Strategies and Best Practices for Data Literacy Education

Knowledge Synthesis Report

Chantel Ridsdale, James Rothwell, Mike Smit,
Hossam Ali-Hassan, Michael Bliemel, Dean Irvine,
Daniel Kelley, Stan Matwin, and Brad Wuetherick

Key Messages

BACKGROUND: We begin with a definition, synthesized from existing literature and refined based on expert input: ***Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner.*** It is an essential ability required in the global knowledge-based economy; the manipulation of data occurs in daily processes across all sectors and disciplines. An understanding of how decisions are informed by data, and how to collect, manage, evaluate, and apply this data in support of evidence-based decision-making, will benefit Canadian citizens, and will increasingly be required in knowledge economy jobs. Data literacy education is currently inconsistent across the public, private, and academic sectors, and data literacy training has not been approached systematically or formally at Canada's post-secondary institutions. There are also per-sector capability gaps, which makes it difficult to set realistic expectations of data-based skills.

CONSIDERATIONS: Developing the solid foundational knowledge of data literacy is integral to building discipline-/domain-specific knowledge and ensuring that citizens are able to use and apply these skills appropriately and diversely throughout their personal and professional lives. The best place to begin this initiative is the undergraduate curriculum in post-secondary institutions, due in part to their overarching goal of producing globally competitive, critically thinking, well-equipped graduates. Post-secondary curricula already introduce students to new theories and practices, and to new forms of literacy such as information literacy and computational thinking. Twenty-first century problems require twenty-first century skills (Pentland, 2013); adding data literacy explicitly to undergraduate curricula will help ensure graduates will be better equipped to meet the data skills gap in Canada's (and the global) workforce.

FINDINGS AND BEST PRACTICES: Data literacy education requires methods that engage and motivate students, as well as encourage task commitment. Best practices for teaching data literacy education include collaboration between educators, organizations, and institutions to ensure goals are being met by all stakeholders; diverse and creative teaching approaches and environment including the effective use of technology; successive/iterative learning with complementary skills integrated (e.g. project-based learning); emphasizing mechanics in addition to concepts (i.e. practical, hands on learning); and increasing engagement with the content by using real world data. Courses built on this model will connect learning with contributing to society or personal interests, and encourages both in-school and lifelong learning. We have also identified gaps in our collective understanding of data literacy education, which will require further research.

DATA LITERACY COMPETENCIES: We have synthesized a set of skills and abilities that together comprise various levels of data literacy, which we present in a ***data literacy competencies matrix,*** organized by the five core aspects of our data literacy definition (data, collection, management, evaluation, application). This matrix is intended to form the basis of ongoing conversations about standards for assessing and evaluating levels of data literacy, and to inform the creation of learning outcomes in data literacy education.

CONCLUSION: For the benefit of students, employers, and society, data literacy must be recognized as a necessary civic skill (Swan et al., 2009). This recognition should come from all levels of government, and from post-secondary institutions. There needs to be agreement on what elements of data literacy are necessary in an undergraduate core curriculum, in order to provide a consistent foundational education for those entering an increasingly data-dependent workforce.

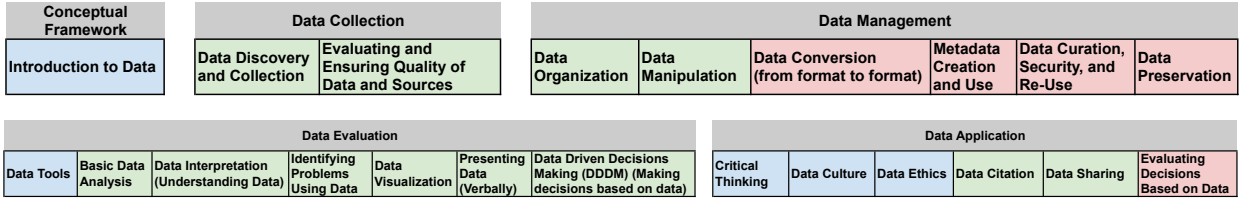
Executive Summary

We are a data-rich society; perhaps even data-driven (Pentland, 2013). In 2012, analysts estimated 90% of the world's data had come into existence within the previous 2 years (Vesset et al., 2014). Organizations in all sectors are struggling with this volume of data, confident that despite the velocity at which it is growing, and the variety of its formats, there is value. The goal is to transition from being data-rich to being information-rich and knowledge-rich, for which we need both data scientists and people capable of working effectively with data. The McKinsey Global Institute suggested that at current training rates, in the US alone there will be 140,000-190,000 more jobs than trained data scientists by 2018 (Manyika et al., 2011). On the literacy, fluency, mastery scale, a data scientist would have achieved mastery. However, the same report also estimated a 1,500,000 employee shortfall of "data-savvy" analysts and managers capable of working with the data to make effective decisions (Manyika et al., 2011); IDC suggests a similar number (Vesset et al., 2014). This latter set of skills is what we refer to as *data literacy*.

Across academic disciplines and throughout the private sector, we are recognizing a growing need for data-literate graduates from all backgrounds. The recent Tri-Council consultation document on digital scholarship (Government of Canada, 2013) recognizes this challenge, and the issue of training in particular: "Digital data are the raw materials of the knowledge economy, and are becoming increasingly important for all areas of society, including industry… The same may be said of the capacity to capture, manage and preserve it, or the requisite training of personnel who can operate effectively in this milieu" (Government of Canada, 2013). This recognition prompts the core question addressed in this report: How can post-secondary institutions in Canada best equip graduates with the knowledge, understanding, and skills required for the data-rich knowledge economy?

We addressed this question by examining existing strategies and best practices for teaching data literacy, synthesizing documented explicit knowledge (from both formal and informal literature) using a narrative-synthesis methodology. When necessary, we used our team's expertise to aid in synthesizing and summarizing; this expertise spans multiple disciplines, including Science, Computer Science, Business, Information Management, Arts and Social Sciences, and Education.

We begin by establishing the skills that comprise data literacy. Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner. 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. This view of data literacy is central to our synthesis, which includes two primary sections: the context and strategic value of data literacy education, and best practices for teaching data literacy across disciplines. There also remains much we do not know, and further steps that need to be taken, to understand data literacy instructions.



Context and Strategic Value of Data Literacy Education

We examined the context of data literacy education in three main areas: Data Literacy as a 21st Century Skill for 21st Century Citizens; Canadian Employers and Economy; and Canadian Students and Universities. These three areas help understand the motivation for ensuring (at minimum) foundational knowledge of data literacy.

Twenty-first century citizens must harness twenty-first century skills to be successful in the knowledge-based economy. Information is in abundance, and information is derived from data. Data comes from innumerable producers, through an increasing number of outlets, in diverse formats. The information/data atmosphere in society requires individuals to employ higher-order thinking, which can be challenging to teach, and often involves non-traditional instruction. Twenty-first century skills include critical thinking, problem solving, and computational thinking. These skills are difficult to hone when not built into curricula with intentionality. Critical thinking is a foundational skill for 21st century thinking and data literacy. Working with data requires the ability to ask the right questions and critically evaluate outcomes. Problem solving requires navigating difficult situations thoughtfully. Computational thinking incorporates a level of both critical thinking and problem solving; Wing describes the fundamental concepts as solving problems, designing systems, and understanding human behavior (2008).

A consistent level of data literacy education across the workforce would have a positive impact on employers, addressing the skills gap and the variance in data-related skills with which students enter the workforce. Acquiring data skills informally can be very difficult, and results in inconsistencies in practice and skill. The level of on-the-job training required would decrease, allowing employers to focus on domain-specific training, or elements of data skill where employees require mastery or fluency. As there is currently not a great deal of information about the specific expectations of employers in various industries and sectors, it is important to consult broadly when designing data literacy courses. The feedback available to date suggests that graduates are expected to be adaptive, with skills that have transferrable application in data, technologies, and methods. There is also a focus on data management, and the related information and knowledge management skills. Data must be findable and usable for subsequent analysis and synthesis; data not effectively managed from the point of collection becomes progressively more expensive to manage. One major gap in existing literature is how to train current members of the workforce in data literacy.

An important societal and student expectation of post-secondary institutions is that they produce globally competitive graduates. Data literacy, and the set of learning outcomes that align with data literacy, is being recognized internationally as a necessary skill in the twenty-first century. While not discussed in the literature, we have the sense that nationally we are behind but getting there; our data literacy competencies matrix is a starting point for discussing national standards. Teaching data literacy early develops foundational knowledge, which provides a basis on which to build disciplinary or domain specific skills and abilities. It also encourages cross-disciplinary thinking and applications, which can help students break out of academic silos, and enable creative and critical thinking. Post-secondary institutions must consider data literacy in its national context, identify how and where elements of data literacy are being taught in their existing courses and programs, systematically identify and fill gaps in this teaching (finding room in their academic timetables as necessary), and help students recognize data literacy (and/or its constituent elements) as a transferable skill.

Best Practices for Data Literacy Education

We identified several best practices for teaching data literacy in the literature, some of which differ from "traditional" strategies but would be consistent with teaching practices already in use in post-secondary institutions.

In any data literacy teaching scenario, the benefits of data, and data skills, must be clearly stated from the beginning. This is particularly true for mid-career learners, who will be more willing to invest their limited time and effort if they see the opportunity to help their community, industry, family, or others.

Hands-on learning in workshops and labs provides students with the necessary practical experience needed to fully understand a technical skill; students need the chance to figure out processes and methods on their own and make mistakes to readjust their own understanding. Mechanics are very important in data literacy; practice is required. Making mistakes can be frustrating, but will encourage critical thinking and problem solving.

Module-based learning allows students to achieve learning outcomes in stages, in a systematic way. Successive, or iterative, learning allows students to build upon previously learned skills, encouraging process over memorization or following rigid instructions, and ultimately making learning an unfamiliar concept more manageable. Beginning small and working up to the more complicated tasks allows students to have confidence in their abilities.

Project-based learning is a helpful way to implement the successive learning approach. Projects that include a wide range of investigation and have real-world applicability will solidify the connection between process/theory and practice. The project will allow evaluators the chance to assess skills practically, instead of formally.

Projects should include real-world data, relevant to the students' interests and in an engaging context, not just data for the sake of data. Increased engagement in working with data can foster innovation, improve learning, and increase the likelihood of lifelong learning. Projects should offer students the opportunity to go further than you expect.

Integrating data literacy teaching into existing subjects that make use of some element of data literacy is a way to integrate the systematic and formal teaching of data literacy into already-full curricula.

Research Gaps and Further Work

There are aspects of data literacy, and data literacy education, which are not addressed sufficiently by existing work. These include geospatial data literacy and GIS; sector-specific and industry-driven data literacy requirements with input from outside of academic institutions; no standard for assessing or evaluating data literacy levels; data security training for students without a computer science background; the ethics of data and data-driven decision-making; and how to provide data literacy training to the existing workforce in addition to new graduates. Our team will continue work in this area; we are developing a data literacy assessment tool, we have applied for academic innovation funding to produce course materials based on the results of this synthesis, and we will share the knowledge we've synthesized in appropriate venues. This report and other resources intended to assist in data literacy education will be posted to dataliteracy.ca.