Three frameworks for data literacy

Downes, Stephen

Three Frameworks for Data Literacy

Stephen Downes

CELDA 20203

Madeira, October 21, 2023

Data Literacy

- 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
- a relatively new field of study, dating only from the 2010s

Three Frameworks

- Data literacy education across three frameworks:
  - the competency model defining data literacy,
  - the assessment of data literacy competencies, and
  - methods for the development of data literacy in an organization.

Literacy

- ‘learning’ a ‘literacy’ involves more than learning about the components of that literacy,
- and that there is an element of ‘being literate’, which is intended as an outcome of that learning.
- To be literate is to:
  - embody a set of skills and competencies typically thought to define that literacy,
  - as reflected in an assessment of that literacy, and
  - which in turn informs the teaching of that literacy.
- The study of data literacy limited
  - This paper seeks to fill that gap

Methodology

- A formal review from Canada's National Science Library for publications related to the definition, application and development of data literacy; a wider search using the same parameters was undertaken using Google Scholar.
- Approximately 150 results were obtained:
  - from which 20 items were found to contain an identifiable data literacy model
  - and three major assessment frameworks were identified
  - A small number of highly specific data literacy development models were also identified.
- The design framework employed draw from previous work by the author on connectivist massive open online courses (cMOOC)

Themes

- 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.

1. The competency model defining data literacy

Data Literacy as Competency Model

- Competencies:
  - “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.

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	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	x	x	
Management		x			x	x	x	x		x	x		x	x		x	x	x	x	
Preservation											x		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					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		x					
Change						x	x	x											x	
Using/Innovating With					x	x		x												
Identifying Problems											x		x	x	x					
Generate Data																	x			

Data Literacy Models

- The list of competencies identified also makes it clear that data literacy does not fall into any single category described above.
- 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:

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).

Analysis and Decision-Making Model

- This model is focused mostly on the use of data to support analytics and decision-making, for example, 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.

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.”

Science and Research Data Literacy Model

- This model of data literacy emphasizes aspects of data related to computer science, mathematics and statistics. 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).

Social Engagement Model

- This model distinguishes between the need for everyday uses of data from the deeper requirements of data science. 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 Workflow

- 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.
- For example, machine learning engineering describes the construction and use of these three elements: data engineering; model engineering; and deployment,

Individual and Group Competencies

- Data literacy is a concept that can be applied equally to both individuals and organizations
  - both the description of data literacy as well as the assessment of data literacy will vary in the given context

2. The assessment of data literacy competencies

Proposal

- Single-factor measure of data literacy ‘levels’ insufficient to account for
  - the variability in both the set of data literacy competencies
  - 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.

Assessment Programs

- We analyzed major skills and data literacy assessment programs, including the following:
  - OECD Programme for the International Assessment of Adult Competencies (PIAAC) literacy (Kirsch & Thorn, 2016, 2.2.1.3)
  - Endorsed by the American Statistical Association, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) (GAISE, 2016, 8).
  - Eckerson Group data literacy assessment including assessment not only of individual data literacy but also of the organization (Wells, 2021).

An Assessment Model

- We obtained an unstructured list of competencies
- 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
- For the sake of consistency with much of the work done previously a slightly modified version of Bloom's taxonomy is used (Bloom, 1956).

Table 1.

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

Role-Defined Dta Literacy

Knowledge
- Know what data is, recognize data vs non-data

Comprehension
- Know methods to read data, comprehend data

Application
- Know how data can be used

Analysis
- Understand parts of data, types of data

Synthesis
- Know ways to join of connect data

Evaluation
- Identify quality data, appropriate data

Creation
- Create data

Perception
- Be able to discover, read, explore data

Set
- Can follow data processes and procedures

Guided Response
- Can follow instructions and respond to data

Mechanism
- Knows about and can use data tools and systems

Complex Overt Response
- Can make decisions using data

Adaptation
- Can create data visualizations, stories

Origination
- Can create and share data from new sources

etc

+

OVERVIEW

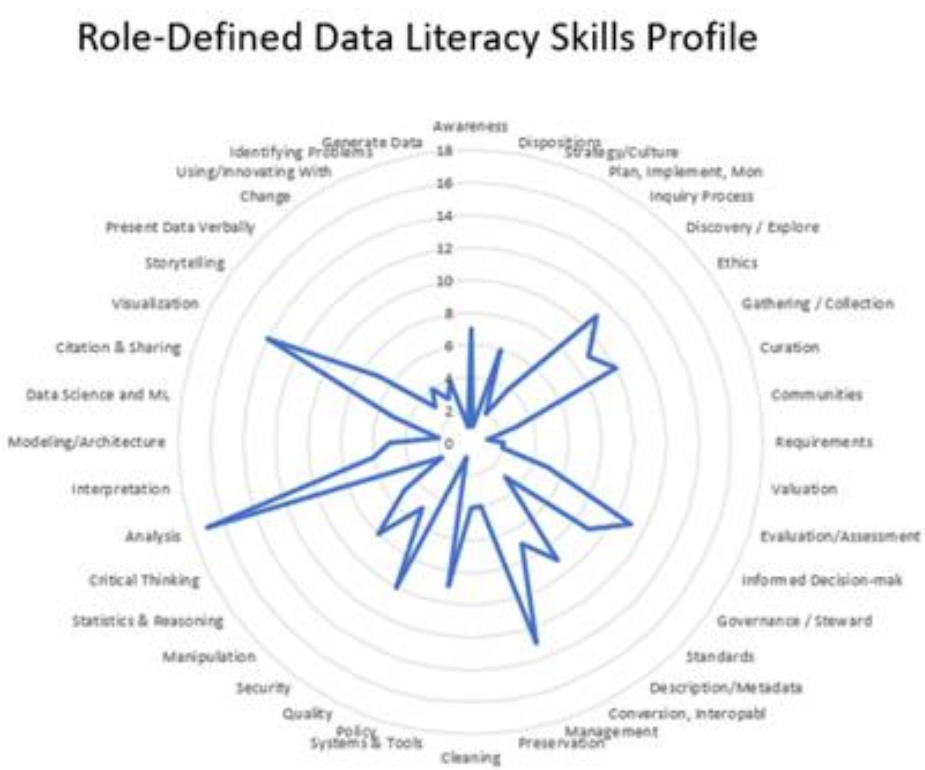
Information Systems Technicians are experts in Information Technologies (IT) who deploy, establish, administer, and maintain multi-platform networking computer environments, and a variety of data and voice networks. They are a part of a larger team that provides the Canadian Armed Forces (CAF) with communications and information services throughout Canada and around the world. They handle communications and information systems equipment, such as:

- Wired and wireless communications and information systems
- Fibre optic and copper wire broadband technology
- Voice and data network equipment and servers

WORK ENVIRONMENT

Information Systems Technicians experience the unique adventures and challenges that come with working outdoors, in military vehicles and server rooms. Information Systems Technicians work across the country and around the world wherever the CAF has a footprint.

=



Competency Profiles

- 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.

3. Methods for the development of data literacy in an organization

Where Data Literacy Fits

- Data literacy seems to fall within two extremes:
  - among other types of information and communication competencies, such as digital literacy or information management programs
  - 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.

Data Literacy Roadmaps

- Numerous data literacy program development roadmaps provided by commercial consultants:
  - The Data Information Literacy project - four-step methodology of planning, development, implementation, and assessment' (Carlson & Johnston, 2015).
  - QuantHub - a series of 'foundational steps' to develop a data literacy vision and roadmap; and an iterative process of assessment, planning, learning and practice (Cowell, 2020).
  - Dave Wells of Eckerson Group - 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 - 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.

More Initiatives

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

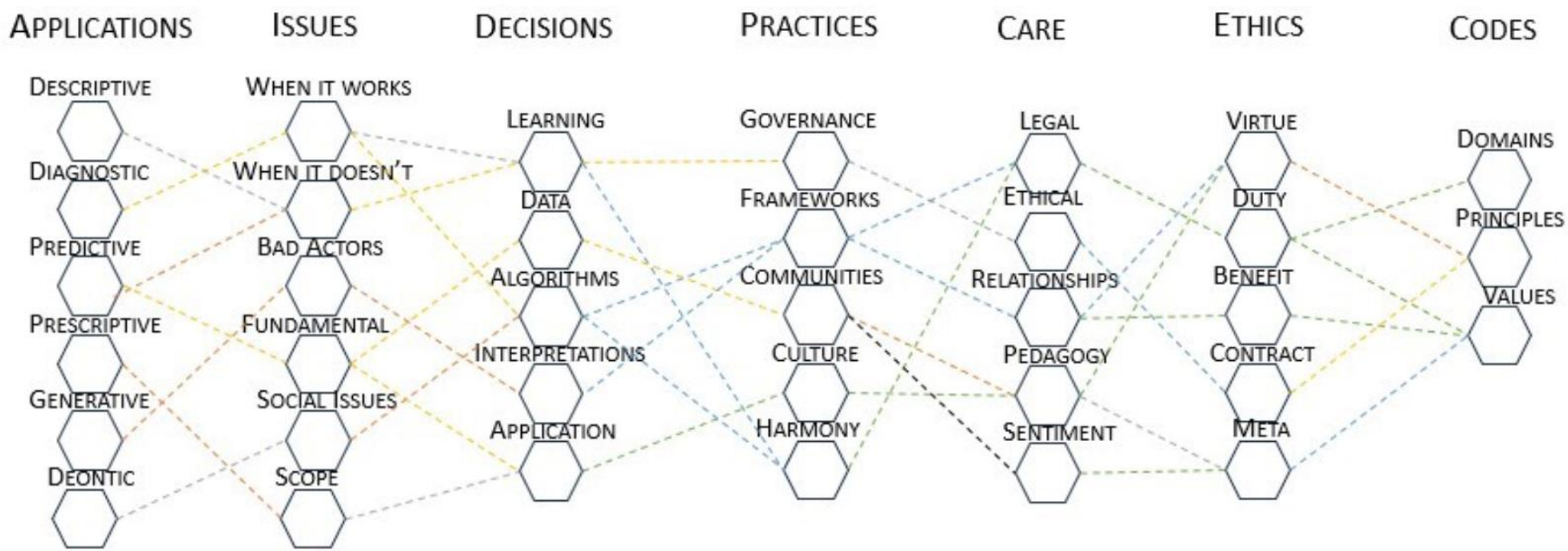
Teaching and Learning Methods

- Datastorming - craft abstract data into hands-on design materials in the form of cards.” (Lim, et al., 2021)
- Simulations and Interactive Technologies - a data visualization and modeling tool developed (Biehler, et al.,2016)
- Case-Based Teaching Method - discuss complex, real-life scenarios (Riddle, et al., 2017).
- Utilising affordances in real-world data - use real-world data from the perspective of the affordances in the data presentation (Chick & Pierce, 2012).
- Data-Driven Decision-Making - team-based approach combines a number of competency requirements in a single activity (Abbott, et al., 2015)

Data Literacy MOOC

- Connectivist MOOC was employed because
  - structured as a graph of connected people, resources, and concepts, in other words, much more like a collection of data.
  - a data-based MOOC (dMOOC) organizes content and resources in a structure suggested by the literature being studied in the course
- The next slide is a sample of the structure used in a similar dMOOC on ethics and analytics (ethics.mooc.ca)

Sample dMOOC



Student Activities in a dMOOC

- Less of learning and remembering content and more of working with relevant data, and specifically:
  - Classifying and labeling major sets and subsets of data
  - Identifying and labeling specific instances of data subjects (for example: an article describing ‘care’ as a legal concept)
  - Identifying and labeling relations between sets and subsets of data, either view argument threds in extant literature, or through data analytics of relevant bodies of literature
  - Assessing the resulting data model, identifying significant threads, and interpreting the resulting model

Participation in a dMOOC

- Does not involve individual study and retention of a pre-defined body of knowledge
- It requires working with others in order to develop not only individual capacities and skills, but also social or community capacities and skills.
  - These typically resist definition prior to the course
  - the consequence of such social interaction and application of a skill or practice is often the development of knew knowledge, approaches, and competencies.

What We Have Learned

- There is no single or simple definition of data literacy.
- But it is not yet taught that way.
- We recommend developing and piloting non-hierarchal cooperative learning environments, such as the cMOOC, for the development of organizational and social competencies required for data literacy.
- That said, these are assertions that need to be empirically tested before being widely adopted and applied.
- This paper offers the conceptual framework within which such assertions may be tested, but does not itself constitute a test of them

Epilogue

- This is a Picasso in the Funchal City Hall and that will later hang in my home
- Data doesn’t have to be rows and columns of text
- Learning how to work with anything, really, is learning how to work with data