-making. It clear that data literacy involves much more than 'reading' and 'writing' with data and includes but is not limited to the framing of the problem or context of use, the data set itself, application, and testing. As such, it is useful to think of data literacy in the context of the full data workflow, models of which can be drawn from data analytics, machine learning, and statistical analytics.

With the full workflow in mind, this study analyzed the sets of data literacy competencies and skills defined across 20 separate studies and, while finding no unanimity or agreement among the full list, was nonetheless able to identify broad similarities. In particular, to evaluate the definition of data literacy already in use by Databilities, this study indicates a need to consider several additional key competencies, for example: data ethics, data governance, and data systems and tools.

Additionally, this study finds that data literacy is a concept that can be applied equally to both *individuals* and *organizations*, though both the description of data literacy as well as the assessment of data literacy will vary in the given context. Framing elements of data literacy as competencies, and employing a widely used model describing *knowledge, skills and attitude*, an overall framework for describing individual data learning competencies and organizational data literacy capabilities is proposed.

Evaluation or Assessment Framework

It is important to be able to evaluate or assess the level of data literacy competencies individually or across the organization for the purpose of assessing operational readiness and for the purpose of planning future training and development. This report first provides an overview of some data literacy assessment programs, then considers some data literacy assessment models, and finally considers some data literacy methods.

Though they vary in content and format, the assessment models considered are consistent in the requirement that assessments be based on a formal, or structured, representation of the knowledge being assessed. There is however little if no agreement on what such a model should look like. Though the formal creation and validation of such a model is beyond the scope of this study, it considers the essential elements of such a model to offer insight on what such a model would look like.

What is offered here is a model based on a slightly modified full list of competencies drawn from the data literacy studies cross-referenced with a comprehensive skills taxonomy. To this end, and for the sake of consistency with much of the work done previously, a slightly modified version of Bloom's taxonomy is used to organize them. This provides a format framework for assessment and evaluation.

It is arguable that a single-factor measure of data literacy 'levels' as employed by numerous data literacy assessment schemes is insufficient to account for the variability in both the set of data literacy competencies and also the varying degree to which each competency is required in different job functions or roles. Accordingly, a role-defined data literacy model is proposed. This model illustrates the calculation of a role-defined data literacy profile, as well as the process used to create *actual* competency profiles.

This report also describes and assesses the relative merits of four major forms of assessment identified in the literature: self-report, skills test (open response), skills test (multiple choice), and analysis. It describes means of independently validating the reliability and validity of assessment instruments already well-established in the literature.

Teaching Framework

There are few data literacy training initiatives extant, and no organization or institution-wide examples were found. So, in the context of data literacy development in a Canadian Forces context, two areas of consideration are important: models and designs for data literacy program development in general, and examples of extant data literacy training programs and curricula.

The development of data literacy in an organization occupies a space between two extremes. On the one hand, we may find data literacy among other types of information and communication competencies, such as digital literacy or information management programs. On the other hand, we might think of data literacy as a first step in the development of higher-level competencies such as data architect or information management. Either approach envisions a large-scale and complex learning initiative.

But it need be neither, provided we think of data literacy not such as knowledge or content to be used, but rather, as a part of other processes and strategies employed to achieve real objectives or outcomes. This accords with the recommendations found in the literature, for example, to focus on performance rather than content knowledge and to ensure it encompasses real operational challenges using authentic data and examples.

By way of example, we look briefly at five data literacy pilot projects evaluated in current literature, and survey some services and resources supporting data literacy training.

Contents

Executive Summary	3
Competency Model or Framework	3
Evaluation or Assessment Framework	3
Teaching Framework	4
Contents	5
Competency Model or Framework	6
Data Literacy	6
Data	7
Data Literacy Models	8
Data Workflows	10
Competency Models	11
Defining Data Literacy	14
Evaluation or Assessment Framework	16
Assessment Programs	16
Data Literacy Model-Based Assessment	19
Levels	22
Role-Defined Data Literacy	23
Assessment Methods	23
Validity and Reliability	27
Teaching Framework	27
Developing Data Literacy	27
Data Literacy Programs	28
Teaching and Learning Methods	30
Individual Learning Resources	31
References	33
Appendices	39
Appendix 1	Error! Bookmark not defined.

Competency Model or Framework

Data Literacy

A Brief History

The topic of data literacy began to emerge from discussions of database and data management, digital literacy and information literacy in the mid 2010s.

A major early definition was offered by Ridsdale, et al. (2015) working at Dalhousie University: “Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner” (p. 2). The authors adopted a competency-based characterization:

“We define the core skills and competencies that comprise data literacy, using a thematic analysis of the elements of data literacy described in peer-reviewed literature. These competencies (23 in total) and their skills, knowledge, and expected tasks (64 in total) are organized under the top-level elements of the definition (data, collect, manage, evaluate, apply) and are categorized as conceptual competencies, core competencies, and advanced competencies.”

Writing from the Open University, Wolff, et al. (2016) expanded the definition. “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.” They define data literacy in terms of skills and outcomes:

“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.”

In a report from Statistics Canada, Bonikowska, Sanmartin and Frenette (2019) adopt a competencies-based approach derived from a comparison of existing digital competencies frameworks. They write that data literacy is the

“ability to derive meaningful information from data” (Sperry 2018). “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...”

Gartner (2019) addresses data literacy from a management perspective. “Data literacy” is formally called out as a new core competency as part of a clear commitment to the organization and leadership valuing “information as a strategic asset.” In a data literate organization, training programs (online and/or in-person; internal and/or external) are available and supported across all required levels of proficiency. (Duncan, et al. 2021)

Canada’s Department of National Defense, observing that “literacy broadly means having competency in a particular area,” argues that “data literacy includes the skills necessary to discover and access data, manipulate data, evaluate data quality, conduct analysis using data, interpret results of analyses, and understand the ethics of using data.” (2019)

Major Themes

The following major themes emerge from the discussion of data literacy over the last decade:

- Data literacy as a set of skills or competencies
- The idea of deriving meaningful information from data
- The data lifecycle or data workflow

- Complexity of skills for differing roles
- Data literacy as individual and corporate capacities

Data

The core concept in data literacy is ‘data’. But what is data? The word has been defined extensively over the years. Some useful definitions include:

- “An object, variable, or piece of information that has the perceived capacity to be collected, stored, and identifiable.” (Bhargava, et al., 2015)
- “The representation of facts as text, numbers, graphics, images, sound, or video” (The Department of National Defence and Canadian Armed Forces Data Strategy, 2019)

What is important here is to note the distinction between data and *statistics*. The latter are quantifications based on counting or surveys. Data, however, may take many forms, and need not be numerical only.

Types of Data

Structure

Data is ‘structured’ if data elements contain a consistence sequence of parts. For example, structured data could consist of a formally defined array of individual values, a hash or table of key-value pairs, or a set of subject-verb-object triples, as found in Resource Description Format (RDF). Semi-structured data is obtained from less formally defined sequences of parts, such as defined by an input form, a news article, or application programming interface (API). Unstructured data, also referred to as ‘blobs’, are, from the perspective of data management, one single undifferentiated whole (Lorent, 2021)

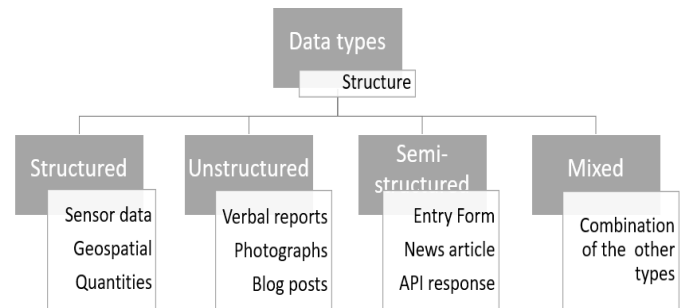


Figure 1 - Data Types: Structure

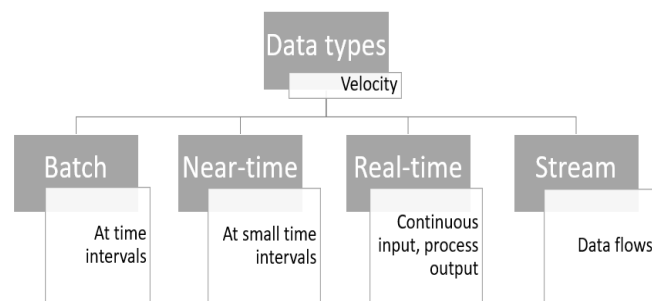


Figure 2 - Data Types: Velocity

Velocity

The ‘velocity’ of data refers to the rate at which it is received, either as original data, or as updates to existing data. Velocity may vary according to how the data is received by a data management system, and how it is output by that system. If all data is received at a specific time interval, it is referred to as ‘batch’ data. Conversely, if data flows in and out of a system continuously, it is referred to as a ‘stream’.

Formats

As mentioned previously, data may exist in non-numerical formats, including text, images, audio, video, multimedia, and more. Data formats are often defined according to their corresponding media format such that a media type (also known as a Multipurpose Internet Mail Extensions or MIME type) indicates the nature and format of a document, file, or assortment of bytes. MIME types are defined and standardized in IETF's RFC 6838 (Freed, et al., 2013). Data types may also be characterized less formally, according to the use to which the data is put, for example, as observational, experimental, derived (or compiled), simulation, or reference (OSU, 2021).

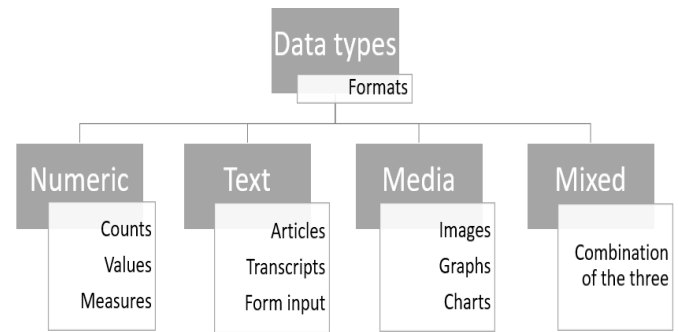


Figure 3 - Data Types: Format

Data Model

A data model is an abstract model that organizes elements of data and standardizes how they relate to one another and to the properties of real-world entities (Nikou & Tziachris, 2022). The data model will define the structure and format of the data, as described above, identify principles or rules embodied in the data, and form the basis for the application of the data in practical contexts. It forms the underlying logic for the 'model-view-controller' (MVC) database model employed by most data management systems today.

Data Literacy Models

Most work in data literacy falls into one of several models or interpretations. "They each have a different focus which tends to reflect the context in which it was derived. They also have a different level of granularity, not just between the definitions, but also within them." (Wolff, et al., 2016) Schield (2004) describes these as 'perspectives', for example, the 'critical thinking' perspective and the 'social science data' perspective.

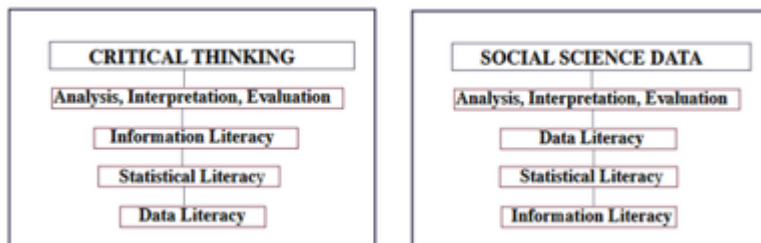


Figure 4 - Data Literacy Models. Schield, 2004, p.8

This section describes the models or perspectives found in the literature.

Data Stewardship Model

This model describes approaches to data literacy that **emphasize data acquisition, curation, quality and deployment**. A prototypical example of this approach is the Statistics Canada descriptions of data quality and the data journey.

"Data literacy also means having the knowledge and skills to be a good data steward including the ability to assess the quality of data, protect and secure data, and their responsible and ethical use." (Statistics Canada, 2020)

Analysis and Decision-Making Model

This model is focused mostly on the ***use of data to support analytics and decision-making***. The best current example of this model is the collection of approaches taken by members of the Data Literacy Project, including Qlik (a data analytics company), Accenture, Cognizant, Experian, Pluralsight, the Chartered Institute of Marketing, and Data to the People. The approach is best summarized by the Data Literacy Project website: “data answers questions faster, provides powerful insights and drives smarter critical decision-making.”

A good example of the sort of definition informing this model is found in Vahey et al. (2006): “data literacy includes the ability to formulate and answer questions using data as part of evidence-based thinking; use appropriate data, tools, and representations to support this thinking; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions.”

While there is overlap between this model and the information literacy model, the academic basis for this model is derived from work in critical thinking, drawing for example from Deahl (2014) who defines a skill set for data literacy that includes ‘understanding data’, checking data for ‘bias and inaccuracy’, and ‘supporting arguments using data’ (pp. 41-42).

Information Literacy Model

“According to Hunt (2004), ***data literacy education should borrow heavily from information literacy education***, even if the domain of data literacy is more fragmented than the field of information literacy.” (Koltay, 2016).

Similarly, Maybee & Zilinski (2016) write, “The emerging construct of data literacy has typically been closely related to information literacy.” Drawing from information literacy coursework, researchers “have used the ACRL’s (2000) information literacy standards and initial research findings to suggest lists of competencies for data literacy... the various models are comprised of skills related to accessing, managing, communicating, preserving, and ethically using data for both undergraduate and graduate education.” (Ibid).

Science and Research Data Literacy Model

This model of data literacy emphasizes ***aspects of data related to computer science, mathematics and statistics***. Though based on principles of data stewardship, it defines a set of data skills including data awareness, forms of statistical representation, the ability to analyze, interpret and evaluate statistical information, and communication of statistical information (Australian Bureau of Statistics, 2010). Koltay (2016) writes that related concepts include research data services, data governance (he notes, “valid data governance may require identifying ‘people who informally already have a level of accountability for the data they define, produce and use to complete their jobs or functions’” p. 305).

Maybee & Zilinski (2016) write, “challenges have been made concerning the efficacy of generic approaches to information literacy, such as the standards, for enabling people to use information in the various contexts in which they live and work (e.g., Bruce, 1997; Lloyd, 2010).” These objections have led to more context-specific models of data literacy development, and in particular to (Bruce, 2008) “data informed learning as an approach to data literacy that shifts the focus from acquiring generic data-related skills to students learning how to use data in disciplinary contexts.”

Social Engagement Model

This model ***distinguishes between the need for everyday uses of data from the deeper requirements of data science***. It may be a nascent model. Thus far, it is only really articulated in a single source (Rahul Bhargava, et. al., 2015), though it has its origins in a broader definition of literacy, as exemplified by Robinson

(2005), who talks of literacy as enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society” (p. 13).

Data Workflows

What *is* data literacy? In this section we look more closely at the nature of artificial intelligence and machine learning, two disciplines already defined by their relation to data, to understand what might be understood as the full ‘data workflow’. This section makes it clear that data literacy involves much more than ‘reading’ and ‘writing’ with data and includes but not limited to the framing of the problem or context of use, the data set itself, application and testing.

Applications of Data Analytics

With the employment of big data analytics for machine learning and artificial intelligence a range of analytics applications have been defined, including especially the following:

- *Diagnostic* - draw an inference about a piece of data based on the patterns detected in sample or training data, for example, to perform recognition, classification or categorization tasks.
- *Predictive* - answer the question, what will (probably) happen, based on an identification of patterns and trends in existing data, and an extrapolation of that pattern or trend to probably future states.
- *Prescriptive* - recommend solutions generated from rules or principles, limits or bounds of operation, equations or balancing mechanisms, or user input.
- *Generative* - employ previous analyses of data in other to generate original content based on parameters or properties of the data studied, combined with predictions or requirements for future data.
- *Deontic* - look at expressions of sentiments, needs, desires, and other such factors to determine what sort of outcome would be best, and then works toward achieving that outcome.

Machine Learning Engineering

The term ‘machine learning’ was coined by Arthur Samuel (1959) for IBM to describe the application of statistical algorithms to learning problems, for example, how to play checkers. In such a case, ‘data’ consists of the sequence of moves and potential moves of an opponent, and the context of application is defined by the rules for playing checkers.

Following this model, the typical machine learning application is comprised of three major elements: the data itself, the machine learning model, and the algorithm or code used to apply the model and data into a final product. Machine learning engineering describes the construction and use of these three elements:

- *Data engineering*, which describes the acquisition, exploration, cleaning, labeling and management of data
- *Model engineering*, which consists of the development or training of the model, testing and evaluation, and packaging for use in an application
- *Deployment*, which describes how the model is served and used, performance evaluation, and performance logging (Visengeriyeva, et al., 2022)

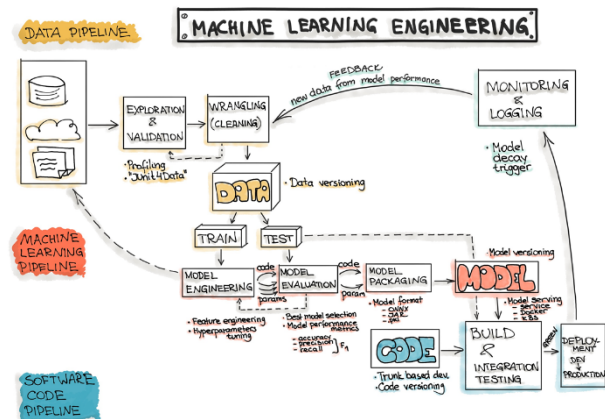


Figure 5 - Machine Learning Engineering. Visengeriyeva, 2021

Statistical Research and Analysis

Statistical research methods workflows emphasize “the importance of asking questions throughout the statistical problem-solving process (formulating a statistical investigative question, collecting or considering data, analyzing data, and interpreting results), and how this process remains at the forefront of statistical reasoning for all studies involving data.” (Bargagliotti, et al., 2020).

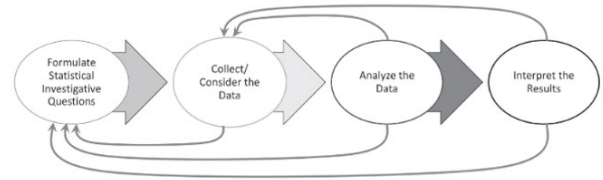


Figure 6 - GAISE Workflow. Bargagliotti, et al., 2020

It also emphasizes “consideration of different data and variable types, the importance of carefully planning how to collect data or how to consider data to help answer statistical investigative questions, and the process of collecting, cleaning, interrogating, and analyzing the data.” In this respect it resembles machine learning and data analytics workflows (above and below).

The Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) by American Statistical Association define four major elements of the workflow:

- *Question* formulation
- *Data collection* and investigation
- *Data analysis*
- *Interpretation* of results

Big Data Analytics

Big data analytics follows the development of machine learning analytics and adapts it for very large bodies of data. Thus, while the workflow is similar, additional elements are required to compose and comprehend the data and data models. A big data analytics workflow management application will consist of the following elements:

- *Data sources*, including a database management system (DBMS), data warehouse (consisting of data ‘streams’, ‘pools’ and ‘lakes’)
- *Data management tools* for preprocessing, filtering, cleaning, and transformation
- *A data modeling tool*, including model definition and evaluation
- *Results analysis* and visualization tools

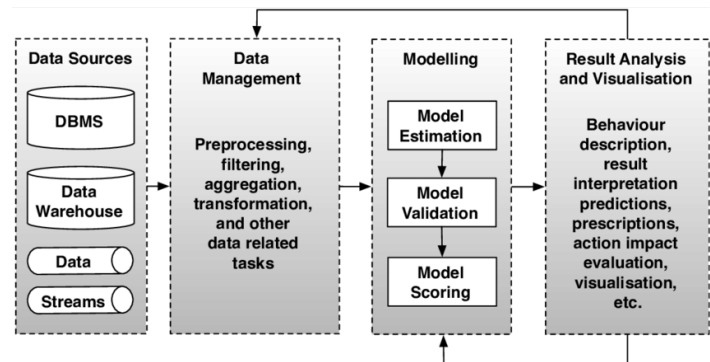


Figure 7 - Big Data Analytics Workflow. Assuncao, et al., 2014, p. 4

(Assuncao, et al., 2014)

Competency Models

The analysis in this section combines the work outlined in the previous sections to achieve a synthesis of the concept of ‘data literacy’. For the purposes of this presentation, data literacy is defined as a set of one or more ‘competencies’, though it should be clear that not all studies reported above do so. We begin with a brief outline of the concept of the ‘competency’ and then proceed with the analysis.

Competencies

Competencies are commonly defined as “a set of basic knowledge, skills, abilities, and other characteristics that enable people at work to efficiently and successfully accomplish their job tasks.” Following Oberländer, et al. (2020) we use the term ‘competencies’ here to draw on a well-established concept that includes knowledge, skills, abilities, and other characteristics (KSAO).

The concept of competencies also includes the requirement of *evidence* for competencies. Thus, employing a definition using competencies is well suited to a discussion of data literacy that includes the fostering and assessment of knowledge, skills, and abilities.

Analysis

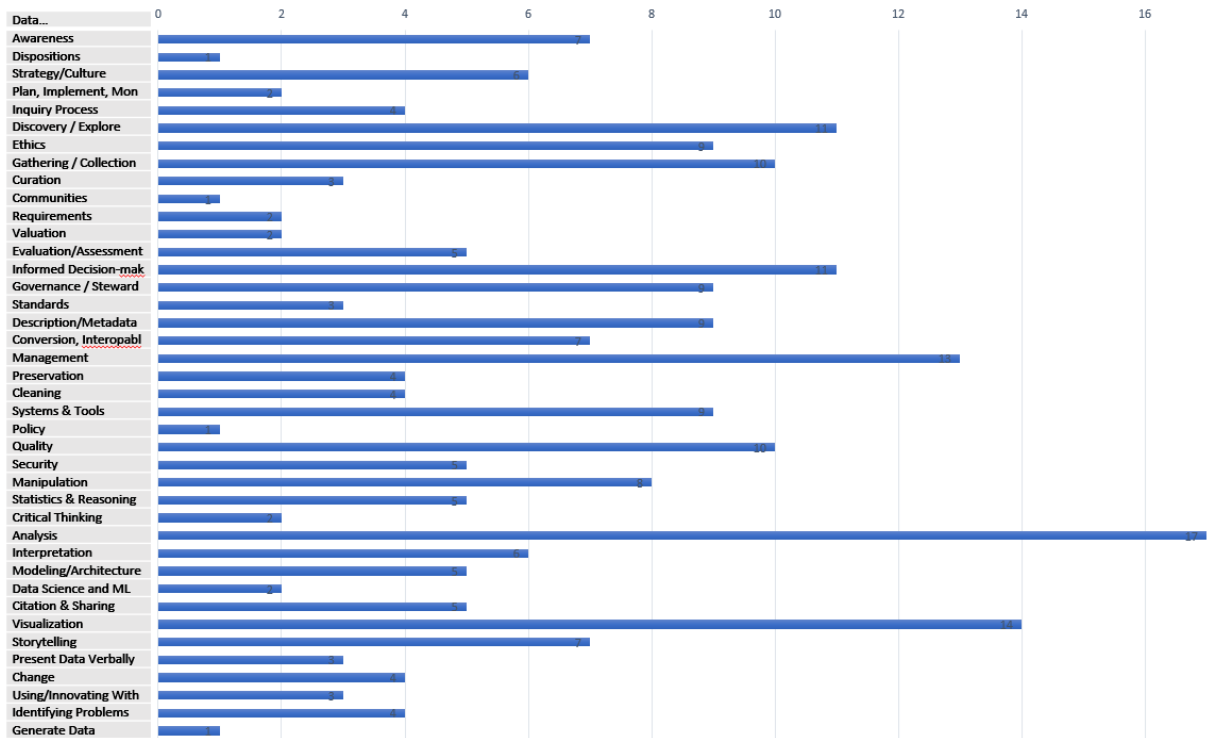
We drew on 20 studies that offered a (more or less) competency-based definition of data literacy and compared the set of competencies each proposed. The selection of sources was intended to draw from and be representative of various data literacy models (see above). The full set of studies is described and summarized in the appendix below.

In assigning the competencies interpretation was required, as the studies did not all employ the same terminology. Accordingly, a brief outline of what is meant by each competency defined is contained in an appendix below. Below is the result of the analysis:

Data...	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Awareness	x		x								x			x			x		x	x
Dispositions																				x
Strategy/Culture	x			x		x		x			x				x					
Plan, Implement, Mon														x				x		
Inquiry Process														x				x	x	
Discovery / Explore			x		x			x	x		x		x	x	x	x	x			x
Ethics	x		x			x			x	x	x			x	x	x				
Gathering / Collection			x	x					x	x	x		x	x		x			x	x
Curation										x	x								x	
Communities						x														
Requirements		x							x											
Valuation								x											x	
Evaluation/Assessment					x	x					x						x		x	
Informed Decision-mak	x		x			x	x	x		x	x		x	x			x		x	
Governance / Steward	x	x	x	x		x	x		x									x	x	
Standards	x					x			x											
Description/Metadata						x		x	x	x	x		x					x	x	x
Conversion, Interopabl	x										x				x			x	x	
Management		x			x	x	x	x		x	x		x	x		x	x	x	x	
Preservation											x					x				x
Cleaning			x			x				x	x									
Systems & Tools	x		x			x					x			x		x	x	x	x	
Policy						x														
Quality	x								x		x	x	x	x	x	x	x	x		
Security	x	x					x			x	x									
Manipulation		x									x		x	x	x	x		x	x	
Statistics & Reasoning								x		x	x							x	x	
Critical Thinking											x									
Analysis	x	x	x			x	x	x	x	x	x				x	x	x	x	x	x
Interpretation											x			x	x	x	x		x	
Modeling/Architecture			x	x			x		x										x	
Data Science and ML							x	x												
Citation & Sharing											x	x				x		x		x
Visualization	x	x	x			x			x		x	x	x		x	x	x	x	x	x
Storytelling	x		x					x	x		x	x							x	
Present Data Verbally											x				x					
Change						x	x	x												x
Using/Innovating With					x	x		x												
Identifying Problems																				
Generate Data											x		x	x	x			x		

Figure 8 - Competency Analysis Table

The diagram below illustrates the frequency of each competency:



25

Figure 9 - Competency Frequency Table

In light of the fact that a small-scale pilot version of a Canadian Armed Forces data literacy survey was conducted last year in conjunction with Databilities, it is particularly relevant to compare the list of competencies identified by that company with the overall list of competencies. Here is the comparison:

Data...	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Awareness	x		x								x			x			x		x	x
Dispositions																				x
Strategy/Culture	x			x		x		x			x				x					
Plan, Implement, Mon														x				x		
Inquiry Process														x			x		x	
Discovery / Explore			x		x			x	x		x		x	x	x	x	x			x
Ethics	x		x		x			x	x	x				x	x	x				
Gathering / Collection			x	x					x	x	x		x	x		x		x	x	
Curation										x	x								x	
Communities						x														
Requirements		x								x										
Valuation								x										x		
Evaluation/Assessment					x	x					x						x		x	
Informed Decision-mak	x		x			x	x	x		x	x		x	x			x		x	
Governance / Steward	x	x	x	x		x	x		x									x	x	
Standards	x					x			x											
Description/Metadata						x		x	x	x			x					x	x	x
Conversion, Interopabl	x									x	x		x		x		x	x	x	
Management		x			x	x	x	x		x	x		x	x		x	x	x	x	
Preservation											x					x				x
Cleaning			x			x			x	x										
Systems & Tools	x		x			x					x			x		x	x	x	x	
Policy						x														
Quality	x								x		x	x	x	x	x	x	x	x		
Security	x	x					x			x	x									
Manipulation		x									x		x	x	x	x		x	x	
Statistics & Reasoning								x		x	x							x	x	
Critical Thinking											x									
Analysis	x	x	x			x	x	x	x	x	x		x	x	x	x	x	x	x	x
Interpretation											x		x	x	x	x				x
Modeling/Architecture			x	x			x		x											x
Data Science and ML							x	x												
Citation & Sharing											x	x				x		x		x
Visualization	x	x	x			x			x	x	x	x			x	x	x	x	x	x
Storytelling	x		x					x	x		x	x								x
Present Data Verbally											x				x					
Change						x	x	x												x
Using/Innovating With					x	x		x												
Identifying Problems											x		x	x	x					
Generate Data																		x		

Figure 10 - Competency Frequencies Mapped to Databilities

Column 13 represents the list of competencies identified by Databilities. We can see that they identify several that are frequently cited by other studies, including data discovery and exploration, data-informed decision-making, data management, data analysis, and data visualization. However, it is significant that three major competency areas are *not* covered by the Databilities model, as highlighted in orange: data ethics, data governance or stewardship, and data systems and tools.

The list of competencies identified also makes it clear that data literacy does not fall into any single category described above. It contains elements of critical thinking, statistical reasoning, data management, and scientific research. Data literacy therefore represents a certain level of competency across a broad range of data-related skills, not a narrowly defined subset of some other type of literacy.

Defining Data Literacy

Individual and Organizational Data Literacy

Another trend evident in the literature review was the distinction between *individual* data literacy and *organizational* data literacy. The former was most often discussed by papers and research studies focused on education and training, while the latter was most often treated by publications, especially from consulting firms, discussing organizational and management strategies. This shows the importance of not depending on one particular type of literature or branch of inquiry.

Just as we say a range of data literacy models, above, organizational data literacy can also be represented using various models. For example, organizational data literacy may be subdivided according to organizational unit, for example, ‘team’, ‘division’ or ‘branch’, and described according to the scope of that unit’s activities.

Alternatively, it can be described according to organizational properties, such as tools and systems, employee skills and capabilities, and organizational procedures and mechanisms.

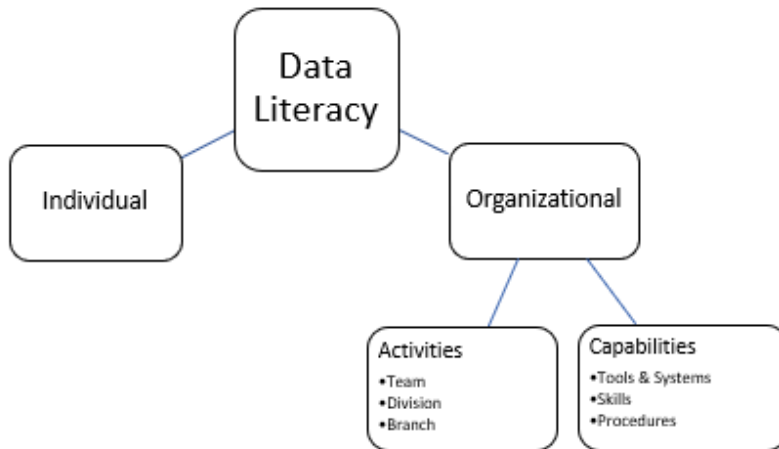


Figure 11 - Classifications of Organizational Data Literacy

For the purposes of this report, it was determined that *each of the competencies* referenced in the previous section could be said to be defined either from an individual or organizational perspective. But what does that mean from the perspective of defining each individual competency?

Competencies

It is beyond the scope of this report to analyze the concept of the competency. To this end, we employ a widely used model describing *knowledge, skills and attitude*. Specifically, “a competency is a set of skills, related knowledge and attributes that allow an individual to successfully perform a task or an activity within a specific function or job” (Salleh, et. al., 2015, United Nations Industrial Development Organization (UNIDO), 2002, 10).

As will be relevant below, it should be noted that competencies are described in ways that are observable and measurable, linked to the context or environment, transferable (i.e., can be learned and applied in new contexts), and based on performance.

To define organizational competencies, an analogous set of elements was defined: organizational or operational definitions; organizational systems, tools and capacities; and organizational practices. This definition is informal, and these terms could be modified or refined through additional research.

Individual	Organizational
Knowledge	Definitions
Skills / Competencies	Capacities
Attitudes	Practices

Figure 12 - Individual and Organizational Data Literacy

Example: Data Visualization

To obtain the full competency definition, and hence, to fully define data literacy, each of these six attributes could be provided. It is understood that the emphasis and importance of these attributes may vary from context to context and according to the competency being defined. Below is an example of the attributed defined in the specific case of 'data visualization':

Individual	Organizational
Knowledge <ul style="list-style-type: none"> - knows visualization formats - understands data representation 	Definitions <ul style="list-style-type: none"> - standard visualizations for key data - visualizations referenced to original data
Skills / Competencies <ul style="list-style-type: none"> - can create visualizations from data - can generate meaning from visualizations 	Capacities <ul style="list-style-type: none"> - staff have access to data visualizations - staff includes data visualization expertise
Attitudes <ul style="list-style-type: none"> - is comfortable working with visualizations - recognizes importance of visualizations 	Practices <ul style="list-style-type: none"> - maintains data visualization software tools - data visualization part of reports workflow

Figure 13 - Individual and Organizational Data Literacy Competencies and Capabilities

Evaluation or Assessment Framework

It is important to be able to evaluate or assess the level of data literacy competencies individually or across the organization for the purpose of assessing operational readiness and for the purpose of planning future training and development.

In this section we first provide an overview of some data literacy assessment programs, then consider some approaches or models of data literacy assessments, and finally consider some data literacy methods.

Assessment Programs

Following are some examples relevant assessment programs illustrating key features of competency assessment in data literacy and related areas.

OECD

Although not addressing data literacy specifically, it is worth considering in this context the OECD Programme for the International Assessment of Adult Competencies (PIAAC) and the Programme for International Student Assessment (PISA). PISA in particular “has become accepted as a reliable instrument for benchmarking student performance worldwide, and that PISA results have had an influence on policy reform in the majority of participating countries/economies” (Breakspear, 2012). While PISA is specifically focused on 15-year-old students, PIAAC, assesses “measures adults’ proficiency in key information-processing skills - literacy, numeracy and problem solving - and gathers information and data on how adults use their skills at home, at work and in the wider community” (Kirsch & Thorn, 2016).

These tests are standardized objective tests, that is, they are defined to measure a specific range of competencies, and the assessments present questions to be answered or tasks to be completed, similar in form to a test or examination at a school or university. Specifically, in the case of PIAAC literacy assessment,

participants were asked to “access and identify tasks require respondents to locate information in a text, integrate and interpret tasks involve relating parts of one or more texts to each other, and evaluate and reflect tasks require the respondent to draw on knowledge, ideas or values” (Kirsch & Thorn, 2016, 2.2.1.3)

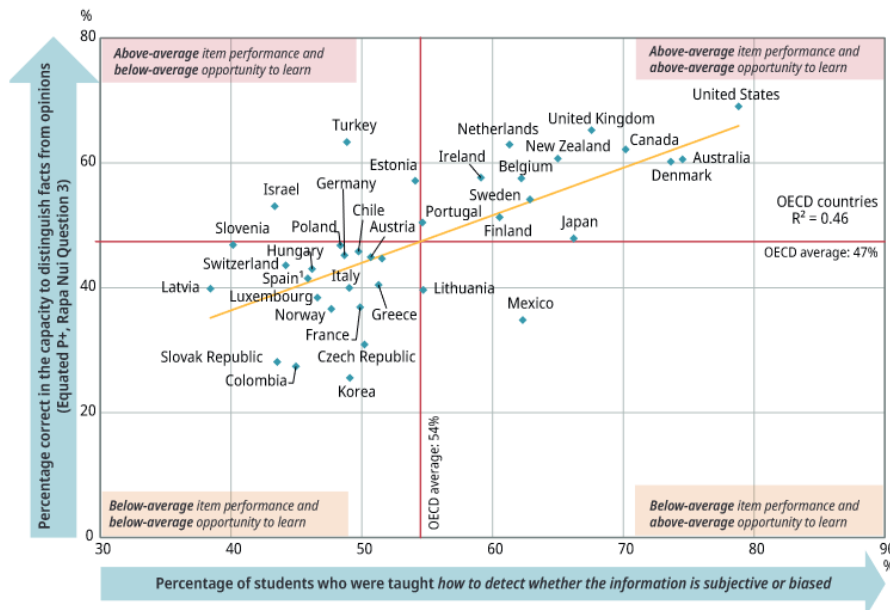


Figure 14 - PIACC Assessment. Kirsch & Thorn, 2016, 2.2.1.3

OECD has begun to address some questions more directly related to competencies related to data literacy, as for example in its study of the question “are 15-year-olds prepared to deal with fake news and misinformation?” It reports that “An average of 54% of students in OECD countries reported being trained at school on how to recognise whether information is biased or not,” with Canada ranking with above-average performance abilities and opportunity to learn (see diagram) and that “education systems with a higher proportion of students who were taught whether information is subjective or biased were more likely to distinguish fact from opinion” (Suarez-Alvarez, 2021 113).

Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Endorsed by the American Statistical Association, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) emphasize that there is no one route to teaching and assessing statistical literacy and notes that “mastering specific techniques is not as important as understanding the statistical concepts and principles that underlie such techniques” (GAISE, 2016, 8).

The framework of essential concepts and 22 examples emphasize the integration of statistical reasoning in the context of real-world examples. Students are asked about the investigative process, which includes formulating questions, considering data, analyzing data, and interpreting results, a pattern resembling the data analytics workflow described above. These goals “require assessments of the students’ statistical understanding through their written communication. For example, students should be able to interpret and draw conclusion from standard output from statistical software” (Johnson, 2018).

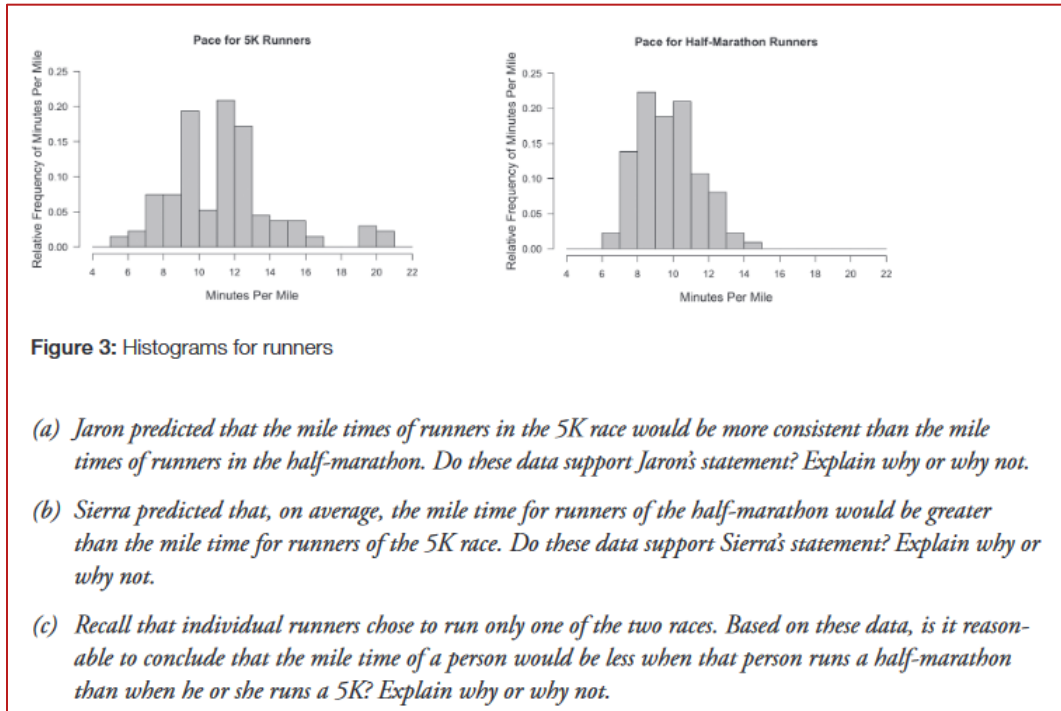


Figure 15 - Sample Statistical Literacy Assessment. Bargagliotti, et al., 2020, p. 110.

For example, in the assessment above, the student is presented statistical data in the form of a graph and asked to answer questions interpreting the information depicted. The question involves recognition of patterns in the data as well as the use of these patterns for prediction or extrapolation.

Eckerson Group Data Literacy Imperative

By contrast with the OECD and GAISE programs, the Eckerson Group describes data literacy assessment specifically and includes assessment not only of individual data literacy but also of the organization (Wells, 2021). Assessments are based initially on a comprehensive Data Literacy Body of Knowledge (DLBOK) defined by the organization. The DLBOK is used for gap analysis and training program development.

“Individual assessment is only the beginning. It is the foundation upon which organizational assessment is built, and organizational assessment is an essential process when building a culture of data literacy. Literacy assessment with business impact is performed at three levels—by individual, by role, and by group.”

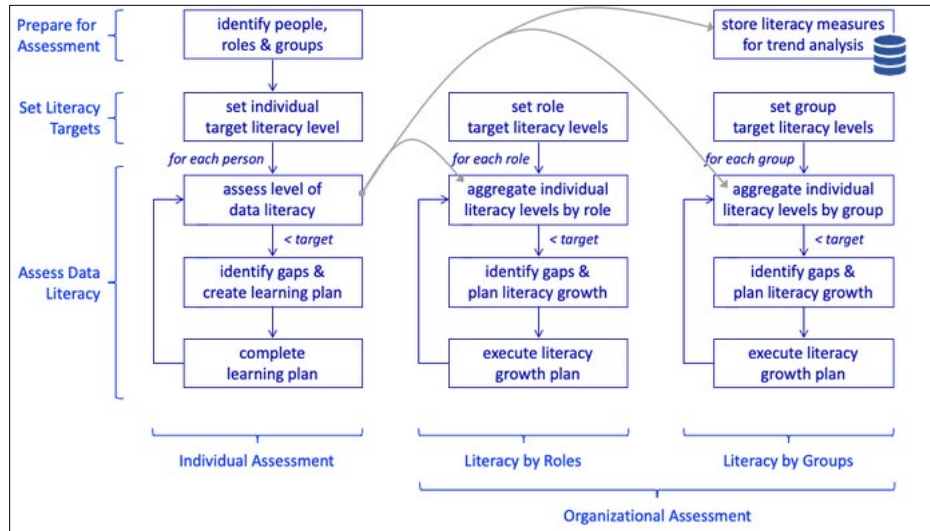


Figure 16 - Data Literacy Assessment Process, Wells, 2022

A data literacy assessment based on the Eckerson DLBOK can be viewed at eLearning Curve (2022). It offers a set of 50 questions testing the respondent’s data literacy knowledge, skills and attitude through questions about terminology, processes, tools, functions, and expectations.

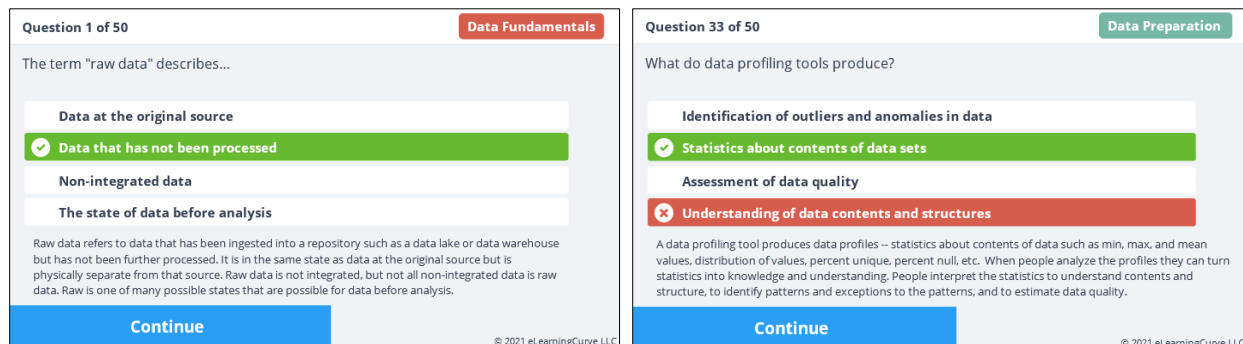


Figure 17 - Data Literacy Assessment Example, eLearningCurve, 2022

Data Literacy Model-Based Assessment

The examples in the previous section, though they vary in content and format, are consistent in the requirement that assessments be based on a formal, or structured, representation of the knowledge being assessed. There is however little if no agreement on what such a model should look like.

The formal creation and validation of such a model is well beyond the scope of this study. However, it is important to consider the essential elements of such a model and to offer insight on what such a model would look like, for the purpose of further discussing the assessment of data literacy and mechanisms for developing or improving it.

Data Literacy Model-Based Assessment

In the analysis of data literacy competencies described in the first section of this report we obtained an unstructured list of competencies. These competencies were organized into different categories by various studies, but there was no consistency whatsoever in the categorization scheme from study to study.

What is offered here is a model based on a slightly modified full list of competencies drawn from the data literacy studies cross-referenced with a comprehensive skills taxonomy. Again, there is a range of taxonomies to choose from, and a detailed discussion of these taxonomies is beyond the scope of this report, therefore for the sake of consistency with much of the work done previously a slightly modified version of Bloom’s taxonomy is used (Bloom, 1956).

While Bloom’s taxonomy is best known for six levels of cognitive skills development (from ‘remembering’ to ‘evaluation’) in fact three separate taxonomies were described: cognitive, affective and psychomotor. These can be thought of corresponding with the already-described taxonomy of knowledge, attitudes and skills, respectively.



Figure 18 - Bloom's Taxonomy

There are some caveats to this use. First, while Bloom’s is often thought of in terms of *levels* of achievement, the elements in each of the three domains are interpreted specifically as a *taxonomy*, with no presupposition as to progression through stages or higher degrees of attainment. Progression, rather, ought to be seen as occurring *within* each element (more on this below). This we would say that these elements can be represented as distinct skills or competencies.

Further, this taxonomy needs to be *extended* to accommodate both individual and organizational competencies. For this purpose, we revisit the definition of data literacy competencies from above.

Bloom’s	Individual	Organizational
Cognitive	Knowledge	Definitions
Psychomotor	Skills / Competencies	Capacities
Affective	Attitudes	Practices

Figure 19 - Bloom's Taxonomy Adapted to Individual and Organizational Competencies

The following three subsections briefly expand this model, mootng possible definitions for each element.

Knowledge

Individual	Organizational
Knowledge - Know what data is, recognize data vs non-data	- Has or uses data in some way
Comprehension - Know methods to read data, comprehend data	- Provides mechanisms for data access
Application - Know how data can be used	- Data can be used as input in tools and systems
Analysis - Understand parts of data, types of data	- Data can be accessed in different views, formats
Synthesis - Know ways to join or connect data	- Data can be pooled or connected
Evaluation - Identify quality data, appropriate data	- There are organizational data quality standards
Creation - Create data	- Data is recorded and produced in the organization

Figure 20 - Bloom's Taxonomy for the Knowledge Domain Adapted to Individual and Organizational Competencies

Skills / Competencies

Individual	Organizational
Perception - Be able to discover, read, explore data	- The organization actively collects data
Set - Can follow data processes and procedures	- There are data management processes
Guided Response - Can follow instructions and respond to data	- There is a capacity to respond to data
Mechanism - Knows about and can use data tools and systems	- Data management tools and systems are supported
Complex Overt Response - Can make decisions using data	- Decisions are driven by data
Adaptation - Can create data visualizations, stories	- Visualizations and data stories are used
Origination - Can create and share data from new sources	- The organization regularly collects and shares data

Figure 21 - Bloom's Taxonomy for the Skills Domain Adapted to Individual and Organizational Competencies

Attitudes

Individual	Organizational
Receiving - Is open to learning from data	- Data is welcomed and sought after
Recognizing - Can detect patterns and regularities in data	- Data is considered and analyzed; there are data-based alerts
Responding - Is willing to act on new data	- Data drives actions and responses to challenges
Framing - Is willing to work in a data-centered way	- Knowledge management is data centered
Valuing - Values and can assign value to different types of data	- Data is valued in the organization and quality controls apply
Organizing - Actively orients data to address challenges	- Key strategies are oriented by data
Characterizing - Develops abstractions, generalizations and principles	- Organizational frameworks, structures, procedures driven by data

Figure 22 - Bloom's Taxonomy for the Attitudes Domain Adapted to Individual and Organizational Competencies

Considerations

It was noted in the first section that the Databilities list of competencies does not include, among others, data ethics. It should be observed that the model just described, produced employing a theory-based approach, also does not include data ethics. It may be that Bloom's is an inappropriate taxonomy to employ for this purpose, it may be that the specific list of competencies is not fully described (perhaps, for example, 'ethics' should be added as an 'attitude'), or it may be that data ethics should not properly be considered a data literacy competency.

What such considerations make clear is that any theory-based conceptualization of data literacy is employed, it should be mapped against existing discussions of data literacy, including the studies cited above, and validated through research and discussion with those responsible for managing and enhancing data literacy capabilities.

Levels

Many data literacy assessment models report their results in terms of 'levels', where the level indicates a degree of proficiency.

For example, Means, et al. (2011) categorize teachers' ability to use data to inform instruction as 'below basic', 'basic', 'proficient' and 'advanced'. This test measured, for example, whether teachers could find relevant data in a complex table or graph (basic) or manipulating data from a complex table or graph to support reasoning (advanced).

The NU Data project, "a professional development intervention aimed at preparing special education teams to use data-based decision making to improve academic outcomes for students with disabilities" (Doll, et al., 2014) finds a "data knowledge scale", a "a single factor measure of data literacy" (Sikorski, 2016).

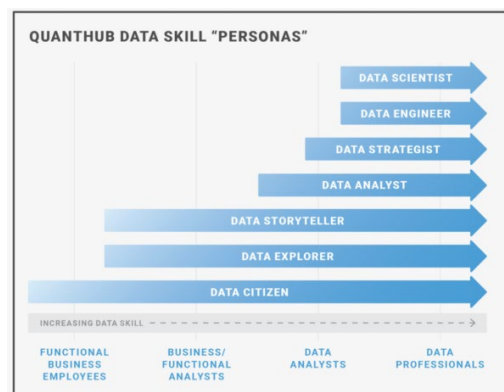


Figure 23 - Levels. DuBois, 2022

Quanthub, which measures data literacy for commercial clients, offers a scale of seven ‘personas’, “each of which represents progressively more sophisticated skill levels up to ‘data scientist’” (DuBois, 2022). DataLiteracy (2021), which “works hand-in-hand with organizations who seek to improve team-based data literacy”, offers a five-level “maturity” rating. Many more examples of data literacy ‘levels’ can be identified in the literature. As in the case of other measures, there is no standard or consistent approach defining how these levels are determined nor even what they mean.

Role-Defined Data Literacy

It is arguable that a single-factor measure of data literacy is insufficient to account for the variability in both the set of data literacy competencies and also the varying degree to which each competency is required in different job functions or roles. Accordingly, a role-defined data literacy model is proposed here.

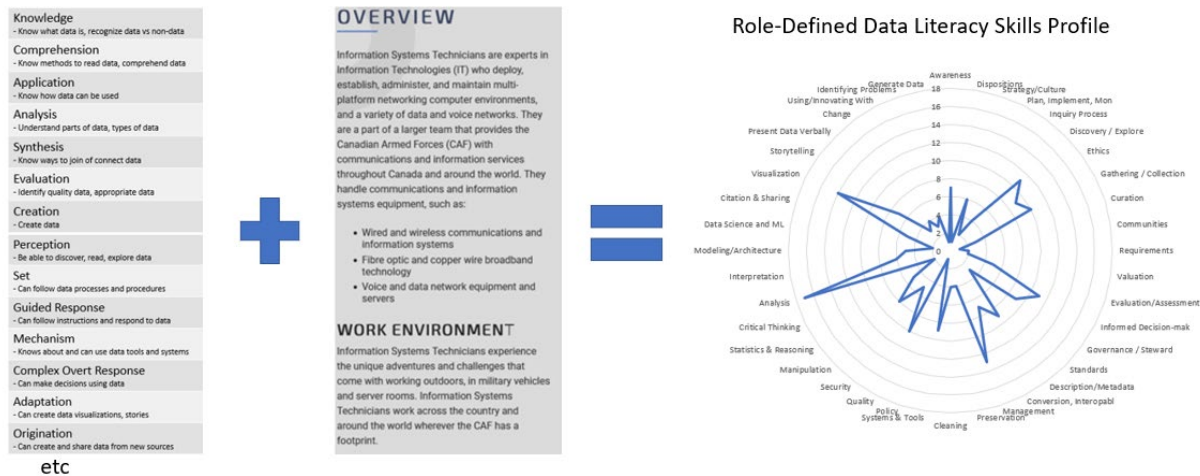


Figure 24 - Defining a Skills Profile

This figure illustrates the calculation of a role-defined data literacy profile. It consists of a combination of the set of competencies as defined in the data literacy model with the actual job or function description. This allows for a definition of the relative importance of each competency for that function, demonstrated here in the form of a radar chart (also known as a spider chart).

As discussed above, the precise definition of data literacy competencies ought to be undertaken in consultation with relevant personnel. Job or function descriptions may be obtained from extant text (the example in the diagram is from the forces.ca Careers page) or drafted as text by managers and those occupying the position. The competency profile may be created by a simply counting of the frequency of relevant terms, or by a more nuanced analysis, perhaps using machine learning.

The same process may be used to create *actual* competency profiles for each individual evaluated, by employing test results or actual communications generated by the person in question (such a process would be subject to ethical and privacy considerations). A similar process may be used to generate organizational level competency profiles.

Assessment Methods

Four major forms of assessment were identified in the literature:

- Self-Report
- Skills Test (Open Response)
- Skills Test (Multiple Choice)

- Analysis

Self-Report

In a self-report assessment, the user is presented with a series of questions about their capabilities to which they respond (presumably) honestly. While used in some cases for individual assessments, such forms of assessment are useful for assessments of organizational capabilities since there is typically no direct or objective means of testing.

The website 'DataLiteracy', for example, offers a "17 Key Traits of Data Literacy Self-Assessment" evaluation

Ask Questions to Collect Data

3. How would you describe your ability to formulate questions to get meaningful insights from data? *

I have no idea what questions can be answered by data.

I am able to formulate questions that can be answered with simple data queries.

I develop projects based on questions that need complex data queries and multiple iterations to resolve.

My 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.

Answered 4/13 questions

Restart Previous Next

Figure 25 - Self-Assessment. Jones, 2019

where for each train respondents use a sliding scale tool to indicate their proficiency and the relative importance of the trait (Jones, 2019). Similarly, a Udemy course offers self-assessment based "Data Literacy Assessment for Every Employee" (Jones, 2021). The Canada School of Public Service offered a very similar "How Data Literate Are You" quiz for federal employees (illustrated below).

There is good reason to be sceptical of self-reported cognitive capabilities, even when respondents are being honest. Subjective assessment may bias responses for a variety of reasons, including a desire to provide the right answer, or as a result of imposter syndrome. But even where no bias is present, "convincing evidence of the association between self-report scales and actual cognitive performance has not been demonstrated" (Williams, et al., 2017).

In research in 2020, NRC and DRDC researchers employed the Databilities data literacy assessment, which employs a self-report method. Aware of this concern, two additional sets of questions were included in the study, a set of 'objective' questions to measure the respondents' actual capabilities (at least in statistical reasoning) and a set of questions intended to measure whether respondents had a bias toward pleasing others. Analysis found a correlation between objective test scores and self-reports. However, such analysis must be conducted with caution to avoid an autocorrelation effect (Fix, 2022).

Skills Test (Open Response)

In an open-response skills test the respondent is asked a question and provided a space in which to provide a response. Typically, there is a correct answer, or minimally, a possibility that some answers may be better than others. Assessment of the response may be based on the factual content expressed in the response (i.e., the answer) or on criteria related to the formulation of the response (e.g., use of evidence or proper argumentation).

For example, students taking the Ontario Secondary School Literacy Test (OSSLT) (EQAO, 2020) might be given a passage to read and asked, “Explain why Montreal’s approach to graffiti is beneficial. Use specific details from the selection to support your answer.” This question would require a specific response (why it’s beneficial) produced in a specific way (using details from the text) (St. Mary’s, 2018). Or the question may require a more complex construction, as in the example illustrated below:

Evaluation criteria for such a question would include completeness (answering all the ‘w’ questions), comprehensiveness (addressing both image and text) and focus (toward a newspaper audience).

Task:	Write a news report based on the headline and picture below. <ul style="list-style-type: none"> You will have to make up the facts and information to answer some or all of the following questions: Who? What? Where? When? Why? How? You must relate your newspaper report to both the headline and the picture.
Purpose & Audience:	to report on an event for the readers of a newspaper
Length:	The lined space provided in the <i>Answer Booklet</i> for your written work indicates the approximate length of the writing expected.

Figure 26 - St. Mary’s Practice Test 2017/2018

An advantage of open-response assessments is that they more closely emulate real-world contexts. For example, being a data-literate teacher means being able to draw open-ended conclusions from a set of data (Athanasas, et al. 2013). The same holds for other professions, including military professions, where there may be no fixed, specific, or ‘right’ answer to questions, only better or worse ways of working with the data.

Indicator	4 Exceeds expectations	3 Meets expectations	2 Needs improvement	1 Inadequate
Describing data: Ability to say what patterns or relationships are seen between data points.	Meets expectations and one or more of the following: Comparison references specific data values beyond the scope of what is asked (e.g., magnitude and/or statistical comparisons). Creatively presents data visualization or uses more advanced tools to show data relationship. Appropriately uses scientific vocabulary to describe patterns.	Description of the data pattern is accurate. Makes quantitative comparison that refers to specific data values. Makes a comparison or describes a pattern that is relevant to the scientific question. Data visualization is properly annotated with descriptions or comparisons, per instructions. <i>A solely qualitative comparison can be scored a 3 only if the scoring guide explicitly notes that a quantitative description is not possible.</i>	Description of the data pattern is attempted, but inaccurate. Makes no quantitative comparison or reference to data values. Makes a qualitative comparison of a pattern relevant to the scientific question. Data visualization is minimally annotated OR annotations are incomplete or incorrect.	Provides little or no description of data patterns. Makes a qualitative comparison, but it is irrelevant to the scientific question. Provides a description that is a summary of background information without use of OT data. Does not include a data visualization OR a visualization with no annotation.

Levels of proficiency for describing data, excerpted from full rubric.

Figure 27 - Levels. Sickler, et al., 2021

Assessment of open-response tests therefore requires a set of evaluation criteria or *rubric*. While accepting that “there is not a clear consensus of what it means to be ‘data literate’,” Sickler, et al. (2021) propose a scoring rubric for the measurement of data literacy skills in undergraduate education. The rubric is based on “data skill indicators,” which on examination are analogous to the data literacy competencies identified in the first section of this report.

A major weakness of open-response assessment is the need for individual and interpretive grading, which requires significant human resources and time,

and may introduce elements of subjectivity and bias. Work is underway in the development of automated assessment, for example, AI-supported essay grading (see, eg. Kumar and Boulanger, 2020).

Skills Test (Multiple Choice)

The intent of multiple-choice tests is to obtain the same quality of assessment as with open-response assessments, but with less effort required on the part of both respondents and graders. Automated grading of multiple-choice tests is well-established and is frequently employed in online courses and web-based resources.

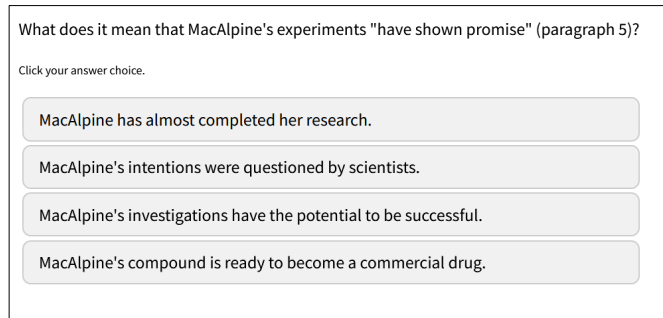


Figure 28 - Multiple Choice. Ontario Secondary School Literacy Test (OSSLT), 2021

While they may be easier to grade, multiple-choice tests are difficult to design. It is important that the responses measure the skills being tested, and not unassociated skills (for example, incidental subject knowledge, or the ability to decipher double negatives). The choices offered need to be plausible, but distinct and easily distinguished by someone with the appropriate skill. It is important to design and develop such tests using recognized methodologies, such as Rasch modeling, which “assumes that the underlying construct that is being measured varies along a single dimension” (Bond & Fox, 2012).

Most data literacy assessments are offered in the form of multiple-choice tests. The eLearning Curve (2022) assessment referenced above is one. Another commercial example can be found by trying the Questionmark Data Literacy test by Cambridge Assessment (2021). Most of the Ontario Secondary School Literacy Test (OSSLT) is in the form of multiple-choice questions, as is the Ontario School data literacy assessment used as ‘objective questions’ in the previously mentioned NRC-DRDC research project.

Analysis

“Content analysis is a method designed to identify and interpret meaning in recorded forms of communication by isolating small pieces of the data that represent salient concepts and then applying or creating a framework to organize the pieces in a way that can be used to describe or explain a phenomenon.” (Kolbe & Burnett, 1991)

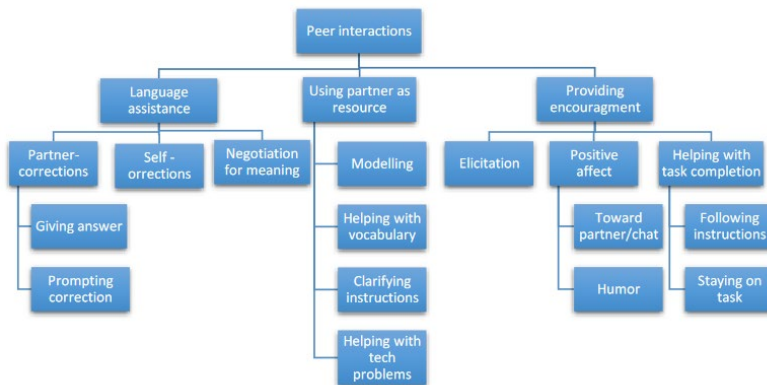


Figure 29 - Content Analysis. Golonka, et al., 2017, p. 161

Content analysis is more often used in the context of research rather than testing and assessment, but as machine learning analysis becomes more prominent, we may expect it to be used more frequently to assess data literacy. In an analysis a body of typically unrelated content is subject to labeling and categorization by researchers to identify semantic, communicative, or cognitive elements. That said, the practice of content analysis as an educational research and assessment tool is well understood in the community (Kleinheksel, et al., 2020).

An illustration of this methodology may be found in the qualitative analysis of chat transcripts to identify peer interaction in text chat (Golonka, et al., 2017). Analysis requires a taxonomy of the entities being identified, as for example provided by a concept map. In the case of data literacy, this set of concepts is analogous to the data literacy competencies described above, and an analysis would contain two parts for each: the name or type of competency being attempted, and the degree to which it was successful.

Some examples of analysis to identify data literacy are extant. Suryadi, et al. (2020) study data literacy in physics students (and based on the analysis find it deficient). Noting that “the meaning of data literacy varies depending on who uses it, and its concept is often conveyed in terms other than data literacy,” Yousef, et al. (2021) use text analysis to identify data literacy communities. Piro and Hutcheson (2014) analyze “changes in perceptions of comfort toward data-literacy behaviors before and after an instructional intervention called a Data Chat.”

Validity and Reliability

As mentioned in passing above, it can be a challenge to ensure that assessments accurately measure the data literacy skills and competencies they are intended to assess. Hence there should be a process to ensure the assessments are valid and reliable.

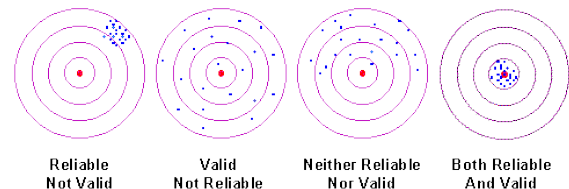


Figure 30 - Validity and Reliability

As is often the case, indices from other disciplines can be useful here. For example, Cohen’s kappa index goes beyond assessor agreement to take into account for the possibility that they actually guess on at least some variables (McHugh, 2012). While used in medicine, it has been used for example by Ebbeler, et al. (2016) to calculate the inter-rater agreement in data literacy assessments.

A similar measure is Cronbach’s alpha coefficient, which measures how closely related a set of items are as a group, used by Delmas, et al. (2007) in an assessment of statistical literacy. It is a function not only of the difference between items but also of the number of items evaluated. Other measures of reliability include Fleiss kappa, the contingency coefficient, the Pearson r and the Spearman Rho, the intra-class correlation coefficient, the concordance correlation coefficient, and Krippendorff’s alpha (McHugh, 2012).

There are several different types of validity (Linn and Millar, 2005):

- Face validity - do the assessment items appear to be appropriate?
- Content validity - does the assessment content cover what you want to assess?
- Criterion-related validity - how well does the test measure what you want it to?
- Construct validity - are you measuring what you think you're measuring?

There is no single test for validity of data literacy assessments; it “involves amassing evidence that supports these interpretations and decisions” to produce a “validity argument”. However, in other disciplines, quantitative approaches to content validity estimations such as Lawshe’s CVR and Aiken’s V are used, which in turn are based on expert assessments (Ikhsanudin & Subal, 2018).

Teaching Framework

Developing Data Literacy

The development of data literacy in the context of this report is tantamount to the development of individual and organizational data literacy, which consist of knowledge, skills and attitudes, or their analogues, in each of the data literacy competencies, defined as described in the first section, such that the achievement of these competencies can be reliably and validly assessed and detected using the assessment methodologies described in the second section.

Bloom's	Individual	Organizational
Cognitive	Knowledge	Definitions
Psychomotor	Skills / Competencies	Capacities
Affective	Attitudes	Practices

Figure 31 - Bloom's Taxonomy for Individuals and Organizations

Data Literacy Programs

There is not yet an established infrastructure for data literacy development; we mostly find commercial training courses and online resources. So, in the context of data literacy development in a Canadian Forces context, two areas of consideration are important: models and designs for data literacy program development in general, and examples of extant data literacy training programs and curricula.

The development of data literacy in an organization occupies a space between two extremes. On the one hand, we may find data literacy among other types of information and communication competencies, such as digital literacy or information management programs. On the other hand, we might think of data literacy as a first step in the development of higher-level competencies such as data architect or information management.

It need be neither. As Deloitte's Madhura Chakrabarti argues, "Basic data literacy means that your HR professionals have skills such as understanding numbers, reading a dashboard, and knowing that correlation doesn't imply cause and effect. It doesn't mean that they need to know how to run a regression analysis or other advanced statistics programs" (Bersin Insights Team, 2018).

Chakrabarti and colleague Jeff Mike (2018) offer five principles for people analytics upskilling in human resources, though these principles could be applied more broadly to data literacy:

- Adopt a performance mindset
- Start using available resources
- Make analytics education interesting and easily accessible
- Don't lose focus on the core analytics team
- Ensure learning encompasses real business challenges

In other words, it is important to think of data literacy not such as knowledge or content to be used, but rather, as a part of other processes and strategies employed to achieve real objectives or outcomes.

Models and designs for data literacy program development

We did not identify a national or institutional data literacy program development methodology, but some universities have conducted background research and there are *numerous* data literacy program development roadmaps provided by commercial consultants. The sample below is representative.

The Data Information Literacy project funded by the Institute of Museum and Library Services (IMLS) funded research project investigating data information literacy (DIL) consisting of research teams from Purdue

University, the University of Minnesota, the University of Oregon, and Cornell University, which aimed to develop and implement a DIL curriculum in conjunction with university faculty to address these needs, and which produced a book of the same title (Carlson & Johnston, 2015) and a series of case studies. They propose a four-step methodology of planning, development, implementation, and assessment.

For example, QuantHub provides a methodology for developing individual and team data literacy learning and development plans. There are two major components: a series of 'foundational steps' to develop a data literacy vision and roadmap; and an iterative process of assessment, planning, learning and practice (Cowell, 2021).

Dave Wells of Eckerson Group offers a comprehensive data literacy program development methodology (2021). It is in overall structure very similar to QuantHub, in that there is a foundational planning stage and then an iterative process of assessment, planning, execution and measurement. It should be noted that this approach resembles training program methodology generally,

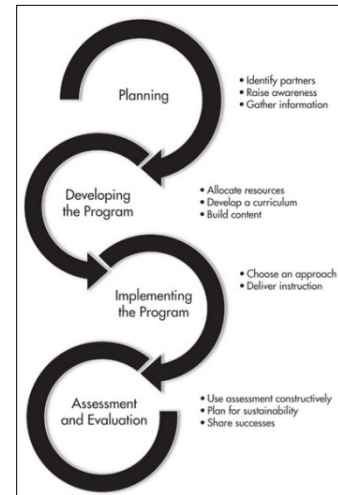


Figure 32 - Learning Development Workflow. Wright, et.al., 2015

	INDIVIDUAL	ORGANIZATION
ASSESS	Individual Literacy Scores Individual Literacy Targets Individual Literacy Gaps	Group/Role Literacy Scores Group/Role Aggregate Targets Group/Role Literacy Gaps
PLAN	Set Knowledge & Skill Goals Identify Learning Resources Develop Learning Plan	Set Capabilities & Cultural Goals Identify Literacy Growth Resources Develop Literacy Growth Plan
EXECUTE	Individual & Group Study Discussion, Coaching, Mentoring Projects & Practice	Incentives & Motivators Organization & Role Changes Process Changes
MEASURE	Study Progress Knowledge & Skills Growth Goals vs Results	Growth of Capabilities Cultural Development Progress Goals vs Results

Annotations: 'periodic reassessment' points to the ASSESS stage; 'continuous planning' points to the PLAN stage.

Figure 33 - Developing Data Literacy. Wells, 2022

and that adaptations specific to data literacy were not needed in the overall framework.

What is key in the Eckerson methodology is that it is recognized that both individual and organizational data literacy competencies and capabilities need to be assessed and planned for. Organizational data literacy is not merely a sum of individual data literacies but requires in addition factors such as tools and systems, incentives and motivators.

Gartner, by contrast, offers a report describing a three-phase methodology for the development of an institutional program (Panetta, 2021) consisting of assessment, data literacy training, and then evaluation of the outcome. While the full program evaluation is an important component, it should be understood once again that organizational *capabilities* constitute more than just training.

Extant Data literacy training programs and curricula

After a brief surge in the mid 2010s, data literacy is enjoying a resurgence in 2022.

While no longer extant, the Data Literacy Project, founded in 2015 at Dalhousie University, proposed “a transdisciplinary examination of existing strategies and best practices for teaching data literacy, synthesizing documented explicit knowledge using a narrative-synthesis methodology and identifying areas where additional research is needed.” (DataLiteracy.ca, Internet Archive, 2021)

This is not to be confused with the Data Literacy Project, launched by Launched by Qlik and including partners Accenture, Cognizant, Experian, Pluralsight, the Chartered Institute of Marketing, and Data to the People. The project has the intention of providing resource-based learning for data literacy and eventually a data literacy certification. This project adopts a competency-based approach similar to that described in this report and makes the point that the competencies “that are relevant to each individual can vary significantly depending on how they interact with data and what part of the data process they are involved in” (Hanegan, 2022).

Conducted online between January and March 2022, the EDUCAUSE Data Literacy Institute consisted of a series of eight synchronous online meetings to discuss resources, activities, and projects in support of seven key data literacy competency areas (Kleitiz & Shelly, 2022).

In the longer term, data literacy initiatives will develop in a manner similar to digital literacy initiatives launched in the previous decade. Now a mature research and development effort, we can draw on the UNESCO assessment that “Six of the national frameworks (Costa Rica, India, Kenya, Philippines, Chile and British Columbia (Canada)) that are most clearly written with regard to the competency areas, as well as the three enterprise frameworks to map against the DigComp 2.0 framework” (Antoninis, 2019).

Teaching and Learning Methods

Pedagogical methods to teach or support data literacy training

Data literacy is new enough that specific pedagogies have not been broadly developed or applied. However, in many ways, data literacy training is similar to that in other disciplines, and especially those characterized as ‘literacies’. Thus, recommendations for, say, digital, information or statistical literacy instruction may apply more broadly to data literacy in general.

The following recommendations are offered by Guidelines for Assessment and Instruction in Statistics Education (GAISE) (GAISE College Report ASA Revision Committee, 2016, Bargagliotti, et al., 2020).

1. Teach statistical thinking.
 - Teach statistics as an investigative process of problem-solving and decision-making.
 - Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Specific trials of different methods in various learning contexts

In this section we survey five specific methods applied to the teaching of data literacy. This is not a comprehensive listing of all methods but serves to illustrate how to apply the principles described just above in specific teaching contexts.

Datastorming

This is a way to think about using how to create designs using data using non-digital media. “To overcome their unfamiliarity to data, we aimed to craft abstract data into hands-on design materials in the form of cards.” (Lim, et al., 2021)

Simulations and Interactive Technologies

Biehler, et al. (2016) describe pre-service teachers' reasoning about modeling a family factory with TinkerPlots, “a data visualization and modeling tool developed for use by middle school through university students.”

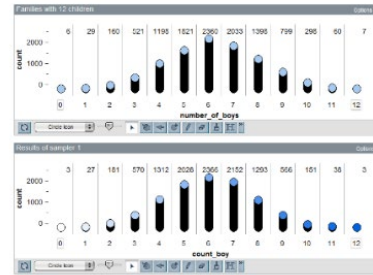


Figure 34 - TinkerPlots

Case-Based Teaching Method

Case-based teaching is “an active learning strategy in which students read and discuss complex, real-life scenarios that call on their analytical thinking skills and decision-making.” (Riddle, et al., 2017).

Utilising affordances in real-world data

Based on the Teaching for Statistical Literacy Hierarchy, this method analyzes statistical literacy lessons that use real-world data from the perspective of the affordances in the data presentation (Chick & Pierce, 2012).

Data-Driven Decision-Making

According to Abbott, et al. (2017), this team-based approach combines a number of competency requirements in a single activity: expertise in data collection, management in a variable environment, allocation of space and time for the process, and the need to ensure process fidelity. This specific activity helps teachers design child literacy instruction, but the approach can be generalized to other data-driven decision-making activities.

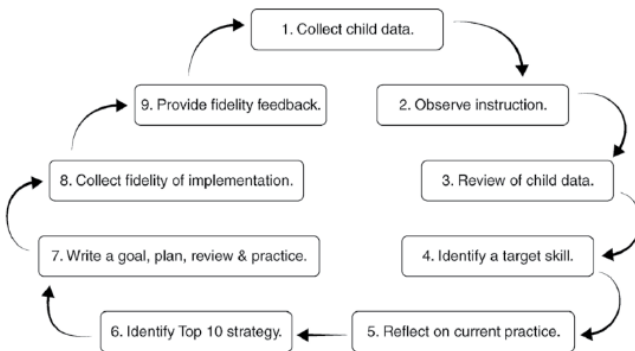


Figure 35 - Data-Driven Decisions. Abbott, et al., 2017, p. 120

Individual Learning Resources

In this section we offer samples of data literacy learning resources. The four types of resources described – lesson plans, help sheets, course libraries, and performance support tools – correspond to four contexts of instruction: in the classroom, in the field, individual or self-study, and while using tools or equipment.

Lessons and Lesson Plans

The Federal Reserve Bank of St. Louis offers a set of clearly written and accessible lesson plans on data literacy. Each lesson reviews data interpretation, analysis, and/or presentation concepts in detail. The source may seem surprising at first but reflects the fact that financial services and insurance agencies have a long history of working with data. A sample lesson is provided as part of this report and may be found here:

<https://www.stlouisfed.org/education/-/media/project/frbstl/stlouisfed/education/lessons/pdf/fred-gdp-stacking.pdf>

Help Sheets and Templates

The Data Visualization Project offers a set of common data visualization formats or templates, with an accompanying instruction page for each one. These templates assist learners consider the varying ways to represent or display a given set of data. <https://datavizproject.com/>

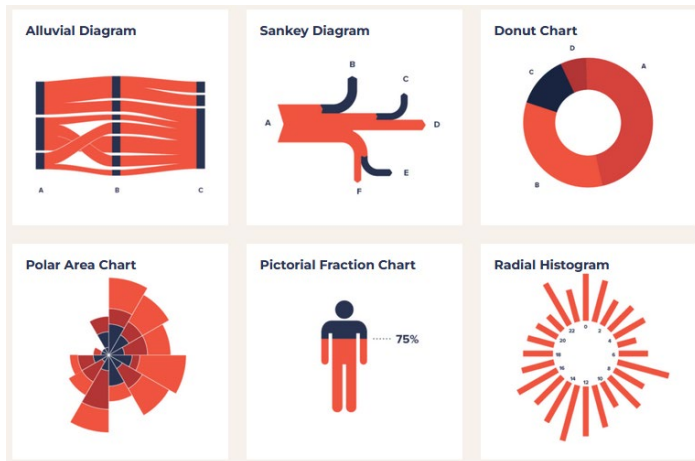


Figure 36 - DataViz Project

Another type of template is the data analysis worksheet created by the Van Andel Education Institute.

Variability questions: Frequency plot (3 kinds) Kind of data: One categorical group and One numeric variable (one axis) Frequency plots show how variable the group is. Describe variability by range, measure of center (mean, median, or mode), and the shape of the distribution.	Dot plot Atlantic Storms - 2013 Max Storm Windspeeds (knots) 2013	Box & whisker plot Atlantic Storms Max Storm Windspeeds (knots) 2013	Histogram Atlantic Storms Max Storm Windspeeds (knots) (2013)
Comparing groups questions: Kind of data: Two or more categorical groups & One numeric variable Frequency plots allow you to compare how variable the groups are. Bar graphs only show a single number (ie. sum, average, percent or count) for each group.	Frequency plots OR Atlantic Storms Max Storm Windspeeds (knots) (To compare two groups of values)	Bar graph Atlantic Storms Max Storm Windspeeds (knots) (To compare two summary values)	Criteria for an informative graph: <input type="checkbox"/> Graph type fits the question <input type="checkbox"/> Axes are drawn & scaled correctly <input type="checkbox"/> Axes are labeled clearly, correctly <input type="checkbox"/> Units are given <input type="checkbox"/> Data are plotted accurately <input type="checkbox"/> Legend is present, if needed <input type="checkbox"/> Graph is overall neat & legible <input type="checkbox"/> Title and/or caption present <input type="checkbox"/> Trend line shown (scatter plot or line graph only) <input type="checkbox"/> Graph helps answer the question
Correlation questions: Kind of data: Two numeric variables Both variables must be continuously numeric. Connect dots only if one variable is linear time (i.e. days, years...) Put time on the X-axis. Show correlation with a "line of best fit".	Scatter plot OR Atlantic Storms Max Windspeeds (knots) vs. Number of Storms	Line graph (for time series) Number of Storms per year Year	

Figure 37 - Van Andel Education Institute, 2018

<https://vaei.vai.org/wp-content/uploads/sites/6/2018/10/Data-Analysis-Strategies.pdf>

Course and Video Libraries

Statistics Canada offers a range of data literacy training products in the form of 20 short-form videos.

<https://www150.statcan.gc.ca/n1/en/catalogue/89200006#wb-auto-2=>

Kubicle offers a subscription-based data literacy and data management course library. Again, the courses are series of videos. <https://kubicle.com/library>

eLearning Curve (2022) offers a data literacy course library in a standard LMS-teaching format. It is part of a larger library of courses on data management, business intelligence and analytics.

Performance Support Tools

Resembling any of the learning resources described above, performance support tools are built into data management applications and help the user progress from basic literacy to advanced user.

For example, Qlik offers a range of performance support learning resources alongside its business intelligence platform. Enterprise-focused data management platforms such as IBM Cognos offer courses and programs that lead to product certification. Progress toward certification is represented through the use of badges to recognize specific skills or competencies.

References

- Abbott, Mary, et al. "A Team Approach to Data-Driven Decision-Making Literacy Instruction in Preschool Classrooms: Child Assessment and Intervention Through Classroom Team Self-Reflection." *Young Exceptional Children*, vol. 20, no. 3, Sept. 2015, pp. 117–32, <https://doi.org/10.1177/1096250615602297>.
- Antoninis, Manos. *Digital Literacy Skills: From a Framework to a Measure*. UNESCO UIS. UNESCO, 6 Mar. 2019, <http://uis.unesco.org/en/blog/digital-literacy-skills-framework-measure>.
- Assuncao, Marcos D., et al. "Big Data Computing and Clouds: Challenges, Solutions, and Future Directions." *Journal of Parallel and Distributed Computing*, vol. 79–80, Dec. 2013, <https://doi.org/10.1016/j.jpdc.2014.08.003>.
- Athanases, Steven Z., et al. "Fostering Data Literacy through Preservice Teacher Inquiry in English Language Arts." *The Teacher Educator*, vol. 48, no. 1, 2013, pp. 8–28.
- Bargagliotti, Anna, et al. *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): A Framework for Statistics and Data Science Education*. American Statistical Association, 2020, https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf.
- Bersin Insights Team. *Insights from IMPACT 2018*. Deloitte Development LLC, 2018, <https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/audit/ca-audit-abm-scotia-insights-from-impact-2018.pdf>.
- Bhargava, Rahul, et al. "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data." *Data-Pop Alliance*, Data-Pop Alliance, Nov. 2015, <https://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data>.
- Biehler, Rolf, et al. "Elementary Preservice Teachers' Reasoning about Modeling a 'Family Factory' with Tinkerplots - A Pilot Study." *Statistics Education Research Journal*, vol. 16, no. 2, Nov. 2017, pp. 244–86, <https://doi.org/10.52041/serj.v16i2.192>.

- Bloom, B. S. "Taxonomy of Educational Objectives", *Handbook: The Cognitive Domain*. David McKay, 1956. [https://www.scirp.org/\(S\(i43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=1506705](https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1506705).
- Bonikowska, Aneta, et al. *Data Literacy: What It Is and How to Measure It in the Public Service*. 2019, <https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2019003-eng.htm>.
- Breakspear, Simon. "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance." *OECD*, Feb. 2012, <https://doi.org/10.1787/5k9fdfqf8-en>.
- Bruce, Christine. *Informed Learning*. Association of College and Research Libraries / American Library Association, 2008, <https://eprints.qut.edu.au/17988/34/17988.pdf>.
- Canada School of Public Service (CSPS). *How Data Literate Are You?* Government of Canada, Apr. 2022, https://catalogue.cspc-efpc.gc.ca/product?catalog=DDN302&cm_locale=en.
- Carlson, Jake, and Lisa R. Johnston. "Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers." Purdue University Press, 2015, <https://www.jstor.org/stable/j.ctt6wq2vh>.
- Chakrabarti, Madhura, and Jeff Mike. *Five Basic Principles for Upskilling HR in People Analytics*. Deloitte Consulting LLP, 2018.
- Chick, Helen L., and Robyn Pierce. "Teaching for Statistical Literacy Utilising Affordances in Real-World Data." *International Journal of Science and Mathematics Education*, vol. 10, no. 2, Apr. 2012, pp. 339–62, <https://doi.org/10.1007/s10763-011-9303-2>.
- Cobb, George W., and David S. Moore. "Mathematics, Statistics, and Teaching." *American Mathematical Monthly*, vol. 104, no. 9, Nov. 1997, pp. 801–23.
- Cowell, Matt. "A Roadmap for Creating a Data Literacy Program - QuantHub." *QuantHub*, QuantHub, 18 June 2020, <https://quanthub.com/data-literacy-program>.
- DataLiteracy. "The Data Literacy Score | Data Literacy." *Data Literacy | Learn the Language of Data*, Feb. 2021, <https://dataliteracy.com/data-literacy-score>.
- . "About This Project - DataLiteracy.Ca." *DataLiteracy.Ca*, Internet Archive, 22 Dec. 2021, <https://web.archive.org/web/20211222131031/http://dataliteracy.ca/about-this-data-literacy-project>.
- Deahl, Erica. *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments*. Massachusetts Institute of Technology, 2007, <https://cmsw.mit.edu/wp/wp-content/uploads/2016/06/233823808-Erica-Deahl-Better-the-Data-You-Know-Developing-Youth-Data-Literacy-in-Schools-and-Informal-Learning-Environments.pdf>.

Delmas, Robert, et al. "Assessing Students' Conceptual Understanding after a First Year Course in Statistics." *Statistics Education Research Journal*, vol. 6, Jan. 2007, [http://iase-web.org/documents/SERJ/SERJ6\(2\)_delMas.pdf](http://iase-web.org/documents/SERJ/SERJ6(2)_delMas.pdf).

Doll, Elizabeth. *Using Data to Foster the School Success of Students with Disabilities*. Institute of Education Sciences, 2011, <https://ies.ed.gov/funding/grantsearch/details.asp?ID=1131>.

DuBois, Jen. "Data Literacy Assessment Is Vital to Data Culture." *QuantHub*, QuantHub, 24 Feb. 2022, <https://quanthub.com/data-literacy-assessment>.

Duncan, Alan D., Donna Medeiros, Aron Clarke, et al. "How to Measure the Value of Data Literacy." *Gartner*, Gartner, Apr. 2022, <https://www.gartner.com/en/documents/4003941>.

---. "Toolkit: Data Literacy Individual Assessment." *Gartner*, Gartner, Apr. 2020, <https://www.gartner.com/en/documents/3983897>.

Ebbeler, Johanna, et al. "Effects of a Data Use Intervention on Educators' Use of Knowledge and Skills." *Studies in Educational Evaluation*, vol. 48, Mar. 2016, pp. 19–31, <https://doi.org/10.1016/j.stueduc.2015.11.002>.

Education Quality and Accountability Office (EQAO). "Ontario Secondary School Literacy Test (OSSLT)." EQAO, Government of Ontario, 2020, <https://www.eqao.com/the-assessments/osslt>.

eLearningCurve. "Request Access to the Data Literacy-Self Assessment". Apr. 2022, <https://ecm.elearningcurve.com/Articles.asp?ID=369>.

Fix, Blair. "The Dunning-Kruger Effect Is Autocorrelation – Economics from the Top Down." *Economics from the Top Down*, 8 Apr. 2022, <https://economicsfromthetopdown.com/2022/04/08/the-dunning-kruger-effect-is-autocorrelation>.

Freed, N., et al. *RFC 6838 - Media Type Specifications and Registration Procedures*. Internet Engineering Task Force (IETF), Jan. 2013, <https://datatracker.ietf.org/doc/html/rfc6838>.

GAISE College Report ASA Revision Committee. *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*. American Statistical Association, 2016, <http://www.amstat.org/education/gaise>.

---. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*. American Statistical Association, 2016, https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf.

Golonka, Ewa M., et al. "Peer Interaction in Text Chat: Qualitative Analysis of Chat Transcripts." *Language Learning & Technology*, vol. 21, no. 2, June 2017, pp. 157–78.

- Hanegan, Kevin. "Establishing a Competency-Based Approach to Data Literacy." *Data Literacy Project*, Qlik, 11 Mar. 2022, <https://thedataliteracyproject.org/posts/establishing-a-competency-based-approach-to-data-literacy>.
- Ikhsanudin, and B. Subali. "Content Validity Analysis of First Semester Formative Test on Biology Subject for Senior High School." *Journal of Physics: Conference Series*, vol. 1097, no. 1, Sept. 2018, p. 012039, <https://doi.org/10.1088/1742-6596/1097/1/012039>.
- Johnson, Beth. "Assessing Written Communication Skills in STEM Courses." *Innovations in Teaching & Learning Conference Proceedings*, vol. 10, Aug. 2018, <https://doi.org/10.13021/G8itlcp.10.2018.2241>.
- Jones, Ben. "A Data Literacy Assessment for Every Employee." *Udemy Business*, Udemy, July 2021, <https://business.udemy.com/resources/data-skills-assessment-template>.
- . "Take the 17 Key Traits of Data Literacy Self-Assessment. DataLiteracy". 2019 <https://dataliteracy.com/take-the-17-key-traits-self-assessment/>
- Kirsch, Irwin, and William Thorn. *Technical Report of the Survey of Adult Skills (PIAAC) (2nd Edition)*. Organisation for Economic Co-operation and Development (OECD), 2016, https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf.
- Kleinheksel, A. J., et al. "Demystifying Content Analysis." *American Journal of Pharmaceutical Education*, vol. 84, no. 1, Jan. 2020, <https://doi.org/10.5688/ajpe7113>.
- Kleitzi, Lauren, and Joe Shelley. *EDUCAUSE Data Literacy Institute*. EDUCAUSE, 24 Jan. 2022, <https://events.educause.edu/educause-institute/data-literacy-institute/2022/online-1>.
- Kolbe, Richard H., and Melissa S. Burnett. "Content-Analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity." *Journal of Consumer Research*, vol. 18, no. 2, Sept. 1991, pp. 243–50, <https://doi.org/10.1086/209256>.
- Koltay, Tibor. "Data Governance, Data Literacy and the Management of Data Quality." *IFLA Journal*, vol. 42, no. 4, Nov. 2016, pp. 303–12, <https://doi.org/10.1177/0340035216672238>.
- Kumar, Vivekanandan, and David Boulanger. "Explainable Automated Essay Scoring: Deep Learning Really Has Pedagogical Value." *Frontiers in Education*, Oct. 2020, <https://doi.org/10.3389/feduc.2020.572367>.
- Linn, Robert L., and David M. Miller. *Measurement and Assessment in Teaching, 9th Edition*. Pearson, 2005, <https://www.pearson.com/us/higher-education/product/Linn-Measurement-and-Assessment-in-Teaching-9th-Edition/9780131137721.html>.
- Lorente, Clara Llebot. *LibGuides: Research Data Services: Data Types & File Formats*. Oregon State University, June 2021, <https://guides.library.oregonstate.edu/research-data-services/data-management-types-formats>.

- Maybee, Clarence, and Lisa Zilinski. "Data Informed Learning: A next Phase Data Literacy Framework for Higher Education." *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, Jan. 2015, pp. 1–4, <https://doi.org/10.1002/pr2.2015.1450520100108>.
- McHugh, Mary L. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica*, vol. 22, no. 3, Oct. 2012, p. 276.
- . "Interrater Reliability: The Kappa Statistic." *Biochemia Medica*, vol. 22, no. 3, Oct. 2012, p. 276.
- Means, Barbara, et al. *Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports*. US Department of Education, Office of Planning, Evaluation and Policy Development, Feb. 2011, <https://eric.ed.gov/?id=ED516494>.
- National Defence. *The Department of National Defence and Canadian Armed Forces Data Strategy - Canada.Ca*. Government of Canada, 3 Dec. 2019, <https://www.canada.ca/en/department-national-defence/corporate/reports-publications/data-strategy.html>.
- Nikou, Melpomeni, and Panagiotis Tziachris. "Prediction and Uncertainty Capabilities of Quantile Regression Forests in Estimating Spatial Distribution of Soil Organic Matter." *ISPRS International Journal of Geo-Information*, vol. 11, no. 2, Feb. 2022, p. 130, <https://doi.org/10.3390/ijgi11020130>.
- Oberländer, Maren, et al. "Digital Competencies: A Review of the Literature and Applications in the Workplace." *Computers & Education*, vol. 146, Mar. 2020, p. 103752, <https://doi.org/10.1016/j.compedu.2019.103752>.
- Panetta, Kasey. "A Data Literacy Guide For D&A Leaders." *Gartner*, Gartner, 26 Aug. 2021, <https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy>.
- Piro, Jody S., and Cynthia J. Hutchinson. "Using a Data Chat to Teach Instructional Interventions: Student Perceptions of Data Literacy in an Assessment Course." *New Educator*, vol. 10, no. 2, 2014, pp. 95–111.
- Riddle, Derek R., et al. "Making a Case for Case-Based Teaching in Data Literacy." *Kappa Delta Pi Record*, vol. 53, no. 3, July 2017, pp. 131–33, <https://doi.org/10.1080/00228958.2017.1334479>.
- Ridsdale, Chantel, et al. *Strategies and Best Practices for Data Literacy Education*. Dalhousie University, 2015, <https://dalspace.library.dal.ca/bitstream/handle/10222/64578/Strategies%20and%20Best%20Practices%20for%20Data%20Literacy%20Education.pdf?sequence=1&isAllowed=y>.
- Robinson, Clinton. *Aspects of Literacy Assessment: Topics and Issues from the UNESCO Expert Meeting*. 2005, <https://unesdoc.unesco.org/ark:/48223/pf0000140125>.
- Salleh, Kahirol Mohd, et al. "Competency of Adult Learners in Learning: Application of the Iceberg Competency Model." *Procedia - Social and Behavioral Sciences*, vol. 204, Aug. 2015, pp. 326–34, <https://doi.org/10.1016/j.sbspro.2015.08.160>.

- Samuel, A. L. "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development*, vol. 3, no. 3, July 1959, pp. 210–29, <https://doi.org/10.1147/rd.33.0210>.
- Schild, Milo. "Information Literacy, Statistical Literacy and Data Literacy." *IASSIST Quarterly*, 2004, <http://www.statlit.org/pdf/2004-Schild-IASSIST.pdf>.
- Sickler, Jessica, et al. "Measuring Data Skills in Undergraduate Student Work." *Journal of College Science Teaching*, vol. 50, no. 4, Apr. 2022, <https://www.nsta.org/journal-college-science-teaching/journal-college-science-teaching-marchapril-2021/measuring-data>.
- Sikorski, Jonathon David. *Examination of the NU Data Knowledge Scale*. University of Nebraska - Lincoln, 2016, <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1268&context=cehsdiss>.
- Sperry, J. *Data Literacy: Exploring Economic Data*. U.S. Census Bureau, 2018.
- St. Mary's. *St. Mary's Practice Test 2017/2018*. 2018, https://drive.google.com/file/d/1LsDbvAy_YHepMVD9VVEEWH6w5U_Dk1-E/view.
- Statistics Canada. *The Daily - Begin Your Data Journey with Data Literacy Training Videos*. Government of Canada, 23 Sept. 2020, <https://www150.statcan.gc.ca/n1/daily-quotidien/200923/dq200923f-eng.htm>.
- Suarez-Alvarez, Javier. "Are 15-Year-Olds Prepared to Deal with Fake News and Misinformation?" *OECD*, May 2021, <https://doi.org/10.1787/6ad5395e-en>.
- Suryadi, et al. "Data Literacy of High School Students on Physics Learning." *Journal of Physics: Conference Series*, vol. 1839, no. 1, Mar. 2021, p. 012025, <https://doi.org/10.1088/1742-6596/1839/1/012025>.
- United Nations Industrial Development Organization (UNIDO). *UNIDO. Competencies. Strengthening Organizational Core Values and Managerial Capabilities - PDF Free Download*. United Nations, 2002, <https://docplayer.net/9459584-Unido-competencies-strengthening-organizational-core-values-and-managerial-capabilities.html>.
- Vahey, Phil, et al. *Mathematizing Middle School: Results From a Cross-Disciplinary Study of Data Literacy*. 5AD, https://archive.sri.com/sites/default/files/publications/imports/TWD_AERA_Mathematizing_Middle_School.pdf.
- Van Andel Education Institute, "Data Analysis", 2018. <https://vaei.vai.org/wp-content/uploads/sites/6/2018/10/Data-Analysis-Strategies.pdf>
- Visengeriyeva, Larysa, et al. *An Overview of the End-to-End Machine Learning Workflow*. MLOps.org, Dec. 2021, <https://ml-ops.org/content/end-to-end-ml-workflow>.

Walsh, John. *Implementing DND/CAF Data Strategy*. Government of Canada, 26 Sept. 2021, <https://publicsectornetwork.co/wp-content/uploads/2021/09/John-Walsh-PDF.pdf>.

Wells, Dave. *Building a Data Literacy Program*. 4 Jan. 2021, <https://www.eckerson.com/articles/the-data-literacy-imperative-part-i-building-a-data-literacy-program>.

Williams, Paula G., et al. "On the Validity of Self-Report Assessment of Cognitive Abilities: Attentional Control Scale Associations With Cognitive Performance, Emotional Adjustment, and Personality." *Psychological Assessment*, vol. 29, no. 5, 2017, pp. 519–30.

Wolff, Annika, et al. "Creating an Understanding of Data Literacy for a Data-Driven Society." *Journal of Community Informatics*, vol. 12, no. 3, Aug. 2016, <https://doi.org/10.15353/joci.v12i3.3275>.

Yi Min Lim, Delia, et al. "Datastorming: Crafting Data into Design Materials for Design Students' Creative Data Literacy." *C&C '21: Creativity and Cognition*, Association for Computing Machinery, 2021, pp. 1–9, <https://doi.org/10.1145/3450741.3465246>.

Appendices

Appendix 1 - Data Literacy Frameworks for Individuals

1. Draft Canada Data Competency Frameworks

Source: Draft Canada Data Competency Frameworks, unpublished, 2021

Competency Framework V3

Data Concepts and Culture			
Data Awareness	Data Culture and Organisational Awareness	Data Ethics	Evidence-Informed Decision-Making

Data Governance, Collection and Stewardship			
Data Governance and Standards	Access, Security and Interoperability	Data Quality, Value and Trust	Methods, Processes and Tools

Analytics and Evaluation						
Asking Questions and Problem Framing	Data Analysis	Advanced Analytics	Storytelling with Data	Data Visualisation	Evaluating Outcomes	Tools

Data Systems and Architecture		
Enterprise Solution Architecture	Data Systems	Methods, Processes and Tools

Level 1 – The Curious	Level 2 – The Competent	Level 3 – The Coach
Has core or foundational knowledge and seeks additional learning Can complete simple tasks on their own or with instruction as needed	Fully grasps foundational knowledge and has a deeper or more advanced understanding of key competencies Can confidently complete defined tasks independently and support work on complex problems	Can teach others and leads by example Can identify, analyse and resolve complex problems independently and support, lead and coach others in their work

2. JNCC Report No. 590, gov.uk Data Skills Framework

2. Source: Heeley, L. & Wilkinson, S., (2018), Data Skills Framework: A generic approach to assessing and developing data related competencies and skills, JNCC Report No. 590, gov.uk <https://hub.jncc.gov.uk/assets/6ef949ed-1c22-438f-b2f7-fd4818ec4566>

1. **Requirements and business analysis** – the ability to understand and prioritise user needs; and identify how data can be efficiently integrated into processes.
2. **Data governance** – the responsibilities associated with collection, handling, ownership, publication and ultimately removal of data.
3. **Data management** – knowledge of data concepts, including quality control, storage, and integration with other sources.
4. **Access and security** – understanding of: the obligations and restrictions around granting and gaining access to data; the principles of open data; different licensing models; security; and the process of risk assessments
5. **Data manipulation** – manipulating, processing, cleansing and combining data for further analysis or use. Automation of complex manipulation on large data volumes
6. **Analysis and modelling** – analysis of data and data modelling. Includes statistical knowledge, processing chains, machine learning and predictive analytics
7. **Communication and visualisation** – interpreting, summarising and communicating data and various analytical outputs for different audiences.

3. Statistics Canada, 2020. Data literacy competencies

Source: Statistics Canada, 2020. Data literacy competencies <https://www.statcan.gc.ca/en/wtc/data-literacy/competencies>

Data analysis - The knowledge and skills required to ask and answer a range of questions by analyzing data including developing an analytical plan; selecting and using appropriate statistical techniques and tools; and interpreting, evaluating and comparing results with other findings.

Data awareness - The knowledge required to know what data is and what are different types of data. This includes understanding the use of data concepts and definitions.

Data cleaning - The knowledge and skills to determine if data are 'clean' and use the best method and tools to take necessary actions to resolve any problems to ensure data are in a suitable form for analysis.

Data discovery - The knowledge and skills to search, identify, locate and access data from a range of sources related to the needs of an organization.

Data ethics - The knowledge that allows a person to acquire, use, interpret and share data in an ethical manner including recognizing legal and ethical issues (e.g., biases, privacy).

Data exploration - The knowledge and skills required to use a range of methods and tools to learn what is in the data. The methods include: summary statistics; frequency tables; outlier detection; and visualization to explore patterns and relationships in the data.

Data gathering - The knowledge and skills to gather data in simple and more complex forms to support the gatherer's needs. This could involve the planning, development and execution of surveys or gathering data from other sources such as administrative data, satellite or social media data.

Data interpretation - The knowledge and skills required to read and understand tables, charts and graphs and identify points of interest. Interpretation of data also involves synthesizing information from related sources.

Data management and organization - The knowledge and skills required to navigate internal and external systems to locate, access, organize, protect and store data related to the organization's needs.

Data modeling - The knowledge and skills required to apply advanced statistical and analytic techniques and tools (e.g. regression, machine learning, data mining) to perform data exploration and build accurate, valid and efficient modelling solutions that can be used to find relationships between data and make predictions about data.

Data stewardship - Knowledge and skills required to effectively manage data assets. This includes the oversight of data to ensure fitness for use, the accessibility of the data, and compliance with policies, directives and regulations.

Data tools - The knowledge and skills required to use appropriate software, tools, and processes to gather, organize, analyze, visualize and manage data.

Data visualization - The knowledge and skills required to create meaningful tables, charts and graphics to visually present data. This also includes evaluating the effectiveness of the visual representation (i.e., using the right chart) while ensuring accuracy to avoid misrepresentation.

Evaluating data quality - The knowledge and skills required to critically assess data sources to ensure they meet the needs of an organization. This includes identifying errors or problems and taking action to correct them. This also includes awareness of organizational policies, procedures and standards to ensure good quality data.

Evaluating decisions based on data - The knowledge and skills required to evaluate a range of data sources and evidence in order to make decisions and take actions. This can include monitoring and evaluating the effectiveness of policies and programs.

Evidence based decision-making - The knowledge and skills required to use data to help in the decision-making and policy making process. This includes thinking critically when working with data; formulating appropriate business questions; identifying appropriate datasets; deciding on measurement priorities; prioritizing information garnered from data; converting data into actionable information; and weighing the merit and impact of possible solutions and decisions.

Metadata creation and use - The knowledge and skills required to extract and create meaningful documentation that will enable the correct usage and interpretation of the data. This includes the documentation of metadata which is the underlying definitions and descriptions about the data.

Storytelling - The knowledge and skills required to describe key points of interest in statistical information (i.e., data that has been analyzed). This includes identifying the desired outcome of the presentation; identifying the audience's needs and level of familiarity with the subject; establishing the context; and selecting effective visualizations.

4. 2021 Data Maturity Model

Source: Palmer, K. 2021. Building a Data Maturity Model + the 4 Stages of Data Maturity. Safegraph. <https://www.safegraph.com/blog/the-four-stages-of-data-maturity>

2021 Data Maturity Model				
	Explorer	User	Leader	Innovator
Business Strategy	Data is used solely for reporting purposes	Data insights are used to inform business decisions	Competitive business strategy is built off of data	Data informs a continuous evolution of business strategy
Data	The organization solely uses its own internal data	The organization uses data providers to enrich & supplement its own data	Third party data is used as a differentiator	The organization is constantly looking to leverage new datasets from non-obvious sources
Culture	The use of data and analysis is left up to the individual	Data is a part of measuring results but not planning	Decision makers are enabled with the results of data analysis to maximize business outcomes	The organization has built true AI/ML algorithms that adapt and improve business objectives
Architecture	The business lacks a cohesive and coherent data architecture	Some architecture exists to automate and analyze data flow	The architecture enables all members of the organization to be data driven	The architecture is built for speed, market distribution and large volumes of data
Data Governance	Governance is largely manual and lacks consistency	Process is in place to protect data quality across the organization	There's universal confidence in the data and resulting insights	Data governance is integrated into all business processes
Procurement & Onboarding	No datasets have been onboarded to the organization	Individual teams are responsible for procuring and onboarding their own data	There's a streamlined process for data procurement, but onboarding isn't universal	The organization has a data procurement team that sources and onboards new data

5. UNESCO Institute for Statistics

Source: Nancy Law, et al. (2018) A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2. UNESCO Institute for Statistics. <http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf>

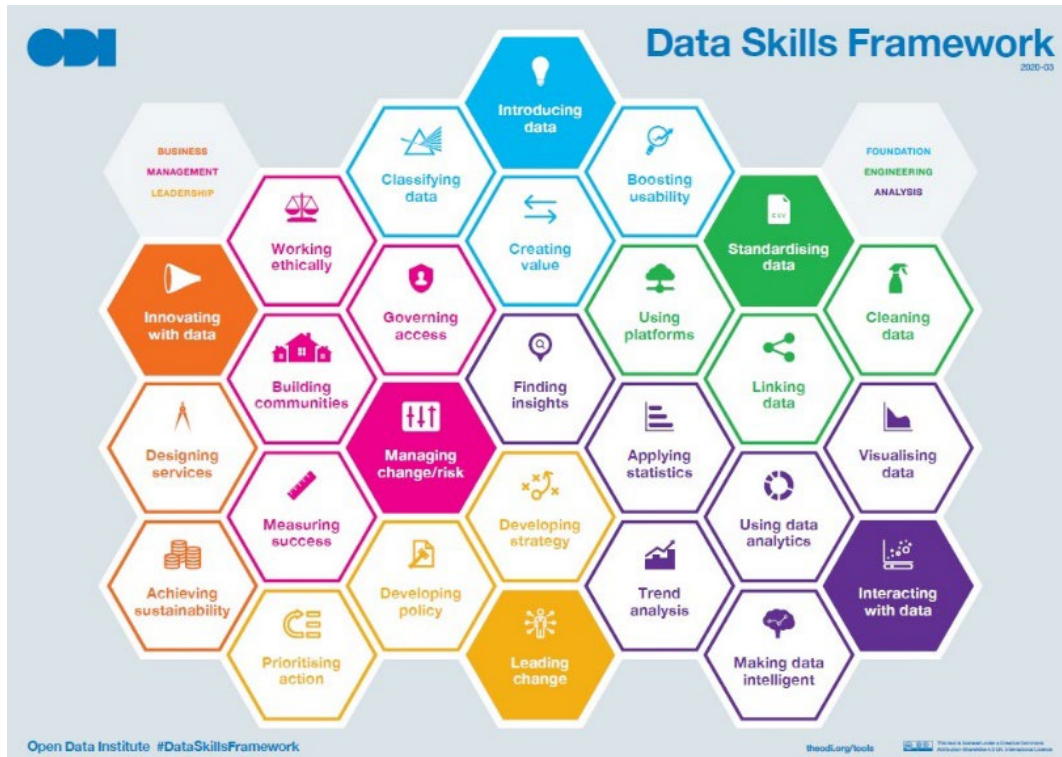
Table 1. DigComp 2.0 competence areas and competences

Competence area	Competences
1. Information and data literacy	1.1 Browsing, searching and filtering data, information and digital content 1.2 Evaluating data, information and digital content 1.3 Managing data, information and digital content

6. Open Data Institute Data Skills Framework

Source: Open Data Institute. (2020) Data Skills Framework.

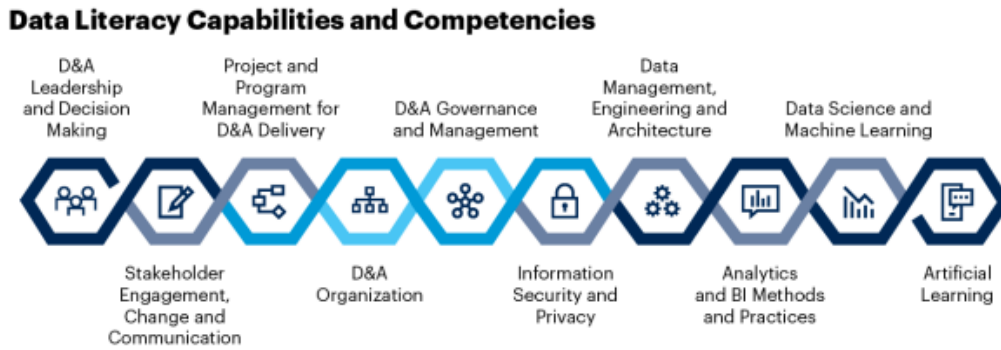
https://www.lrsd.net/SiteGovernImages/LRSD/News/Why_Data_Literacy_2a6ce980_8d6e_4bd2_9c14_428dd0b4f273_cmp/Data-Skills-Framework.jpg



7. Duncan, et al., Data Literacy Capabilities and Competencies

Source: Alan D. Duncan, Donna Medeiros, Aron Clarke, Sally Parker. (2021) How to Measure the Value of Data Literacy. Gartner. <https://www.gartner.com/en/documents/4003941-how-to-measure-the-value-of-data-literacy>

Figure 4. Data Literacy Skills and Competencies



Source: Gartner
750950_C



8. Gartner Data Literacy Toolkit

Source: Gartner. (2019). Toolkit: Data Literacy Individual Assessment.

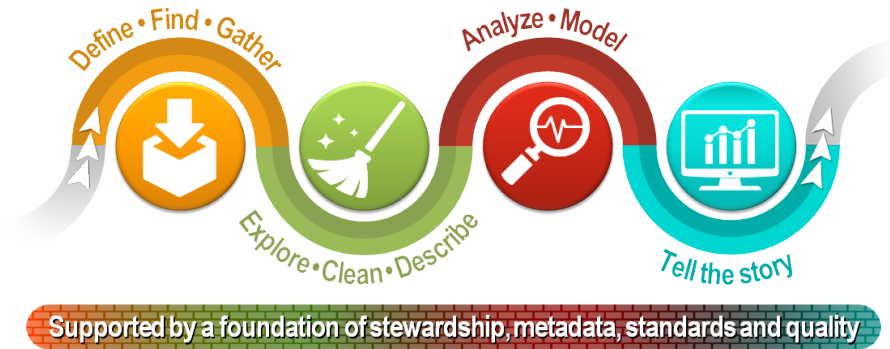
<https://www.gartner.com/en/documents/3983897/toolkit-data-literacy-individual-assessment>

Assessment Questions	Response	T
1. I have identified data literacy as a core skill set and can define its business value for my organization in the modern workplace.	0 1 2 3 4	
2. I can "speak data" and cite examples of data-driven decisions within my organization.	0 1 2 3 4	
3. I understand the business value of data scientists, data engineers and business analysts and the importance of meeting them frequently and productively.	0 1 2 3 4	
4. I can identify myself as a specialist (in data management, data science, information governance and my organization's business domains) or generalist (who can translate and work across the specialties within my organization).	0 1 2 3 4	
5. I am empowered by my HR department to develop specific skills related to data literacy and can articulate its contribution to my personal development.	0 1 2 3 4	
6. I understand how data adds value to business decisions, participate in formally measuring the value of data and analytics investments, and can cite examples of measurable outcomes powered by data and analytics.	0 1 2 3 4	
7. I have confidence in our data management professionals to have strong business acumen and their ability to articulate our strategy, business process areas, key metrics and business analytics.	0 1 2 3 4	
8. I regularly participate in meetings where we commonly share and discuss data, metrics, analytics and the decisions they support, and processes and outcomes they improve.	0 1 2 3 4	
9. In my role I am empowered to innovate with data, design new data-enabled products and processes, and explore new business models, including monetisation (e.g., selling of data).	0 1 2 3 4	
10. I can describe how my organization fits within a business ecosystem, naming examples of our partners, customers and providers.	0 1 2 3 4	
11. I understand that information is a strategic asset, and can each explain three examples of how it is, or is not, treated or accounted for as such.	0 1 2 3 4	
12. I am able to have a conversation with our data & analytics professionals and understand the differences of a data warehouse, data mart, data lake and data hub.	0 1 2 3 4	
13. I have a good understanding of data quality, master data management, application data management, information governance/stewardship and metadata management principles, and can explain the basic value of each with examples in terms that matter.	0 1 2 3 4	
14. I can name five data sources (either external or internal) that are relevant to our business now but were not prevalent 10 years ago.	0 1 2 3 4	
15. I regularly leverage data-discovery capabilities and tools to accelerate exploration, ingestion and management of new data.	0 1 2 3 4	
16. I understand modern technologies such as augmented analytics, can explain the difference between predictive and prescriptive analytics, and can give an example of each.	0 1 2 3 4	
17. I commonly use data visualisation and storytelling techniques and can confidently stand up and tell a story with data and visualisation.	0 1 2 3 4	
18. I can explain the following terms confidently: mean, median, mode and standard deviation.	0 1 2 3 4	
19. I understand how natural-language processing (NLP) and natural-language generation (NLG) are applied, and can describe use cases of each technique.	0 1 2 3 4	
20. I understand machine learning (ML) and artificial intelligence (AI) capabilities, and can each describe 3 use cases of ML/AI relevant to my organization.	0 1 2 3 4	
21. I understand change is required in the modern workplace and how I need to change to embrace the concept of data literacy.	0 1 2 3 4	
22. In addition to classroom based training, I am active in helping others learn and develop e.g., through social learning, on the job coaching, mentoring etc.	0 1 2 3 4	
23. In addition to classroom based training, my colleagues are active in helping me and others learn and develop e.g., through social learning, on the job coaching, mentoring etc.	0 1 2 3 4	
24. I feel I actively promote and encourage social learning, on the job coaching and mentoring within my organization.	0 1 2 3 4	
25. I am given freedom to be experimental in my ways of working and empowered to learn from mistakes.	0 1 2 3 4	

9. Apolitical Data Literacy Boot Camp

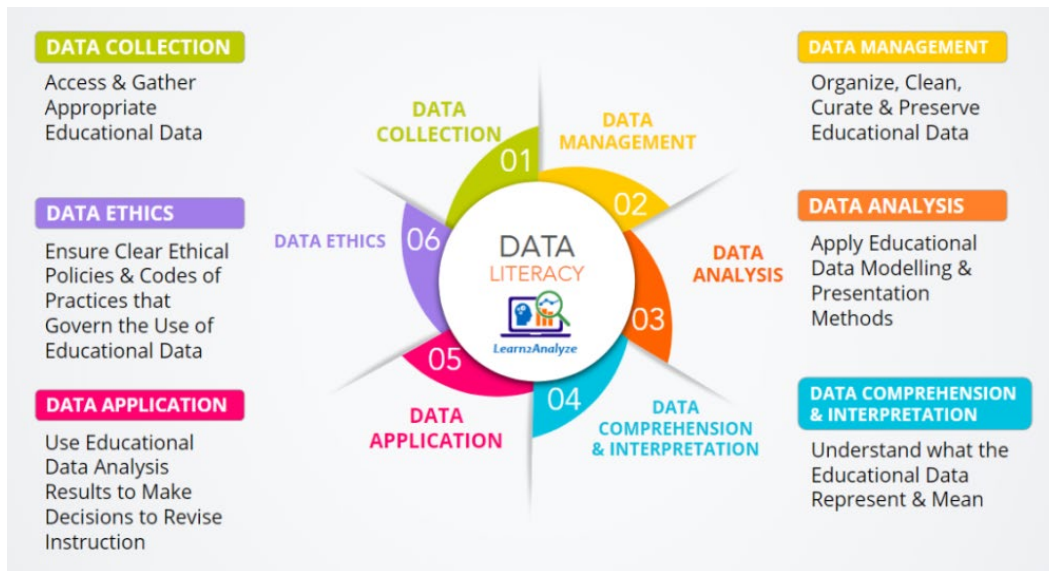
Source: Apolitical. (2021). Apolitical Data Literacy Boot Camp With Statistics Canada.

<https://apolitical.co/events/data-literacy-an-online-boot-camp-with-statistics-canada>



10. Learn2Analyze Educational Data Literacy Competence Profile

Source: Learn2Analyze. (2017) The Learn2Analyze Educational Data Literacy Competence Profile. European Commission / Erasmus+ funded project. <https://learn2analyze.eu/proj/l2a-edl-cp/>



11. Dalhousie University Data Literacy Education

Source: Chantel Ridsdale, et al. (2015). Strategies and Best Practices for Data Literacy Education. Dalhousie University.

<https://dalspace.library.dal.ca/bitstream/handle/10222/64578/Strategies%20and%20Best%20Practices%20for%20Data%20Literacy%20Education.pdf?sequence=1&isAllowed=y> Also referenced in: Bonikowska, Sanmartin, and Frenette, Statistics Canada 2019

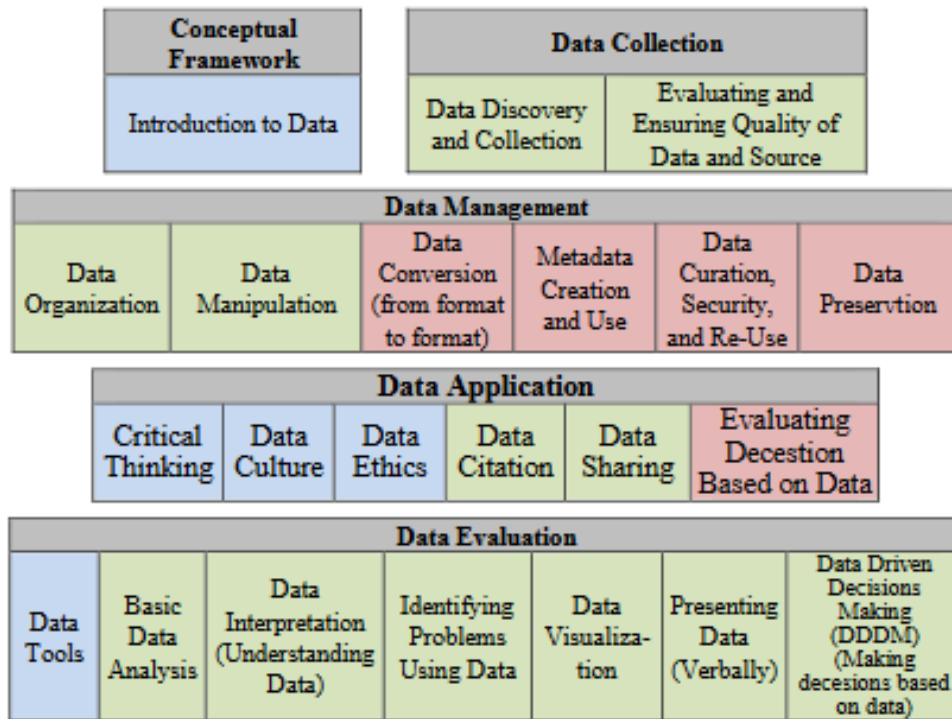
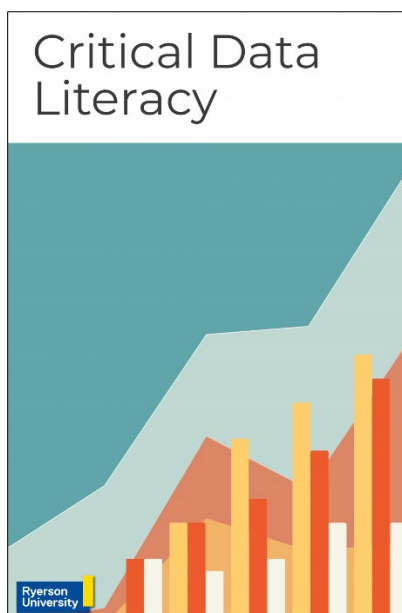


Image via https://www.researchgate.net/publication/341968247_The_profile_of_students%27_data_literacy_at_SMA_Negeri_1_Karanganyar/fulltext/5ee559ada6fdcc73be7b7701/The-profile-of-students-data-literacy-at-SMA-Negeri-1-Karanganyar.pdf?origin=publication_detail

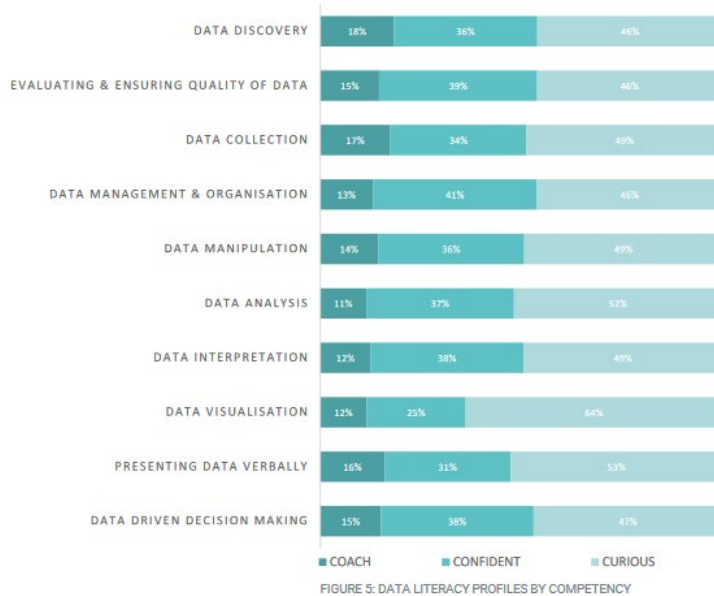
12. Critical Data Literacy

Source: Nora Mulvaney, Audrey Wubbenhorst, Amtoj Kaur (2022). Critical Data Literacy. Ryerson University. <https://pressbooks.library.ryerson.ca/criticaldataliteracy/>



13. Databilities

Source: Data to the People. 2019. Databilities <https://www.datatothepeople.org/databilities> See also: https://www.datatothepeople.org/files/ugd/1ff4ae_3e3b4a50ead44939b50713c3c9af4b66.pdf . Also referenced in: Bonikowska, Sanmartin, and Frenette, Statistics Canada 2019



14. Wolff, et al.

14. Annika Wolff, et al. (2016). Creating an understanding of data literacy for a data-driven society. <https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3275> Also referenced in: Bonikowska, Sanmartin, and Frenette, Statistics Canada 2019

Table 1. categorization of data literacy skills across multiple definitions

PPDA C	Competence	Foundational competence
<i>Inquiry process</i>		
	Plan, implement and monitor courses of action [14]	
	Undertake data inquiry process [34]	
<i>Foundational knowledge</i>		
	Understand the ethics of using data [34]	Ethics
	Use data to solve (real) problems [29] [2] [34] [14] [7]	Real-world problem-solving context
	Understand the role and impact of data in society in different contexts [7]	
P	Identify problems or questions that can be solved with data [14] [7]	Ask questions from data
P	Develop hypotheses [14] Identify data [14]	Develop hypotheses and identify potential sources of data
D	Collect or acquire data [6] [34] [21] [14] [7] <i>Critique data [21][7]</i>	Collect or acquire data

A	Transform data into information and ultimately actionable knowledge [29] [34] [14] [7] Create explanations from data [29] [33] [14] [7] [2] [4] [6] [21] <i>Access data [22]; Analyse data [2] [34] [21] [14] Understand data types [7]; Convert data [22]; Prepare data for analysis [7]; Combine quantitative and qualitative data [34] [7]; Use appropriate tools [29]; Work with large data sets [6]; Summarise data [14]</i>	Analyse and create explanations from data
C	Interpret information derived from datasets [29] [4] [34] [4][22][3][30] Critique presented interpretations of data [29] [4] [3][30]	Evaluate the validity of explanations based on data and formulate new questions

15. Maturity Model for Data Literacy

15. Helena Sternkopf, Roland M. Mueller. (2018). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. University of Hawai'i at Manoa, Proceedings of the 51st Hawaii International Conference on System Sciences. <https://scholarspace.manoa.hawaii.edu/handle/10125/50519> Also referenced in: Bonikowska, Sanmartin, and Frenette, Statistics Canada 2019



Figure 3. Results of the DLMM self-evaluation

16. Grillenberger & Romeike, 2018

16. Andreas Grillenberger and Ralf Romeike. (2018). Developing a theoretically founded data literacy competency model. WiPSCE '18: Proceedings of the 13th Workshop in Primary and Secondary Computing Education. <https://dl.acm.org/doi/10.1145/3265757.3265766> Also referenced in: Bonikowska, Sanmartin, and Frenette, Statistics Canada 2019

Table 1: Matrix of exemplary competencies for the different combinations of process (P1–P4) and content areas (C1–C4).

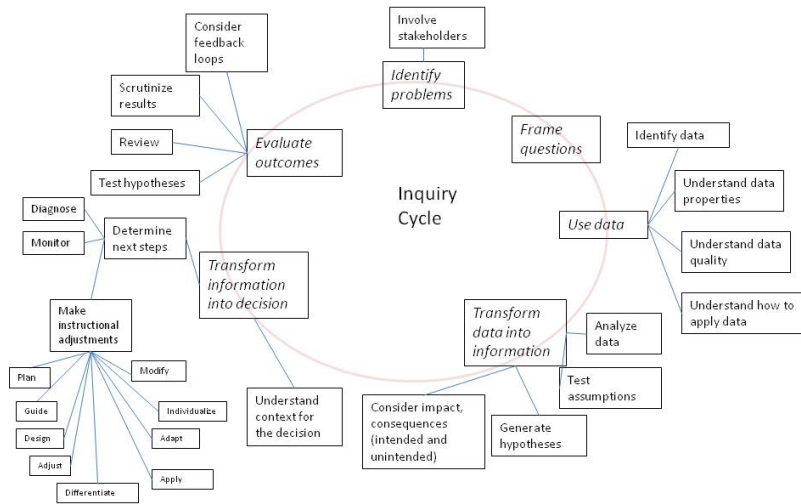
	P1 gathering, modeling and cleansing	P2 implementing and optimizing	P3 analyzing, visualizing and interpreting	P4 sharing, archiving and erasing
C1 data and information	<ul style="list-style-type: none"> - choose suitable sensors for gathering the desired information as data - structure the gathered data in a suitable way for later analysis - evaluate if the captured data represents the original information correctly 	<ul style="list-style-type: none"> - implement algorithms for gathering the desired data - implement simple algorithms to download data from web APIs - discuss optimizations and limits of data gathering 	<ul style="list-style-type: none"> - combine data to gain new information - emphasize the desired information in visualizations - interpret data and analysis results to get new information 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - select a suitable data model - structure the gathered data in a suitable way for storage - visualize data models in a suitable way 	<ul style="list-style-type: none"> - decide on a suitable data storage and store the data - use possibilities for enabling efficient access to data - increase storage efficiency using compression 	<ul style="list-style-type: none"> - access the data in a suitable way for analysis - use suitable data formats for the data to analyze - store their analysis results appropriately 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - decide whether specific data influences results of analysis - structure data appropriately for analysis - connect data from different sources for analysis purposes 	<ul style="list-style-type: none"> - implement simple analysis algorithms - determine adjustment screws for analysis - optimize data analyses in order to gain higher quality results 	<ul style="list-style-type: none"> - decide for appropriate analysis methods - visualize data and analysis results - interpret the results of analyses 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - discuss the ethical impacts of the conducted data analyses and their results - decide whether analysis results are sufficiently anonymized - reflect whether analyzing specific data raises privacy issues 	<ul style="list-style-type: none"> - 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

17. Conceptual Framework for Data Literacy for Teachers (DLFT)

Source: Edith Gummer & Ellen B. Mandinach. (2015). Building a Conceptual Framework for Data Literacy. Teachers College Record. <https://www.tcrecord.org/content.asp?contentid=17856>

Also: Ellen B.Mandinach & Edith S.Gummer (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions.

<https://www.sciencedirect.com/science/article/pii/S0742051X16301391#bib14>










18. APS Data Capability Framework

Source: . (2021). APS Data Capability Framework. Government of Australia. <https://www.apsc.gov.au/publication/aps-data-capability-framework>



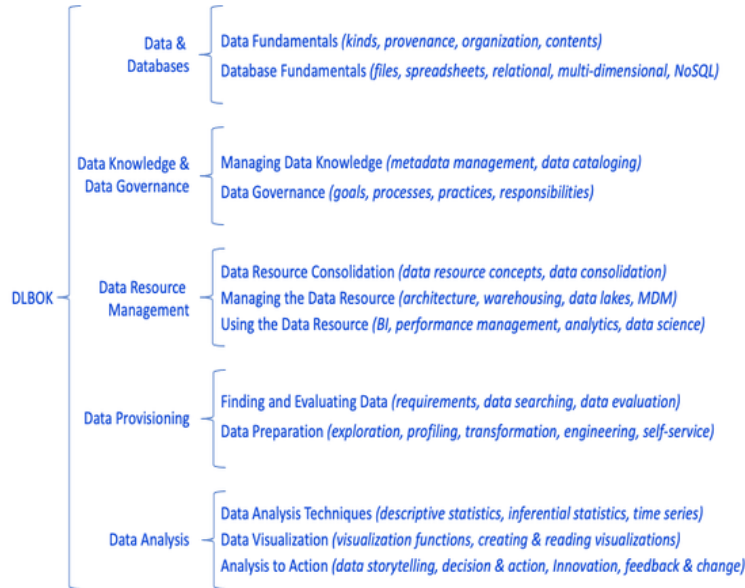
Categories

-  **Plan** - The processes and resources are mapped out for the lifecycle of the data. The project's goals are stated, and a full data management plan is created.
-  **Collect** - Data is gathered or generated by the individuals/organisation wanting to use it.
-  **Describe** - The data is accurately described using the appropriate metadata standards.
-  **Store** - The data is stored in a digital repository, is made secure and reusable. This often very quickly follows collection.
-  **Prepare** - The data is prepared, made ready for analysis and use.
-  **Analyse & Use** - The data is analysed and used for the purpose for which it was collected or generated and reused for additional value.
-  **Save / Destroy** - Actions are taken to safeguard the long-term viability and availability of the data.

19. Data Literacy Body of Knowledge (DLBOK)

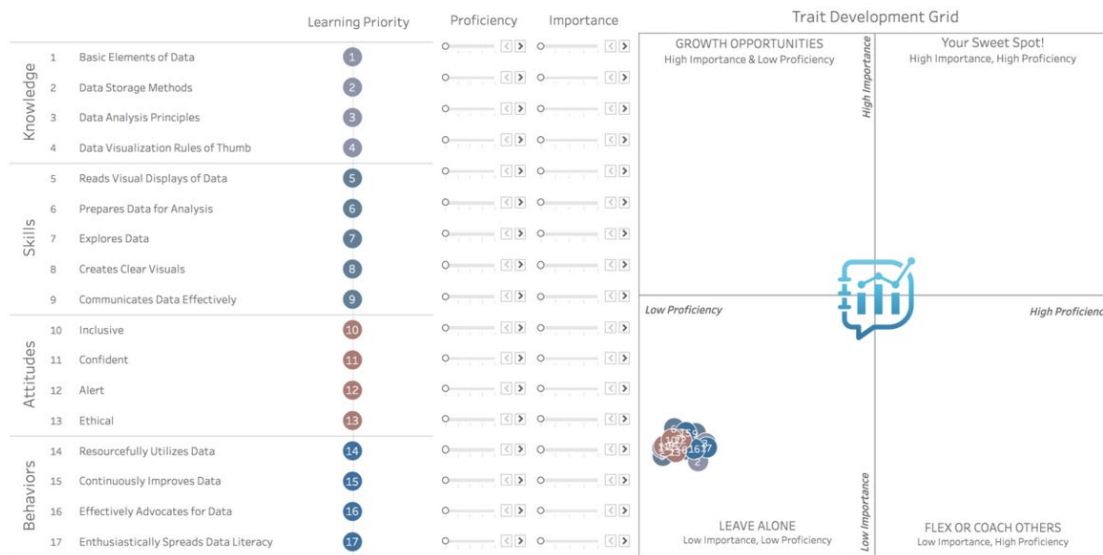
Source: Dave Wells. (2022). The Data Literacy Imperative - Part II: The Data Literacy Body of Knowledge. Eckerson Group. <https://www.eckerson.com/articles/the-data-literacy-imperative-part-ii-the-data-literacy-body-of-knowledge>

Figure 1. The Data Literacy Body of Knowledge (DLBOK)



20. DataLiteracy 17 Key Traits of Data Literacy

Source: Ben Jones (2019). Take the 17 Key Traits of Data Literacy Self-Assessment. DataLiteracy. <https://dataliteracy.com/take-the-17-key-traits-self-assessment/>



Appendix 2 - Corporate Data Literacy Frameworks

1. TDWI Data Literacy Maturity Model Assessment Guide

Source: Fern Halper & Chris Adamson. (2021). TDWI Data Literacy Maturity Model Assessment Guide. TDWI. <https://tdwi.org/-/media/D98D941F36214D0FA9501FCF3B5F3A79.pdf>

Culture and Resources	Data Infrastructure	Skills and Talent	Tooling	Data Governance and Processes
<ul style="list-style-type: none"> Leadership and commitment Culture and adoption Funding 	<ul style="list-style-type: none"> Data foundations Data architecture Data access 	<ul style="list-style-type: none"> Assessment initiatives Learning programs Individual skills Expertise of supporting data and analytics teams 	<ul style="list-style-type: none"> Scope Proficiency Role alignment 	<ul style="list-style-type: none"> Data governance foundations Governance roles Governance ownership Security and privacy Audit

Figure 1. TDWI Data Literacy Maturity Model dimensions.

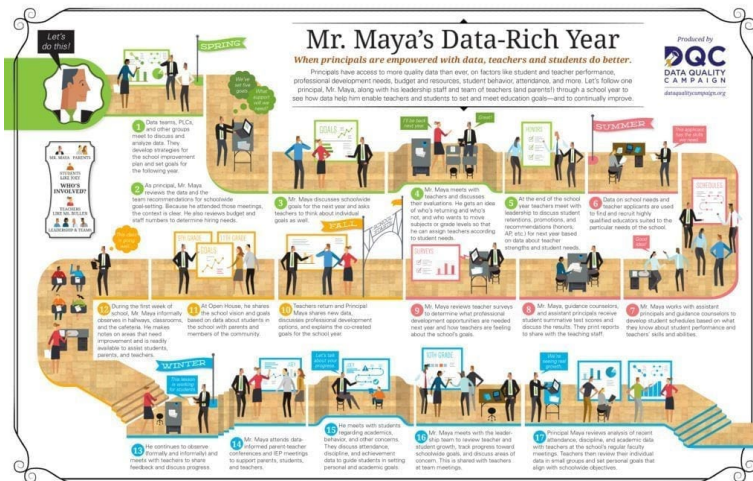
2. The Data Literacy Index

Source: The Data Literacy Project. (2018). The Data Literacy Index.

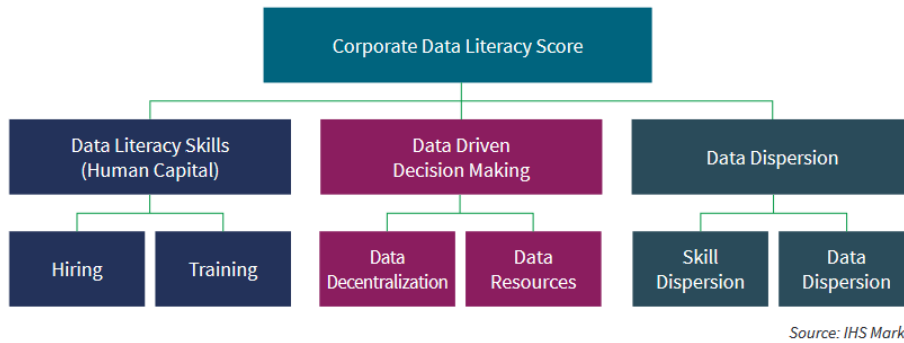
<https://thedataliteracyproject.org/files/download/downloads/qlik%20-%20The%20Data%20Literacy%20Index%20October%202018.pdf>

3. Data Quality Campaign

Source: Data Quality Campaign. (2019). Data Literacy 101. <https://dataqualitycampaign.org/resource/data-literacy-101/>



THE PILLARS OF CORPORATE DATA LITERACY



Appendix 3 – Data Literacy Models

Data Stewardship Model

This model describes approaches to data literacy that emphasize data acquisition, curation, quality and deployment. A prototypical example of this approach is the Statistics Canada descriptions of data quality and the data journey.

Statistics Canada, 2020-09-23: “Data literacy also means having the knowledge and skills to be a good data steward including the ability to assess the quality of data, protect and secure data, and their responsible and ethical use.”

Statistics Canada, 2020-09-23a: Data Quality in Six Dimensions: Relevance; Accuracy; Timeliness; Interpretability; Coherence; Accessibility

Statistics Canada 2020-09-30: Steps of the data journey... The data journey is supported by a foundation of stewardship, metadata, standards and quality.

Panetta (Gartner), 2019 Digital Dexterity: “Further, data literacy is an underlying component of digital dexterity, which is an employee’s ability and desire to use existing and emerging technology to drive better business outcomes, another important skill for digital business.”

Bonikowska, et al. 2019.: “Data literacy also means having the knowledge and skills to be a good data steward including the ability to assess the quality of data, protect and secure data, and their responsible and ethical use.”

Bryla (2018): “With data everywhere, it will become the means of communication between IT and business, between the citizen data scientists and the domain experts... Understanding of where the data came from, how it was collected, and how it was analyzed and visualized are all essential elements of data literacy.”

Qlik, 2018 - Corporate Data Literacy: “We established the first definition for Corporate Data Literacy as the ability of an organization to read, analyze, utilize for decisions, argue with and communicate data throughout the organization.... The Data Literacy Index, commissioned by Qlik and produced by Wharton School academics and IHS Markit, reveals correlation between company performance and workforce data literacy.”

Sperry 2018: data literacy is the “ability to derive meaningful information from data”. (Cited in Bonikowska, et al.; caution - this cannot be found by Google Search)

Wolff et al., 2016: “Current definitions of data literacy are not fit for purpose - they don’t account for changes in the nature of data sets, which are becoming larger and more complex. Nor do they account for the different roles in which people must apply data literacy skills.” p. 10 “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) Cited in Bonikowska, et al. 2019.

Deahl, 2014: “The concept of information literacy has been criticized for its focus on building technical skills that are easy to acquire but quickly become obsolete. Discussions of media literacy, on the other hand, deemphasize skill acquisition and focus instead on ways of supporting media production and a deeper critical understanding of media focused on modes of representation, language, production, and audience.”

Mandinach & Gummer (2013): Data literacy is “The ability to understand and use data effectively to inform decisions. It is composed of a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge. These skills include knowing how to identify, collect, organise, analyse, summarise and prioritise data. They also include how to develop hypotheses, identify problems, interpret the data, and determine, plan, implement, and monitor courses of action.” Cited by Wolff et. al. 2016.

Vahey et al. 2006: “data literacy includes the ability to formulate and answer questions using data as part of evidence-based thinking; use appropriate data, tools, and representations to support this thinking; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions.” Cited by Wolff et. al. 2016.

Information Literacy Model

Drawing from information literacy coursework, researchers “have used the ACRL's (2000) information literacy standards and initial research findings to suggest lists of competencies for data literacy... the various models are comprised of skills related to accessing, managing, communicating, preserving and ethically using data for both undergraduate and graduate education.” Maybee & Zilinski (2016).

Hunt (2004): data literacy education should borrow heavily from information literacy education, even if the domain of data literacy is more fragmented than the field of information literacy.” (Koltay, 2016). As Maybee & Zilinski (2016) write, “The emerging construct of data literacy has typically been closely related to information literacy.”

Framework for Information Literacy for Higher Education (ACRL, 2015) - see <http://www.ala.org/acrl/standards/ilframework>

Maybee & Zilinski (2016): “challenges have been made concerning the efficacy of generic approaches to information literacy, such as the standards, for enabling people to use information in the various contexts in which they live and work (e.g., Bruce, 1997; Lloyd, 2010).”

	Description
Awareness	Understand what data are and its role in communities or society
Access	Understand how to identify, locate, and use appropriate datasets and databases
Engage	Evaluate, analyze, organize, and interpret existing data; make decisions based on data
Manage	Plan for and manage data, including organizing and analyzing, developing protocols for security and storage, sharing, and documentation
Communicate	Synthesize, visualize, and represent data
Ethical use	Acknowledge data sources, evaluate and manage risks, understand issues related to data reuse
Preserve	Be aware of curation practices for long-term storage and use

Elements of generic data literacy models (Maybee & Zilinski, (2016)

Pothier & Condon, 2020: Seven Data Competencies

Carretero, Vuorikari & Punie, 2017: European Digital Competencies Framework: Information and data literacy; Browsing, searching, filtering data, information and digital content; Evaluating data, information and digital content; Managing data, information and digital content.

Maybee & Zilinski, 2016: Data Informed Learning: proposed Elements of generic data literacy models (see above); “we advance data informed learning as a new framework for data literacy for use in higher education.”

Carlson and StowellBracke, 2015: A pilot data literacy program on data literacy offered at Purdue University was built around information literacy skills.

Calzada Prado and Marzal, 2015: data literacy as a form of information literacy. Precedent concepts:

- quantitative literacy (“using simple mathematical concepts to solve everyday problems” (Steele and Kiliç-Bahi 2008, 1))
- statistical literacy (“the ability to understand and critically evaluate statistical results that permeate our daily lives - coupled with the ability to appreciate the contribution that statistical thinking can make in public and private, professional and personal decisions” (Wallman 1993, 1))

Data Literacy: “their similarities may lead to regarding data literacy as a complement to or a form of information literacy, as respectively suggested by Stephenson and Caravello (2007) and Otto (2012), which makes us think that data literacy would be the umbrella concept covering statistical literacy, rather than the vice versa.” (p. 125)

Carlson & Johnston (eds) 2015: Data Information Literacy (Purdue). "Information literacy implies the ability to find, manage, and use information in any format, and editors Carlson and Johnston apply it to the format of raw

data. They coined the term data information literacy as an application of information literacy in the context of research."

Alton Grizzle and Maria Carme Torras Calvo, eds. 2013: Media information literacy. "MIL (Media information literacy), understood as a composite concept, encompasses knowledge, skills and attitudes that enable citizens to understand the role and functions of media and other information providers in democratic societies."

Calzada Prado & Marzal, 2013: Data literacy as component of information literacy "identified a number of abilities, some of which clearly show their origin in the best-known definition of information literacy (ALA, 1989) and the Information Literacy Competency Standards for Higher Education (ACRL, 2000). They also emphasize the ability to identify the context in which data is produced and reused. By mentioning these two main components of the data lifecycle they are in line with contemporary views of information literacy that incorporate the understanding of how information is produced (ACRL,2013)." "Data literacy can be defined, then, as the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data." (p. 126)

Andretta, et al. 2008: data as a learning outcome of information literacy. "Andretta et al. (2008) identified presenting, evaluating and interpreting qualitative and quantitative data as a learning outcome of information literacy." (cited by Koltay, 2016)

Statistical Literacy

Schild, 2004: "data literacy is the part of statistical literacy that involves training individuals to access, assess, manipulate, summarize and present data, whereas statistical literacy (also) aims to teach how to think critically about descriptive statistics." (cited by Calzada Prado & Marzal (2013), p. 125)

Bundy, 2004: "The Australian and New Zealand Information Literacy Framework, edited by Alan Bundy (2004) states that information literate persons obtain, store and disseminate not only text, but data as well." (Koltay, 2016)

Australia Bureau of Statistics, 2010: statistical literacy standards. "Statistical literacy is essentially the ability to find, access, utilise, understand and communicate the story contained within the data. Sound understanding, interpretation and critical evaluation of statistical information can then contribute to decision making. The importance of statistical literacy in our information-rich society means that it has now become a core competency like reading and writing."

Science and Research Data Literacy

This model of data literacy emphasizes aspects of data related to computer science, mathematics and statistics. Though based on principles of data stewardship, it defines a set of data skills including data awareness, forms of statistical representation, the ability to analyze, interpret and evaluate statistical information, and communication of statistical information (Australian Bureau of Statistics, 2010). Koltay (2016) writes that related concepts include research data services, data governance (he notes, "valid data governance may require identifying 'people who informally already have a level of accountability for the data they define, produce and use to complete their jobs or functions'" p. 305).

Maybee & Zilinski (2016) write, "challenges have been made concerning the efficacy of generic approaches to information literacy, such as the standards, for enabling people to use information in the various contexts in which they live and work (e.g., Bruce, 1997; Lloyd, 2010)." These objections have led to more context-specific models of data literacy development, and in particular to (Bruce, 2008) "data informed learning as an approach to data literacy that shifts the focus from acquiring generic data-related skills to students learning how to use data in disciplinary contexts."

Chantel Ridsdale, et al., 2015: matrix of data literacy competencies: "Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner." p.2 "Ridsdale et al. (2015) set up a matrix of data literacy competencies with the intention to foster an ongoing conversation about standards of data literacy and learning outcomes in data literacy education." (Koltay, 2016)

Koltay, 2016: research data services and data governance: "Data literacy is closely related to research data services that include research data management (RDM)." p. 303 Related concepts: research data services, data governance ("valid data governance may require identifying 'people who informally already have a level of accountability for the data they define, produce and use to complete their jobs or functions'" p. 305)

Koltay, 2015a: transform data into information: "data literacy can be defined as a specific skill set and knowledgebase, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage and ethically use data (Koltay, 2015a)." (need to verify: Koltay T (2015a) Data literacy: In search of a name and identity. Journal of Documentation 71 (2): 401-415.)

Searle et al., 2015: "identify data literacy as one of RDSs activities that support researchers in building the skills and knowledge required to manage data well. Therefore, we can say that data literacy is related to practically all processes that are covered by RDSs, and build the main framework of libraries' involvement in supporting the data-intensive paradigm of research (Tenopir et al., 2014). RDSs are undoubtedly comprehensive, thus covering their aspects makes data literacy overarching and comprehensive." (Koltay, 2016)

Johnson, 2012: "While the various definitions of data literacy will be discussed below, we define it here as the ability to process, sort and filter vast quantities of information, which requires knowing how to search, how to filter and process, to produce and synthesize it (Johnson, 2012)."

Social Engagement Model

This may be a nascent model. Thus far, it is only really articulated in a single source (Rahul Bhargava, et. al., 2015), though it has its origins in a broader definition of literacy, as exemplified by Robinson (2005), who talks of literacy as enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society" (p. 13). It distinguishes between the need for everyday uses of data from the deeper requirements of data science.

Rahul Bhargava, et. al. 2015: engage in society through and about data: "We define data literacy as "the desire and ability to constructively engage in society through and about data."

Robinson, 2005: "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." (p. 13)

Appendix 4 – Data Literacy Competencies

Data literacy competencies are the knowledge and skills you need to effectively work with data. The list of competencies was extracted from the literature surveyed in this report. The number represents the number of times (out of 20) the competency was listed. Please note that references in this section are listed in Appendix 1.

Change Management

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 (Statistics Canada, 2020).

Some types of data – master data and reference data – should have tightly controlled sets of valid values. These values appear in thousands and millions of transactions; without change control, different repositories storing master and reference data get out of sync (Data Governance Institute, 2020).

Open Data Institute, 2020; Duncan, et al., 2021; Gartner, 2019; Wells, 2022; Data Governance Institute, 2022;

Citation and Sharing

Knowledge of widely-accepted data citation methods, creates correct citations for secondary data sets (Ridsdale, et al., 2015).

Although you are likely creating your own, original data visualizations, they are based on external data sources. Any reader who is looking at your data visualization should be able to find its original source. Don't forget to cite your data source that you used to create your visualization. Review some tips on citing data from the Ryerson University library (Mulvaney, et al., 2022)

Ridsdale, et al., 2015; Mulvaney, et al., 2022; Grillenberger & Romeike, 2018; Australian Public Service Commission, 2021;

Critical Thinking

Conceptually, data literacy requires critical thinking, gaining knowledge from abstraction, and application of results. This critical, and often abstract reasoning is similar to computational thinking. Computational thinking involves “defining abstractions, working with multiple layers of abstraction and understanding the relationships among the different layers” (Ridsdale, et al., 2015).

Ridsdale, et al., 2015; Grillenberger & Romeike, 2018; Bowne-Anderson & Loukides, 2022;

Data Analysis

The knowledge and skills required to ask and answer a range of questions by analyzing data including developing an analytical plan; selecting and using appropriate statistical techniques and tools; and interpreting, evaluating and comparing results with other findings. (Statistics Canada, 2020). Eg. predictive and prescriptive analytics.

Statistics Canada, 2021; Heeley & Wilkinson, 2018; Statistics Canada, 2020; Open Data Institute, 2020; Duncan, et al., 2021; Gartner, 2019; Apolitical, 2021; Learn2Analyze, 2017; Ridsdale, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022; Van Andel Education Institute;

Data Awareness

The knowledge required to know what data is and what are different types of data. This includes understanding the use of data concepts and definitions (Statistics Canada, 2020).

Statistics Canada, 2021; Statistics Canada, 2020; Ridsdale, et al., 2015; Wolff, et al., 2016; Gummer & Mandinach, 2015; Wells, 2022;

Data Cleaning

The knowledge and skills to determine if data are 'clean' and use the best method and tools to take necessary actions to resolve any problems to ensure data are in a suitable form for analysis (Statistics Canada, 2020).

Statistics Canada, 2020; Open Data Institute, 2020; Apolitical, 2021; Learn2Analyze, 2017;

Data Communities

Data communities - networks of engaged data users within an organization - represent a way for businesses to create conditions where people can immerse themselves in the language of data, encouraging data literacy and fueling excitement around data and analytics... to generate data insights that are truly valuable, people need to become fluent in data—to understand the data they see and participate in conversations where data is the lingua franca. Just as a professional who takes a job abroad needs to immerse herself in the native tongue, businesses who value data literacy need ways to immerse their people in the language of data (Compton, 2020).

Open Data Institute, 2020; Jason Compton, Forbes, 2020;

Data Conversion and Interoperability

Data processing methodology: understanding and/or applying statistical procedures used to deal with intermediate data and statistical outputs, e.g., weighting schemes, statistical adjustment, or methods for imputing missing values or source data (APS, 2021)

Knowledge of different data types and conversion methods, converts data from one format or file type to another (Ridsdale, et al., 2015)

Australian Public Service Commission, 2021; Ridsdale, et al., 2015; Databilities, 2020; Sternkopf & Mueller, 2018; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022;

Data Curation

Ensure that data is reliably retrievable for future reuse, and to determine what data is worth saving and for how long (Learn2Analyze, 2017).

You need to find data that is well-matched with analysis goals. Data seeking is not practical without first stating the goals and understanding the business, information, and data requirements. With known requirements, you can search for data. The datasets that you find should be evaluated for quality and trustworthiness, and sometimes to select the best fit among multiple datasets that are available (Wells, 2022).

Learn2Analyze, 2017; Ridsdale, et al., 2015; Wells, 2022; Koltay, 2022;

Data Description or Metadata

Metadata management: The discipline of managing information that describes various facets of a data asset to improve its usability. Shorthand: "Data about data." (Gartner, 2019, Toolkit).

The knowledge and skills required to extract and create meaningful documentation that will enable the correct usage and interpretation of the data. This includes the documentation of metadata which is the underlying definitions and descriptions about the data (Statistics Canada, 2020).

Open Data Institute, 2020; Gartner, 2019; Apolitical, 2021; Learn2Analyze, 2017; Ridsdate, et al., 2015; Databilities, 2020; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022; elearning Curve, 2022;

Data Discovery and Exploration

Automatically finding, visualizing and narrating important findings within datasets (such as correlations, exceptions, clusters, links and predictions) that are relevant to users without requiring them to build models or write algorithms (Gartner, 2019, Toolkit).

The knowledge and skills to search, identify, locate, and access data from a range of sources related to the needs of an organization (Statistics Canada, 2020). The methods include: summary statistics; frequency tables; outlier detection; and visualization to explore patterns and relationships in the data (Statistics Canada, 2020).

Statistics Canada, 2020; Nancy Law, David Woo, Jimmy de la Torre, Gary Wong, UNESCO Institute for Statistics, 2018; Gartner, 2019; Apolitical, 2021; Ridsdate, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015;

Data Ethics

The knowledge that allows a person to acquire, use, interpret and share data in an ethical manner including recognizing legal and ethical issues (e.g., biases, privacy) (Statistics Canada, 2020).

Be able to use the informed consent, be able to protect individuals' data privacy, confidentiality, integrity and security, be able to apply authorship, ownership, data access (governance), re-negotiation and data-sharing (Learn2Analyze, 2017).

Australian Public Service Commission, 2021; Statistics Canada, 2020; Open Data Institute, 2020; Apolitical, 2021; Learn2Analyze, 2017; Ridsdate, et al., 2015; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018;

Data Evaluation or Assessment

To analyse, compare and critically evaluate the credibility and reliability of sources of data, information and digital content. To analyse, interpret and critically evaluate the data, information and digital content (Law, et al., 2018).

Nancy Law, David Woo, Jimmy de la Torre, Gary Wong, UNESCO Institute for Statistics, 2018; Open Data Institute, 2020; Ridsdate, et al., 2015; Gummer & Mandinach, 2015; Wells, 2022;

Data Gathering or Data Collection

The knowledge and skills to gather data in simple and more complex forms to support the gatherer's needs. This could involve the planning, development and execution of surveys or gathering data from other sources such as administrative data, satellite or social media data (Statistics Canada, 2020).

Be able to obtain, access and gather the appropriate data and/or data sources, and be able to apply data limitations and quality measures (e.g., validity, reliability, biases in the data, difficulty in collection, accuracy, completeness)(Learn2Analyze, 2017).

Data collection, manipulation, and analysis processes: These behaviours that can be qualified mainly as motivational can be supplemented by distinct steps that are needed for data sharing. Some of them are described by Buckland (2011) (Koltay, 2015).

Statistics Canada, 2020; Korri Palmer, Safegraph, 2021; Apolitical, 2021; Learn2Analyze, 2017; Ridsdate, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Grillenberger & Romeike, 2018; Australian Public Service Commission, 2021; Wells, 2022; Koltay, 2022;

Data Governance or Stewardship

Knowledge and skills required to effectively manage data assets. This includes the oversight of data to ensure fitness for use, the accessibility of the data, and compliance with policies, directives and regulations (Statistics Canada, 2020).

Australian Public Service Commission, 2021; Heeley & Wilkinson, 2018; Statistics Canada, 2020; Korri Palmer, Safegraph, 2021; Open Data Institute, 2020; Duncan, et al., 2021; Apolitical, 2021; Australian Public Service Commission, 2021; Wells, 2022; elearning Curve, 2022;

Data Interpretation

The knowledge and skills required to read and understand tables, charts and graphs and identify points of interest. Interpretation of data also involves synthesizing information from related sources (Statistics Canada, 2020).

Databilities, 2020; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Ridsdate, et al., 2015; Grillenberger & Romeike, 2018; Wells, 2022;

Data Management

Consistently describing the core entities of an organization across different views/users of the same data, including: customers, prospects, citizens, suppliers, sites, hierarchies, chart of accounts etc. (Gartner, 2019, Toolkit).

The knowledge and skills required to navigate internal and external systems to locate, access, organize, protect and store data related to the organization's needs (Statistics Canada, 2020).

VAULTIS, short for data that is visible, accessible, understandable, linked, trustworthy, interoperable and secure. (Houston, AMCOM, 2021)

Heeley & Wilkinson, 2018; Nancy Law, David Woo, Jimmy de la Torre, Gary Wong, UNESCO Institute for Statistics, 2018; Open Data Institute, 2020; Duncan, et al., 2021; Gartner, 2019; Learn2Analyze, 2017; Ridsdate, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015; Houston, 2022;

Data Manipulation

Manipulating, processing, cleansing and combining data for further analysis or use. Automation of complex manipulation on large data volumes... covers reformatting, cleansing and combining of data from different sources for further analysis, storage or use. This includes simple processes to check and transform data, through to automating complex manipulation of large data volumes. This excludes analysis, in-depth domain knowledge and ongoing management of datasets (Heeley & Wilkinson, 2018).

Heeley & Wilkinson, 2018; Ridsdale, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018; Australian Public Service Commission, 2021; Wells, 2022;

Data Modeling or Architecture

The knowledge and skills required to apply advanced statistical and analytic techniques and tools (e.g. regression, machine learning, data mining) to perform data exploration and build accurate, valid and efficient modelling solutions that can be used to find relationships between data and make predictions about data (Statistics Canada, 2020).

Statistics Canada, 2020; Korri Palmer, Safegraph, 2021; Duncan, et al., 2021; Apolitical, 2021; Wells, 2022;

Data Policy

A data policy contains a set of rules, principles, and guidelines that provide a framework for different areas of data management throughout the enterprise, including but not limited to data governance, data quality, and data architecture. Osthus

A data policy enables an organization to consistently address the broad range of potential developments and scenarios that may arise related to its creation, processing, use, and sharing of digital data.

Open Data Institute, 2020; Chartered Professional Accountants (CPA), 2020;

Data Preservation and Reuse

Assesses requirements for preservation, assesses methods and tools for data preservation, preserves data (Ridsdale, et al., 2015).

Data preservation is the act of conserving and maintaining both the safety and integrity of data. Preservation is done through formal activities that are governed by policies, regulations and strategies directed towards protecting and prolonging the existence and authenticity of data and its metadata. The main goal of data preservation is to protect data from being lost or destroyed and to contribute to the reuse and progression of the data. (Wikipedia)

Databilities, 2020; Ridsdale, et al., 2015; Grillenberger & Romeike, 2018;

Data Quality

The knowledge and skills required to critically assess data sources to ensure they meet the needs of an organization. This includes identifying errors or problems and taking action to correct them. This also includes awareness of organizational policies, procedures and standards to ensure good quality data (Statistics Canada, 2020).

Data quality is one of the cornerstones of the data-intensive paradigm of scientific research that is determined by multiple factors. The first one is trust, which is complex in itself. The elements of trust include the lineage, version and error rate of data and the fact that they are understood and acceptable (Buckland, 2011).

Cognitive authority, which has two levels. At an operational level, cognitive authority is the extent to which users think that they can trust the information. On a more general level, cognitive authority refers to influences that a user would recognize as proper because the information therein is thought to be credible and worthy of belief (Rieh, 2002).

Authenticity, which measures the extent to which the data is judged to represent the proper ways of conducting scientific research, including the reliability of the instruments used to gather the data, the soundness of underlying theoretical frameworks, the completeness, accuracy, and validity of the data. In order to evaluate authenticity, the data must be understandable (Koltay, 2015).

Australian Public Service Commission, 2021; Apolitical, 2021; Ridsdate, et al., 2015; Mulvaney, et al., 2022; Databilities, 2020; Wolff, et al., 2016; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; elearning Curve, 2022; Koltay, 2022; Statistics Canada, 2020;

Data Requirements

The ability to understand and prioritise user needs; and identify how data can be efficiently integrated into processes... By understanding the processes and data with which they work, a data specialist can effectively improve the efficiency and effectiveness of their work. To do this they need to understand the dependencies of their parts of the business and the importance of clarity in understanding data and other problems. (Heeley & Wilkinson, 2018)

Heeley & Wilkinson, 2018; Apolitical, 2021;

Data Science and Machine Learning

Includes the following:

Natural language processing: NLP is a way for computers to analyze, understand and derive meaning from human language in a smart and useful way. NLP is a subset of artificial intelligence (AI). Examples: Sentiment analysis, speech-to-text recognition, automatic summarization and language translation.

Natural language generation: NLG automates the creation of language or content from data inputs. Examples: Weather reports, form letters and financial reports.

Artificial intelligence: AI is a set of related technologies that seems to emulate human thinking and action by learning, coming to its own conclusions and enhancing human cognitive performance (also known as cognitive computing) or replacing people on execution of nonroutine tasks.

Machine learning: ML algorithms are composed of many technologies (such as deep learning, neural networks and natural-language processing), used in unsupervised and supervised learning, that operate guided by lessons from existing information inputs. Example use cases: Autonomous vehicles; automatic speech recognition and generation; detecting novel concepts and abstractions.

Duncan, et al., 2021; Gartner, 2019;

Data Security

Individual is aware of the main policies around data security, sharing and licensing. This includes awareness of the Data Protection Act and the appropriate action to take in the instance of a suspected or actual breach (Heeley & Wilkinson, 2018).

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 (Statistics Canada, 2020).

Australian Public Service Commission, 2021; Heeley & Wilkinson, 2018; Duncan, et al., 2021; Learn2Analyze, 2017;

Data Standards

Generic procedures and standards on how to handle data are formalized and widespread, and benefits are understood at all levels of the organization (Statistics Canada, 2020).

Make use of open data standards to build technology that's easier to expand, upgrade and use with other technologies; meet API technical and data standards to help deliver better services (DSA, 2022)

Australian Public Service Commission, 2021; Open Data Institute, 2020; Apolitical, 2021; Data Standards Authority (DSA), 2022;

Data Storytelling

A combination of data visualization, narrative (the plotline) and context (the surrounding situation/scenario). (Gartner, 2019, Toolkit) [More]

A combination of data visualization, narrative (the plotline) and context (the surrounding situation/scenario). (Gartner, 2019, Toolkit).

The knowledge and skills required to describe key points of interest in statistical information (i.e., data that has been analyzed). This includes identifying the desired outcome of the presentation; identifying the audience's needs and level of familiarity with the subject; establishing the context; and selecting effective visualizations." (Statistics Canada, 2020).

Australian Public Service Commission, 2021; Statistics Canada, 2020; Open Data Institute, 2020; Gartner, 2019; Apolitical, 2021; Ridsdate, et al., 2015; Mulvaney, et al., 2022; Wells, 2022;

Data Strategy or Culture

Example: "Meetings are data-driven and analytically rich. Metrics and analytics are at the forefront of business decision-making, not an afterthought to validate an opinion. Data is trusted, and context of data is understood and appreciated. We discuss outcomes and moments powered by data and insight." (Gartner, 2019, Toolkit)

A data culture is a "learning environment within a school or district that includes attitudes, values, goals, norms of behavior and practices, accompanied by an explicit vision for data use by leadership for the importance and power that data can bring to the decision-making process. (Hamilton et al., 2009, p. 46) (Mandinach, 2012)

The rules of using and caring for data, offered by Goodman (2014), can guide researchers in their effort to ensure that their data and analyzes continue to be of value. (Cited by Koltay, 2015)

Australian Public Service Commission, 2021; Korri Palmer, Safegraph, 2021; Open Data Institute, 2020; Gartner, 2019; Sternkopf & Mueller, 2018; Mandinach, 2012; Koltay, 2022;

Data Systems and Tools

The knowledge and skills required to use appropriate software, tools, and processes to gather, organize, analyze, visualize and manage data (Statistics Canada, 2020).

Many different types of databases exist. The most common types in use today include flat files, spreadsheets, relational databases, multi-dimensional databases, and NoSQL databases. Data literate individuals need to be capable of working with databases of many different kinds (Wells, 2020)

Tools include:

- Spreadsheets (such as Excel and Open Office) – essential tools for quickly scanning and checking data as well as providing simple data manipulation and combination.
- Geographic Information Systems (such as QGIS and ESRI's Arc toolset) – tools for holding, managing and integrating spatial information.
- Databases (such as Access, SQL Server, Oracle, Postgres, PostGIS) – powerful tools for manipulating and querying larger volumes of data.
- Database management tools (such as Toad, SQL Developer) – provides clear access to database structure for exploring and querying databases (Heeley & Wilkinson, 2018).

Australian Public Service Commission, 2021; Statistics Canada, 2020; Open Data Institute, 2020; Ridsdate, et al., 2015; Wolff, et al., 2016; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022;

Data Valuation

Understanding the business value of data scientists, data engineers and business analysts and the importance of meeting them frequently and productively. ... understanding how data adds value to business decisions. (Gartner, 2019, Toolkit)

Gartner, 2019; Australian Public Service Commission, 2021;

Data Visualization

Use of dashboards (e.g., dials, gauges, charts and maps), infographics, flow charts, decision trees, slide show/series (Gartner, 2019).

The knowledge and skills required to create meaningful tables, charts and graphics to visually present data. This also includes evaluating the effectiveness of the visual representation (i.e., using the right chart) while ensuring accuracy to avoid misrepresentation (Statistics Canada, 2020).

Australian Public Service Commission, 2021; Heeley & Wilkinson, 2018; Statistics Canada, 2020; Open Data Institute, 2020; Apolitical, 2021; Ridsdate, et al., 2015; Mulvaney, et al., 2022; Databilities, 2020; Sternkopf & Mueller, 2018; Grillenberger & Romeike, 2018; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022; Statistics Canada, 2020;

Data-Informed Decision-Making

Examples: Leadership presentations include: key performance metrics; related analysis, visualization and storytelling; roles and moments affected are described and data-driven actions taken; explanation of results, business impact and outcomes achieved. (Gartner, 2019, Toolkit)

Cramer, Little & McHatton (2015) describe a recursive five-step data-based decision-making model: goal identification, data collection, data reflection, identifying areas for improvement, and dissemination. DDDM is not just about the numbers or the data. It is about making actionable the data by transforming them into usable knowledge (Mandinach, 2012).

The knowledge and skills required to use data to help in the decision-making and policy making process. This includes thinking critically when working with data; formulating appropriate business questions; identifying appropriate datasets; deciding on measurement priorities; prioritizing information garnered from data; converting data into actionable information; and weighing the merit and impact of possible solutions and decisions (Statistics Canada, 2020). Also the knowledge and skills required to evaluate a range of data sources and evidence in order to make decisions and take actions. This can include monitoring and evaluating the effectiveness of policies and programs (Statistics Canada, 2020).

Australian Public Service Commission, 2021; Statistics Canada, 2020; Open Data Institute, 2020; Duncan, et al., 2021; Gartner, 2019; Learn2Analyze, 2017; Ridsdate, et al., 2015; Databilities, 2020; Wolff, et al., 2016; Gummer & Mandinach, 2015; Cramer, et al., 2020; Swan, et al., 2011; Mandinach, 2012; Wells, 2022;

Generate Data

"Generate data is another subcomponent of the 'use data' component, and it includes 'understand assessment', which expands into understand statistics and psychometrics, and 'use formative and summative assessments.'" (Gummer & Mandinach, 2015)

Gummer & Mandinach, 2015;

Identifying Problems With Data

Mandinach & Gummer (2016) identify six dispositions: belief that all students can learn; belief in data/think critically; belief that improvement in education requires a continuous inquiry cycle; ethical use of data, including the protection of privacy and confidentiality of data; collaboration (vertically and horizontally); and communication skills with multiple audiences.

Bocala & Boudett (2015) identify three 'habits of mind': shared commitment to action, assessment, & adjustment; intentional collaboration; and relentless focus on the evidence.

Edith Gummer, Ellen B. Mandinach, *Teaching and Teacher Education*, 2016; Candice Bocala, Kathryn Parker Boudett, *Teachers College Record*, ; Elizabeth H. Schultheis, Melissa K. Kjelvik, *The American Biology Teacher*, 2020;

Inquiry Process

In the Inquiry Process "the components identify problems and frame questions lack subcomponents and elements... subelements (include) planning, guiding, designing, adjusting, differentiating, and individualizing instruction associated with the element." (Gummer & Mandinach, 2015).

Identifying and implementing change to create efficiencies and new opportunities by making existing processes, systems, tools and products better and/or creating new ones. (e.g.) Can identify deficiencies in current processes/systems and tools/ products, gain the required approval to make changes, and lead the implementation of those changes (APOS, 2021).

Wolff, et al., 2016; Gummer & Mandinach, 2015; Australian Public Service Commission, 2021; Wells, 2022;

Plan, Implement and Monitor

In order to identify their own knowledge gap in the first place, they must have an understanding of how to plan for data collection, know how data can be identified and obtained and conceive how this data might eventually provide an answer to their initial question (Wolff, et al., 2016).

Wolff, et al., 2016; Australian Public Service Commission, 2021;

Present 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 (Ridsdale, 2015).

Databilities, 2020; Sternkopf & Mueller, 2018; Ridsdale, et al., 2015;

Statistics and Critical Reasoning

Become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted; recognize questions for which the investigative process in statistics would be useful and should be able to answer questions using the investigative; produce graphical displays and numerical summaries and interpret what graphs do and do not reveal; recognize and be able to explain the central role of variability in the field; recognize and be able to explain the central role of randomness in designing studies and drawing conclusions; gain experience with how statistical models, including multivariable (GAISE, 2020).

Descriptive analysis develops statistics to illustrate the shape of the data, describing characteristics such as the distribution of values (Wells, 2022).

Know – understand – be able to interpret statistics commonly used with educational data (e.g., randomness, central tendencies, mean, standard deviation, significance) (Learn2Analyze, 2017).

Gartner, 2019; Learn2Analyze, 2017; Ridsdale, et al., 2015; Robert Carver, et al., American Statistical Association, 2016; Bargagliotti, et al., 2020; Australian Public Service Commission, 2021; Wells, 2022;

Using or Innovating With Data

Example: "Data is a prevalent element of ideation and how we explore new business ideas. In our meetings, it is common to hear: "What if we had access to that data? Could others leverage this data? Can we blend this data with that data? Who else might benefit from this data? What insights does this data provide? What if we share this data, and are we allowed to? What data is available from our partners?" (Gartner, 2019, Toolkit)

Nancy Law, David Woo, Jimmy de la Torre, Gary Wong, UNESCO Institute for Statistics, 2018; Open Data Institute, 2020; Gartner, 2019;

Appendix 3 – Assessment Frameworks

Guidelines for Assessment and Instruction in Statistics Education (GAISE)

- Robert Carver, et al.. (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf
- Anna Bargagliotti, et al. (2020). Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II). American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf

Rasch Model

“Rasch models make it possible to test the hypothesis that the particular challenges posed in a curriculum and on a test coherently represent the infinite population of all possible challenges in that domain.” - Wikipedia

R Ferdhiana, I Santigana, F Abdurrahman, Nurhasanah. (2021). A Rasch model analysis on the ability to analyze descriptive statistics: A case Study of FMIPA Universitas Syiah Kuala new students. Journal of Physics: Conference Series. <https://iopscience.iop.org/article/10.1088/1742-6596/1882/1/012040/meta>

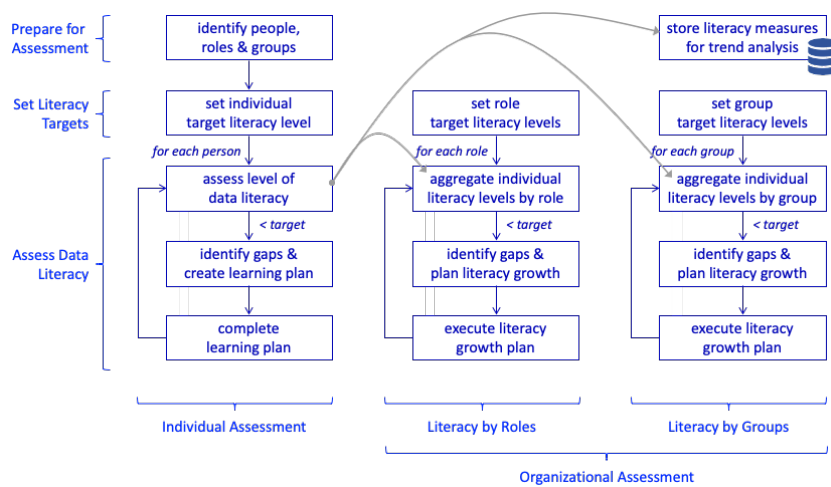
Essay Test

Suryadi, I K Mahardika, Supeno, Sudarti. (2021). Data literacy of high school students on physics learning. Journal of Physics: Conference Series.

https://www.researchgate.net/publication/350169382_Data_literacy_of_high_school_students_on_physics_learning

Data Literacy Assessment Process

Dave Wells. (2022). The Data Literacy Imperative - Part III: Data Literacy Assessment. Eckerson Group. <https://www.eckerson.com/articles/the-data-literacy-imperative-part-iii-data-literacy-assessment>



Appendix 4 – Teaching Frameworks

Data Wise Project

- Candice Bocala, Kathryn Parker Boudett, (2015). Teaching Educators Habits of Mind for Using Data Wisely, Teachers College Record. <https://www.tcrecord.org/Content.asp?ContentID=17853>
- Boudett, K. P., City, E. A., & Murnane, R. J. (2013). *Data Wise: A step-by-step guide to using assessment results to improve learning and teaching* (revised and expanded ed.). Cambridge, MA: Harvard Education Press.

Literacy Data-Driven Decisions (L3D)

- Mary Abbott, et al. (2017). A Team Approach to Data-Driven Decision-Making Literacy Instruction in Preschool Classrooms: Child Assessment and Intervention Through Classroom Team Self-Reflection. Young Exceptional Children. <https://files.eric.ed.gov/fulltext/EJ1151410.pdf>



Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Robert Carver, et al.. (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Anna Bargagliotti, et al. (2020). Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II). American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf

Case-Based Teaching Method

Derek R. Riddle, Jori S. Beck, Joseph John Morgan, Nancy Brown, Heather Whitesides. (2017). Making a case for case-based teaching in data literacy. Kappa Delta Pi Record. <https://www.tandfonline.com/doi/full/10.1080/00228958.2017.1334479>

Simulation

Rolf Biehler, Daniel Frischemeier, Susanne Podworny. (2015). Elementary preservice teachers' reasoning about modeling a family factory with TinkerPlots - A pilot study. Statistics Education Research Journal. http://iase-web.org/documents/SERJ/SERJ16%282%29_Biehler.pdf

Teaching for Statistical Literacy Hierarchy

Distinguishes between critical, widely used, and idiosyncratic concepts of statistical literacy.

Helen L. Chick, Robyn Pierce. (2011). Teaching for statistical literacy: Utilising affordances in real-world data.. International Journal of Science and Mathematics Education. <https://link.springer.com/article/10.1007/s10763-011-9303-2>

TABLE 1

A Teaching for Statistical Literacy Hierarchy (adapted from the statistical literacy hierarchy, Table 3, Watson & Callingham, 2003)

Level	Lesson characteristics for participating students	Examples with the Melbourne Water Data
6. Critical mathematical	Lessons at this level demand critical, questioning engagement with context, using sophisticated mathematical reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions and interpreting subtle aspects of language.	Lesson requires students to interpret the variation and trends in water levels throughout a year and over several years (e.g. seasonal changes or compare same months in different years), with differences discussed in percentage terms (not absolute terms); students make predictions while also discussing what may influence the actual outcomes.
5. Critical	Lessons at this level require critical, questioning engagement in familiar and unfamiliar contexts that do not involve sophisticated mathematical reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance and appreciation of variation.	Lesson requires students to interpret the variation and trends in water levels throughout a year and over several years (e.g. seasonal changes or compare same months in different years), with differences discussed in absolute terms; has students make predictions while also discussing what may influence the actual outcomes.
4. Consistent non-critical	Lessons at this level require appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only and statistical skills associated with the mean, simple probabilities and graph characteristics.	Lesson requires students to describe simple characteristics of the graph, but without discussing the variation or seasonal effects. Some comparisons might be made, but without questioning or encouraging explanation.
3. Inconsistent	Lessons at this level expect selective but inconsistent engagement with context, appropriate recognition of conclusions but without justification and qualitative rather than quantitative use of statistical ideas. Some statistical ideas evident, but not connected to each other or related to the context in any depth.	Lesson requires students to produce a table of data from the graph, e.g. from 1 year only, and may observe trends but without discussing reasons. Students also required to read graph data, but without being related to the broader context of the drought.
2. Informal	Lessons at this level require only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings and basic one-step straightforward tables, graphs and chance calculations.	Lesson requires students to read off the highest and the lowest water storage levels and calculate the difference; discussion may be about water evaporating.
1. Idiosyncratic	Lessons at this level suggest idiosyncratic engagement with context, tautological use of terminology and basic mathematical skills associated with one-to-one counting and reading cell values in tables.	Lesson fails to address any statistical concepts; or asks students to "read values" without attending to context or meaning; or incorporates activities that are not related to the given data (e.g. collect rainfall data from around the school)

Datastorming

"where data is materialized into a set of tangible artifacts to support design student's hands-on thinking [16] and collaboration for creating new design concepts."

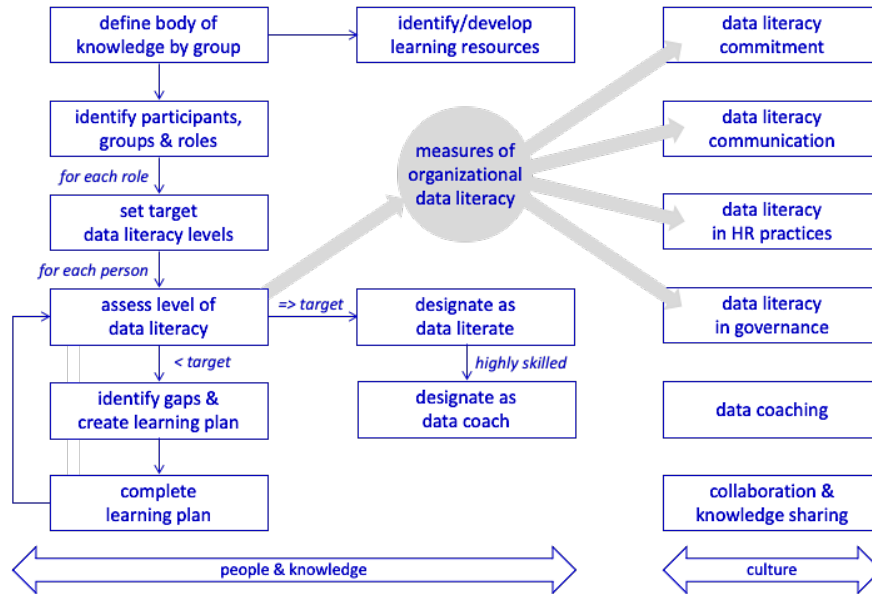
Delia Yi Min Lim, Christine Ee Ling Yap, Jung-Joo Lee. (2021). Datastorming: Crafting Data into Design Materials for Design Students' Creative Data Literacy. C&C '21: Creativity and Cognition. <https://dl.acm.org/doi/pdf/10.1145/3450741.3465246>

Authentic Data

Elizabeth H. Schultheis, Melissa K. Kjølvik. (2020). Using Messy, Authentic Data to Promote Data Literacy & Reveal the Nature of Science. The American Biology Teacher. https://www.researchgate.net/publication/345422494_Using_Messy_Authentic_Data_to_Promote_Data_Literacy_Reveal_the_Nature_of_Science

Data Literacy Program

Dave Wells. (2022). The Data Literacy Imperative - Part I: Building a Data Literacy Program. Eckerson Group. <https://www.eckerson.com/articles/the-data-literacy-imperative-part-i-building-a-data-literacy-program>



Appendix 5 – Global Data Literacy Initiatives

Government Initiatives

International

UNESCO

- Published by the UNESCO Institute for Statistics, is part of the Global Alliance to Monitor Learning (GAML), a Digital Literacy Global Framework was developed, <http://uis.unesco.org/en/blog/digital-literacy-skills-framework-measure>
- *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*
<http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf>
- *Recommendations on Assessment Tools for Monitoring Digital Literacy within UNESCO DLGF*
http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/12/4.4.2_02-Assessment-tools-for-monitoring-digital-literacy.pdf

OECD

- OECD has a website on skills development at <https://www.oecd.org/skills/R>
- Although not focused on Data Literacy, this report has some policy on dealing with stakeholders: *Strengthening the Governance of Skills System: a self- assessment tool.*
https://www.oecd.org/skills/centre-for-skills/Strengthening_the_Governance_of_Skills_Systems_Self_Assessment_Tool.pdf

The World Bank

- The World Bank has a Data Use and Literacy Program, “which delivers a range of activities to build capacity for data literacy and data use; enable data-driven decision-making; and democratize participation in the data revolution across low and middle income countries”.
- <https://www.worldbank.org/en/programs/data-use-and-literacy-program> Website has a blog, reports, and toolkit.

Australia

Government of Australia

Launch of the Data Capability Framework, October, 2021

- Australian Public Service Commission (2021). <https://www.apsc.gov.au/initiatives-and-programs/aps-professional-streams/aps-data-professional-stream/aps-data-professional-stream-news/launch-data-capability-framework>
- <https://www.apsc.gov.au/publication/aps-data-capability-framework>

Data Fluency

- Department of Industry, Science, Energy and Resources. (2021). Data Strategy 2021–2024. Government of Australia. <https://www.industry.gov.au/data-and-publications/data-strategy-2021-2024/1-capability-and-culture-from-data-literacy-to-data-fluency>

Queensland State Schools

- Data Literacy Framework provided by Queensland State Schools (<https://education.qld.gov.au/initiativesstrategies/Documents/data-literacy-framework.PDF>)

Canada

NSERC Award for 2020-2021 – Computing tools for learning and teaching data literacy –

- Fanny Chevalier, University of Toronto, Computer Science - \$39,000 (Discover Grants Program – Individual) Award Summary - Data literacy is a fundamental part of evidence-informed reasoning ..
- https://www.nserc-crsng.gc.ca/ase-oro/Details-Detailles_eng.asp?id=713367

NSERC Award for 2020-2021 – From Personal Data visualization to Data –Empowered Citizens

- Charles Perin, University of Victoria, Computer Science - \$33,000 (Discover Grants Program – Individual)
 - Award Summary - This research focuses on empowering citizens with data through promoting visualization of personal data.
 - https://www.nserc-crsng.gc.ca/ase-oro/Details-Detailles_eng.asp?id=715541

Statistics Canada

- Data literacy competencies at <https://www.statcan.gc.ca/en/wtc/data-literacy/competencies>
- Data journey: The data journey represents the key stages of the data process starting with finding and exploring data through to telling the data story. <https://www.statcan.gc.ca/en/wtc/data-literacy/journey>
- Learning catalogue with data literacy training <https://www.statcan.gc.ca/en/wtc/data-literacy>

Europe

European Commission's comprehensive and widely used framework for digital skills. (2016)

- JRC Publications Repository - DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model. <https://publications.jrc.ec.europa.eu/repository/handle/JRC101254>
- Gallery of Implementations (**30 case studies and 20 tools**) <https://ec.europa.eu/jrc/en/digcomp/implementation>

DataBuzz Project

- DataBuzz project is a high-tech, mobile educational lab, which is housed in a 13 meter electric bus. Its specific goal is to increase the data literacy of different segments of society in the Brussels region through inclusive and participatory games and workshops.
- <https://digitalcommons.uri.edu/jmle/vol12/iss3/9/> and <https://doi.org/10.23860/JMLE-2020-12-3-9>

SELFIE

- [SELFIE](#), has been developed with reference to an extension of DigComp for education institutions. It is targeted at school leaders, teachers and students to help schools identify digital literacy strengths and weaknesses and build a school improvement strategy. It has gone through a 2017 pilot phase with 67,000 users and its aim, included in the European Union's [Digital Education Action Plan](#), is to reach 1 million users by the end of 2019.
- https://ec.europa.eu/education/schools-go-digital/how-selfie-works_en

Data.Europa Academy (was European Data Portal)

- This is an open data portal for Europe, but it also includes some data training webinars, reading, videos and e-learning on topics such as techniques and technologies in the field of data visualization, understanding geo-information, understanding APIs, cleaning data, etc.
- <https://data.europa.eu/en/academy> More info at <https://www.capgemini.com/ca-en/news/capgemini-invent-awarded-contract-to-continue-developing-the-european-data-portal/>

France

PIX

- Pix is an online platform for assessment and certification of DigComp skills, managed by the French Ministry of National Education, Higher Education and Research
- <https://jill-jenn.net/static/works/pix-a-platform-for-certification-of-digital-competencies.pdf>

- www.pix.fr

Germany

Hochschulforum Digitalisierung (HFD)

- The Hochschulforum Digitalisierung (HFD) is sponsored by Germany's Federal Ministry of Education and Research (BMBF). <https://hochschulforumdigitalisierung.de/en>
- Future Skills; a Framework for Data Literacy; Competence Framework and Research Report , July 2020 (Working Paper #53)
- https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_53_Data_Literacy_Framework.pdf

U.K.

- UK National Data Strategy (NDS) <https://www.gov.uk/government/publications/uk-national-data-strategy>
- Open Data Institute - Data literacy and the UK government - <https://docs.google.com/document/d/19acuJFOnJEWzh21ahfgCYqCJ0KZo837mA2FAaWJqFPU/edit#>

United States

Air Force USAF

- Data Governance boot camp pilot <https://fcw.com/Articles/2020/11/30/vidrine-usaf-data-strategy.aspx?Page=2>
- A partnership with the Air Force Institute of Technology piloting an online graduate certificate program in data science
- Partnering with AFWERX. They virtually brought together groups of air and space professionals and worked on a single problem set, specifically, on the C-17 scheduling. That datathon offered a 92% scheduling accuracy rate using automated capabilities, using data sets.

Department of Health and Human Services (HHS)

- The Department of Health and Human Services launched its Data Science CoLab in 2017 to boost basic and intermediate data skills. <https://fedtechmagazine.com/article/2021/07/data-literacy-government-how-are-agencies-enhancing-data-skills-perfcon>
- HHS Data Science CoLab at <https://www.hhs.gov/cto/initiatives/data-science-colab/index.html>

Federal Government

- Federal Data Strategy 2020 Action Plan at <https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf>

Federal Cyber Reskilling Academy, CIO Council

- In 2019, the federal CIO Council launched the Federal Cyber Reskilling Academy
- <https://fedtechmagazine.com/article/2021/07/data-literacy-government-how-are-agencies-enhancing-data-skills-perfcon>
- <https://fedtechmagazine.com/article/2019/01/new-cybersecurity-reskilling-academy-train-feds-security-roles>

Federal Data Science Training Program

- The Office of Management and Budget launched the federal Data Science Training Program in the fall of 2020 to retrain workers to use new skills including data mining, data visualization, statistics, and enhanced presentation skills.
- <https://www.cio.gov/programs-and-events/data-science-training-program/>

Wisconsin Department of Public Instruction

- Wisconsin Department of Public Instruction provides an example instruction for teachers in the school system
- <https://dpi.wi.gov/strategic-assessment/professional-learning/assessment-and-data-literacy-e-learning-series/data-literacy-module>

Industry Initiatives

Data Literacy Foundation

- Data Literacy Foundation provides a Data Literacy tool kit,
- (New York, New York)
- EDLP – Enterprise Data Literacy Program framework
 - *Data Literates* podcast <https://dataliteracyfoundation.org/>

Data Literacy Project

- Data Literacy Project launched by Qlik includes founding partners Accenture, Cognizant, Experian, Pluralsight, the Chartered Institute of Marketing, and Data to the People
- <https://thedataliteracyproject.org/>

QuantHub

- QuantHub, a data skills platform, has a Roadmap for Creating a Data Literacy Program with additional articles and guides on the topic.
- <https://quanthub.com/data-literacy-program/>

SAS

- SAS has an education Outreach entitled Building a data literate future which includes apps for school age children.
- <https://curiosity.sas.com/en/programs/edu-outreach-building-a-data-literate-future.html>

State of Open Data

- State of Open Data has included a chapter on Issues on Open Data – Data Literacy
- <https://www.stateofopendata.od4d.net/chapters/issues/data-literacy.html>

Sigma

- Sigma has created a step by step program on developing data literacy.
- <https://www.sigmacomputing.com/blog/how-to-build-a-data-literacy-program-a-step-by-step-framework-for-data-and-business-leaders/>

Sport Scientist Canada

- Sport Scientist Canada has a 60-minute online course, to learn about and apply data science concepts and skills related to the following topics:
- Learn more at <https://www.sportscientistcanada.ca/en-CA/Programs/Data-Science-Module>

Tableau

- Tableau is providing free data skills training for individuals and organizations, a 5 hour e-learning program that helps anyone learn foundational data skills– announced on Oct. 14, 2020.
- Data Literacy for ALL <https://www.tableau.com/learn/data-literacy>

Universities Initiatives

International

DataPop Alliance

- Data-Pop Alliance is a collaborative laboratory created by the Harvard Humanitarian Initiative, MIT Connection Science, and Overseas Development Institute.
- <https://datapopalliance.org/dataliteracy/workshops/>
- Also have an open learning hub at <https://datapopalliance.org/dataliteracy/open-learning-hub/>

United States

University of Georgia

- The Data Literacy Committee proposal.” <https://ovpi.uga.edu/initiatives/data-literacy/>

Michigan State University - Data Literacy for Community Development Webinar Series

- MSU Extension’s Data Literacy for Community Development Webinar Series is an ongoing professional development opportunity to train community leaders in data collection, visualization and processing using open source applications that are widely available to the public.

MIT Management Sloan School

- This website has links to many informative articles on data literacy.
- <https://mitsloan.mit.edu/ideas-made-to-matter/how-to-build-data-literacy-your-company>

Rutgers University

- This website has a comprehensive list of courses and workshops at Rutgers and elsewhere.
- https://libguides.rutgers.edu/data_literacy/online_courses

