Data Literacy: A Multidisciplinary Synthesis of the Literature

Chantel Ridsdale  
Dalhousie University  
Chantel.Ridsdale@Dal.Ca

James Rothwell  
Dalhousie University  
James.Rothwell@Dal.Ca

Hossam Ali-Hassan  
Dalhousie University  
Hossam@Dal.Ca

Michael Bliemel*  
Dalhousie University  
M.Bliemel@Dal.Ca

Dean Irvine  
Dalhousie University  
Dean.Irvine@Dal.Ca

Daniel Kelley  
Dalhousie University  
Dan.Kelley@Dal.Ca

Stan Matwin  
Dalhousie University  
Stan@cs.Dal.Ca

Michael Smit  
Dalhousie University  
Mike.Smit@Dal.Ca

Bradley Wuetherick  
Dalhousie University  
Brad.Wuetherick@Dal.Ca

Abstract

Data literacy is the ability to comprehend, create, and communicate data. Data literacy transcends any single discipline, and is still an emerging concept. This team was assembled to provide expert, multidisciplinary input in interpreting existing literature and identifying gaps. We conducted a systematic review, searching relevant electronic databases, grey literature, white papers, and governmental reports and policies for quantitative and qualitative studies to determine data literacy competencies, skills, and abilities, as well as teaching practices for undergraduate students. A synthesis of the literature allowed us to define detailed competencies for data literacy in key knowledge areas around data collection, data management, data evaluation, and data application; these competencies are broken down into conceptual, core and advanced competency levels, and further broken down into individual tasks. 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; emphasizing mechanics in addition to concepts (i.e. practical, hands-on learning); and increasing engagement with the content by using real-world data. We conclude with a discussion of the implications and recommendations for academic institutions including suggestions of how SAP University Alliances can be leveraged to teach data literacy competencies.

Keywords: Data Literacy, Education, Competencies
**Introduction**

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 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.

Our research attempts to answer the question of “How can post-secondary institutions best equip graduates with the knowledge, understanding, and skills required for the data-rich knowledge economy?” We examined existing strategies and best practices for teaching data literacy, synthesizing documented explicit knowledge (from both formal and informal literature) using a narrative-synthesis methodology. 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.

**Defining Data Literacy and its’ Competencies**

Data has become the currency of the new ‘Knowledge Economy’, and a critical driver of decision making in business, government, and social spheres (Chinien & Boutin, 2011; Cowan, Alencar, & McGarry, 2014; Mitrovic, 2015; Ikemoto & Marsh, 2008; Mandinach & Gummer, 2013). In this 21st century context, it is crucial that our citizens have the ability to contribute, interact with, and understand data (Mitrovic, 2015). In other words, citizens must be data literate; based on our synthesis of research, we have crafted the following definition: “Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner.”

In order for citizens to effectively engage and work with data, they must possess knowledge of the requisite theory and competencies. Data literacy shares the same theoretical grounding as information and statistical literacies, which are often taught at the postsecondary level (Hogenboom, Holler Phillips, & Hensley, 2011; Hunt, 2004; Koltay, 2014). 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 are organized under the top-level elements of the definition and are categorized as conceptual competencies, core competencies, and advanced competencies shown in Figure 1.

---

**Figure 1. Data Literacy Competencies**
**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 such as incorporating both formal and informal teaching methods into education, and providing students with tools and encouragement to develop skills outside of class, tutorials, and labs. A combination of these methods will improve students’ understanding and ability to use data (Wing, 2008; and Doucette & Fyfe, 2013).

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. Hands-on activities also help students refine their skills with experience (Johnson, 2012; Erwin, 2015; and MacMillen, 2014). Using critical thinking to work their way through their processes appropriately allows for mistakes, which enhancing the level and quality of their learning and long-term skill.

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. Building upon skills can be effective by letting increasingly complex data inform content (MacMillen, 2014). Students enjoy discovering their own conclusions, and this approach encourages exactly this.

Real-world data provides students with the opportunity for diverse experiences and caters to a wide range of skills (Carlson, Johnson, Westra, & Nichols, 2013; Romani, 2009; and Davenport & Kim, 2013). One approach to successive learning is project-based learning (PBL), a frequently tested and approach to engaging students to build their technical data skills (Romani, 2009). PBL is useful when teaching students critical thinking and problem solving (P21, 2012), which is useful because data manipulation requires creative thinking. An important aspect of PBL is choice (Burdette & McLaughlin, 2010); if students are going to be working on a project for extended periods of time, learning and building on technical skills, they must be engaged, and choice of the topics and data to examine in depth is an important component of this.

**Conclusions**

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. We have proposed a working definition for data literacy and identified its' related competencies; in addition to this, we have found several best practices for teaching data literacy that can serve to guide faculty designing curriculum in this field.

Working with real data and data systems is paramount to successful teaching of data literacy competencies. This requires partnerships between post-secondary educators, organizations with data that they are willing to share, and access to data processing technologies (hardware and software tools). Ideally curriculum is structured as progressively challenging modules with hands-on learning and working with real data in a problem based learning approach to foster critical thinking skills.

**Acknowledgements**

This research was funded by a Social Sciences and Humanities Research Council Knowledge Synthesis grant, and we would like to thank the graduate students Chantel Ridsdale and James Rothwell who were instrumental in conducting this analysis under the most excellent leadership of Dr. Mike Smit.
References


Systems Education, Vol. 17, pp. 441-448