

# 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.

Conceptual Framework		Data Collection			Data Management							
Introduction to Data		Data Discovery and Collection	Evaluating and Ensuring Quality of Data and Sources		Data Organization	Data Manipulation	Data Conversion (from format to format)	Metadata Creation and Use	Data Curation, Security, and Re-Use	Data Preservation		
Data Evaluation							Data Application					
Data Tools	Basic Data Analysis	Data Interpretation (Understanding Data)	Identifying Problems Using Data	Data Visualization	Presenting Data (Verbally)	Data Driven Decisions Making (DDDM) (Making decisions based on data)	Critical Thinking	Data Culture	Data Ethics	Data Citation	Data Sharing	Evaluating Decisions Based on Data

## **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 21<sup>st</sup> 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 21<sup>st</sup> 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](http://dataliteracy.ca).

## Table of Contents

Key Messages .....	1
Executive Summary .....	2
Context and Strategic Value of Data Literacy Education .....	3
Best Practices for Data Literacy Education .....	4
Research Gaps and Further Work .....	4
Table of Contents.....	5
Implications.....	7
Results .....	10
Toward a Shared Understanding of Data Literacy.....	10
Context and Strategic Value of Data Literacy Education .....	11
Building a 21st Century Citizenry .....	11
Barriers and Challenges .....	12
Canadian Employers and Economy.....	14
Right Tools, Right Job: Data Literacy Skills for the Workforce .....	14
Go Fish: Matching Skills with Requirements.....	15
Making the Call: Data Management and Decision Making.....	15
Barriers and Challenges .....	16
Canadian Universities and Graduates.....	17
Team Effort: Collaborating to Deliver Data Literacy Education.....	17
Barriers and Challenges .....	18
Best Practices for Teaching Data Literacy .....	18
Appropriate Timing of Data Literacy Education .....	19
Delivery: From Courses to Workshops .....	19
Integration of Data Literacy into Curricula .....	20
Emerging Teaching Approaches and Learning Environments .....	21
Engaging content with real world data to foster innovation .....	21
Successive/Iterative, Practical, Hands-on Learning.....	21

Assessment and Evaluation .....	22
Additional Resources.....	23
Further Research/Research Gaps .....	26
Geospatial/Temporal Data Management .....	26
Lack of Agreed Upon Standards and Best Practices .....	26
Data Security .....	27
Data Ethics .....	27
Data Literacy for the Existing Workforce.....	27
Knowledge Mobilization.....	27
References.....	29
Appendices.....	36

## Context

With the advent of advanced information communication technologies (ICTs), and the creation of true, globalized networks of commerce, communication, and transport, the flow and importance of information has become crucial to the socio-economic well-being of nation states. This connectivity, coupled with new technologies, has allowed for a dramatic increase in collection, analysis, sharing, and use of data (Pentland, 2013). Developments such as open data and 'Big Data' have led to a fundamental shift in how individuals and groups perceive and utilize data.

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). Innovative use of data can serve to improve services, and add value to existing products and processes (Cowan, Alencar, & McGarry, 2014; Gunther, 2007). In this 21st century context, it is crucial to Canada's socio-economic well-being 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 to date, 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). Thus, our team has drafted a knowledge synthesis based on a review of formal and informal literature that seeks to understand and share best practices for teaching data literacy at the postsecondary level. The purpose of this knowledge synthesis is two-fold:

- Identify key themes, knowledge gaps, and data literacy core competencies, so as to determine new avenues for research that can broaden our deeper understanding; and
- Provide the foundational knowledge necessary for the creation of an introductory data literacy course that could be taught at Canadian universities.

Data literacy transcends any single discipline, and is still an emerging concept. This team was assembled to provide expert, transdisciplinary input in interpreting existing literature and identifying gaps.

## Implications

The implications of data literacy education at the postsecondary level are far-reaching. The volume of data in the world is continuing to grow at an incredible rate. It was estimated in 2012 that 90% of the world's data had come into existence within the previous two years (Vesset, et al., 2014). The society of the 21st century is arguably a data rich one. Any country that does not have a technology and data-savvy citizenry will ultimately be left behind both socially and economically (Chinien & Boutin, 2011; Organization for Economic Co-Operation and Development, 2013; Pentland, 2013). Implications can be further subdivided into the following main areas:

1. Building a Shared Understanding of Data Literacy
2. Context and Strategic Value of Data Literacy Education
  - 2.1. Building a 21st Century Citizenry



- 2.1.1. Looking Forward: 21st Century Thinking in Action
- 2.1.2. Nuts and Bolts: Data Skills for the 21st Century
- 2.1.3. Barriers and Challenges
- 2.2. Canadian employers and economy
  - 2.2.1. Right Tools, Right Job: Data Literacy Skills for the Workforce
  - 2.2.2. Go Fish: Matching Skills with Requirements
  - 2.2.3. Making the Call: Data Management and Decision Making
  - 2.2.4. Barriers and Challenges
- 2.3. Canadian universities and graduates
  - 2.3.1. Team Effort: Collaborating to Deliver Data Literacy Education
  - 2.3.2. Barriers and Challenges
- 3. Best Practices for Data Literacy Education
  - 3.1. Delivery methods
  - 3.2. Creative teaching approaches
  - 3.3. Success/iterative learning
  - 3.4. Practical, hands-on learning
  - 3.5. Relevance to real-world data
  - 3.6. Assessment and Evaluation

21st Century Citizens are considered to be individuals who possess a high capability to interact with and utilize technology, think critically, make connections, and apply these skills within social and professional contexts (Chinien & Boutin, 2011; Wanner, 2015). Due to the interconnectivity brought on by intense globalization, challenges facing the next century will involve multiple actors, and cross disciplines and borders. These uniquely 21st century problems will require 21st century thinking (Pentland, 2013). Data literacy is considered one of the most relevant and “essential survival skills for the 21st century” (Chinien & Boutin, 2011, p. 8). It entails competencies that strengthen individuals’ aptitude to process complex cognitive problems, including the ability to analyze, create abstractions, and propose effective solutions to said problems (Chinien & Boutin, 2011; Cowan, Alencar, & McGarry, 2014; Yeh, Xie, Ke, 2011). Data literacy will allow Canadians to fully engage and tackle the new challenges, threats, and opportunities that will face Canada in the years to come.

Canadian universities (including Dalhousie University, University of Ottawa, Simon Fraser University, York University, Ryerson University, and others) offer specialized courses and degrees related to data (e.g. Master of Science in Big Data). However, there are few course offerings at the undergraduate level that directly focus on or explicitly incorporate comprehensive data literacy education. This is despite the increased importance and prevalence of research data management and data sharing within the social and applied sciences (Doucette & Fyfe, 2013; Martin & Leger-Hornby, 2012; Wanner, 2015). In order to stay on the cutting edge of research and teaching Canadian universities must ensure that graduates are encountering the core principles of data literacy into their core curriculum.

In addition to producing better graduates, data literacy education can help produce better students with improved learning skills and study habits. Data literacy is strongly intertwined with and shares overlapping competencies with information, statistical, digital, media, computational, and visual literacies (Hattwig et al., 2013; Mackey & Jacobson, 2011). This has been dubbed metaliteracy or transliteracy by scholars (Frau-Meigs, 2012; Hattwig et al., 2013; Mackey & Jacobson, 2011; Vahey et al., 2012; Zalles, 2005). Students who are exposed to meta or transliteracy in their learning will be

better equipped to apply their skills diversely, in varying settings, and carry out higher order thinking that utilizes multiple tools and systems (Frau-Meigs, 2012; Hattwig et al., 2013; Mackey & Jacobson, 2011). Data literacy is a critical aspect of this ability to evaluate, test, validate, and produce meaningful content (Frau-Meigs, 2012).

An individual would be hard-pressed to find a professional position in today's world that does not include working with data. Whether it is conducting interviews, performing observations, filling spreadsheets, reading charts, working with key performance indicator (KPI) dashboards, or assessing sales information, new graduates must have the basic skills to collect, interpret, analyze, evaluate, and produce data (Ikemoto & Marsh, 2008). On a higher level, data driven decision-making (DDDM) and analysis allows managers to identify actionable insights from the copious amount of data available (Ikemoto & Marsh, 2008; McKendrick, 2015). These skills are crucial for participating in the knowledge economy. Utilizing data effectively can also improve efficiency, reduce costs, and drive innovative product development (McKendrick, 2015). However, there is currently a gap in the skills that employees require, and those that graduates possess.

At current training rates, it is estimated in the United States alone there will be 140,000 - 190,000 more jobs than trained data scientists by 2018 (Manyika et al., 2011). The McKinsey Global Institute further estimates a 1.5 million shortfall of data-savvy (i.e. data literate) analysts and managers capable of working with the data (Ibid). On the job training can only go so far. By teaching data foundational literacy competencies early on in a postsecondary student's academic career, students are more likely to learn lifelong skills that will allow them to effectively and dynamically work with data and fully participate in the economy of today, and tomorrow.

## Approach

We conducted a transdisciplinary examination of existing strategies and best practices for teaching data literacy, synthesizing documented explicit knowledge using a narrative-synthesis methodology, informed by our transdisciplinary team with expertise in data in their fields. This process began with 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.

Our initial focus on synthesizing a definition allowed us to expand the terminology used in our searches. We found that 'data information literacy' was used often in relation to overlapping competencies like information literacy, and taught concurrently; however, information literacy is not within the scope of our review. A similar term, 'data management literacy', refers to the data creator, as opposed to data users (and was thus of limited use). There were also many articles written on specific, targeted areas of data literacy, such as 'data analytics'. We included some of these articles in order to round out our focus on foundational data literacy core competencies.

Following our initial sweep we began to narrow our focus, and began searching electronic databases for keywords, including: 'data litera\* OR data science OR data skills OR data-driven OR data savvy'; 'teach\* OR train\* OR develop\* OR educat\* OR instruct\*'; and university OR college OR higher education OR post-secondary'. The searches were narrowed to the years 2000-2015, because information prior to 2000 would not accurately reflect the current technological, academic, and professional situation.

This search proved difficult, because the interdisciplinary nature of our study required a foundational view of competencies. Looking at specific disciplines was more challenging than helpful, but there were a few databases that were most helpful: EbscoHost, ProQuest, SpringerLink, Taylor & Francis, and IEEE. This is a relatively new area of interest, and there were relatively few peer-reviewed articles available through this format. We further examined the bibliographies of the relevant articles, which widened our result set and introduced us to organizations and initiatives advocating data literacy to a variety of audiences and levels.

We searched for grey literature (not peer-reviewed but still formally published) using Google, and identified articles and white papers. As this is an emerging field, we generally found that the websites, courses and workshops, and associations and organizations were more current. These sources range in appropriateness from formal published articles to personal blogs, but we chose from the very beginning to include formal and informal literature to a comprehensive view of 'data literacy' as it is seen today.

We broadened our original scope for search to include some terms that proved helpful, but did not seem relevant before the initial literature sweep, including 'transliteracy', 'metaliteracy', 'computational thinking' and '21st century literacies'. These terms have overlapping competencies, and relate to data literacy directly, without using those words, because each clearly states that overarching need for critical thinking, problem solving, and communication. Additionally, we included articles on the skills gap impacting the use and adoption of Open Data and Big Data, particularly on those articles defining the necessary skills or best practices for imparting those skills.

We collected reports from provincial and federal governments, to provide context on the Canadian landscape. The most recent PISA report ranked Ontarians consistently above average in all literacies being taught K-12. For this reason, we believed that using their policies and reports would provide a better, more comprehensive and targeted outline of a successful data literacy core course that reflects the expectations of the 21st century.

We produced an annotated bibliography including all of the papers found as of the midpoint of this project (Appendix 4). Our thematic analysis of these resources is included in Appendix 3.

## Results

We have organized our findings functionally: what do we know about data literacy, and what do we establish as a shared understanding of what data literacy is and which competencies it includes; what is the current context for and strategic value of data literacy education; and finally (and most importantly) best practices for data literacy education.

### **Toward a Shared Understanding of Data Literacy**

We define data literacy briefly as "the ability to collect, manage, evaluate, and apply data, in a critical manner". This definition was synthesized from dozens of existing definitions (a word cloud generated from these definitions is in Appendix 2) and refined by our team.

This brief definition uses several loaded terms, which we use as the top-level of a hierarchy of competencies and tasks that comprise data literacy: collect, manage, evaluate, apply, and data. These terms are broadly defined and include a variety of elements considered core to data literacy. In Appendix 1, we present the complete set of competencies (23) and associated tasks/skills (64) in a

Data Literacy Competencies Matrix, and our summary table listing which of our 32 articles included each competency.

We note later in this synthesis that it is important that the definition of data literacy be left open, and informed by employers, students, and universities. We present this Data Literacy Competencies Matrix not to end the discussion, but to further it.

## **Context and Strategic Value of Data Literacy Education**

Data literacy is an increasingly necessary skill, required in a variety of wider communities, academia, and industries. The Association of College and Research Libraries (ACRL) posits that society has reached a critical intersection between societal/economic demand and academic demand (2013). Society is now a data rich environment, and data skills extend to all citizens, throughout all levels of society (Maycotte, 2014; Mitrovic, 2014). It is becoming increasingly important for the everyday person to have the skill to distinguish between ‘good’ and ‘bad’ representations of data (Twindale, Blake, & Grant, 2013; and Swan, Vahey, Kratcoski, van ‘t Hooft, Rafanan, & Stanford, 2009). Academia is beginning to recognize the necessity of preparing their graduates with more data-based skills for the workforce, and society as a whole, which is increasingly data-centered (Koltay, 2014; and Gunter, 2007). Providing graduates with foundational knowledge of data literacy allows for those students entering the 21st century workforce to apply diverse skills to a variety of situations. Industry feedback has also provided insight into what skills are valued in the workforce, and the overwhelming response was focused on data-related skills (Harris, 2012; Hu, 2012; and Koltay, 2014).

In this section, we examine these stakeholders in more detail: society (by understanding the 21<sup>st</sup> century citizen); employers and the economy; and students and universities.

### **Building a 21st Century Citizenry**

21st century challenges require 21st century skills. These skills can be more accurately described as ways of thinking (e.g. critical thinking, problem-solving, and computational thinking). There has been a shift in perspective from learning facts, to acquiring inquiry skills; these soft skills can be difficult to cultivate, but are necessary for success in the 21st century, and an increasingly data driven society (Pentland, 2013; Swan, et al., 2009; Yeh, Xie, & Ke, 2011; Boyles, 2012; Romani, 2009; and P21, 2012). Humans are social by nature, and making the connection between human behaviour and societal interactions can increase productivity, innovation, and creativity (Pentland, 2013; Wyner, 2013; Erwin, 2015; and Liqueste, 2012), as well as bridge the gap between abstract and reality, thereby helping the learning process.

Critical thinking is integral in this process and is considered a civic skill, as well as one of the foundations for 21st century thinking and data literacy (Pentland, 2013; and Swan, et al., 2009). Citizens’ ability to ask the right questions, evaluate findings, and be critical of concepts, claims, and arguments is essential, and the basis for success in the 21st century (Johnson, 2012; Koltay, 2015; Gunter, 2007; Schield, 2004; Burdette, 2010; and Liqueste, 2012). This is especially true since the explosion of uncurated information perpetuated by the onset of Web 2.0.

Due in part to this massive and continuous flow of information, problem-solving is no longer the act of making simple logical decisions; it has evolved into a series of complex issues, often with multiple layers. The type of elevated thinking required to tackle these issues is in increasing demand

throughout society and industry. The knowledge-based economy requires people to be able to navigate difficult situations in diverse ways, which is challenging, but critical to success in the 21st century (Liquete, 2012; Gunter, 2007; and Erwin, 2015).

Computational Thinking (CT) incorporates a level of both critical thinking and problem solving, and much more. Wing describes the fundamental concepts as solving problems, designing systems, and understanding human behaviour. It involves science, math, and engineering as drivers in the “push/pull” loop: “scientific discovery feeds technological innovation, which feeds new societal applications; in the reverse direction, new technology inspires new creative societal issues, which may demand new scientific discovery” (2008, p. 3722). Yeh, Xie, and Ke, (2011) as well as Johnson, et al., (2015) go further in explaining that tasks are growing in scope and complexity, and CT skills are very useful in navigating through these high-order problems.

Other transferable skills related to CT such as critical thinking and problem-solving encourage engagement, which motivates individuals to be creative and curious, providing foundations for lifelong learning. People are much more likely to become lifelong learners if they are engaged in inquiry that interest them, impact them, or relate to everyday life (Erwin, 2015; Burdette, 2010; and Wyner, 2013). Contributing to society and solving real problems effectively provides individuals with analytical skills to use throughout their career.

Project-based learning (PBL), with multiple levels of discovery, is a great way for people and students to learn data literacy. Utilizing their 21st century skills, PBL with real-world data engages individuals in higher order thinking, connects procedure to practice, and helps bridge the gap between learning facts and acquiring inquiry skills, critical reasoning, argumentation, and communication (Wyner, 2013; Romani, 2009; Erwin, 2015; and Swan, et al., 2009).

Twindale, Blake, and Grant take this further, explaining that data literacy can improve citizen engagement in the democratic process, as well as help them understand and participate in data-driven decision-making processes (2013). Ontario’s Green Button Initiative, whereby citizens can access their energy usage data online is an real-life example of how data literate users can explore opportunities to save and minimize their footprint (2015). Data literacy allows for data to be an everyday benefit to everyone. Littlejohn Shinder puts emphasis on the connection between knowledge and power, and states that the knowledge of how to deal with data may be one of the most powerful weapons professionals (and everyone) can have (2013, para. 8).

### *Barriers and Challenges*

Despite the arguments for data literacy being a necessity in the 21st century, there are several barriers impeding progress. First is society’s misconception that people born post-1983, referred to as ‘digital natives’ or the ‘net generation’, have inherent technological skills and abilities; in reality there is a complex and diverse range of skills in these students, which requires formal education to bridge the gaps (Jones, Ramanau, Cross & Healing, 2009; Thompson, 2012; and Romani, 2009). This misconception has resulted in a major skills gap in industry (Manyika, et al., 2011), and the daunting realization that this must be remedied.

Moreover, with little to no formal education, people already in the workforce are expected to seek practical knowledge and skills out independently through alternative avenues, creating disparity in practice as well as breadth of knowledge attained (Doucette & Fyfe, 2013; and Teal, Cranston, Lapp, White, Wilson, Ram & Pawlik, 2015). Romani asserts that people frequently rank themselves higher

on self-evaluation of skills, but actually rank lower in reality (2009). The scope, quality, and availability of resources is a challenge to self-learners as well; as Carlson, Johnson & Westra found self-directed learning through trial and error often results in focusing on mechanics rather than concepts (2013). Seeking out additional learning has value, but reinventing the wheel can be wasteful and expensive. The disparity in skills is a risk to society going forward.

The general need to educate the Canadian population about data is not a new phenomenon; for example, there have been initiatives throughout Canada encouraging public use of open data. Open Data and Big Data are common terms in academic circles and the general. This is in part due to the promise and ability to create transparency, and the potential to gain insights into patterns, trends, and relationships to predict behaviours. The opportunity for discovery and optimizing of raw data through open data and big data will only continue to grow (Littlejohn Shinder, 2013; McAulay, Rahemtulla, Goulding, & Souch, 2010; Cowan, Alencar & McGarry, 2014; and Manyika, Chui, Brown, Bughin, Roxborough & Hung Byers, 2011); however, many people do not have the skills necessary to engage with data in this way (Mitrovic, 2015). The Environment Commissioner of Ontario published a detailed report on this barrier with several perspectives from employers across multiple sectors reporting this fact, and arguing for government support for data literacy education, but it does not exist yet, and has been slow to develop (2015). This must be a priority for government, and should be done in collaboration with education systems to ensure consistent education across disciplines and institutions.

A suggestion to bridge the gap in formal education includes incentivizing learning for the Canadian population. This would involve courses that are non-credit based (Johnston & Jeffryes, 2014; Schneider, 2013; and ACRL, 2014), but that have an indicator of knowledge gained. This would be most effective if there were predetermined standards and skill levels. Our Competencies Matrix provides a base for skill levels concerning data literacy, and could act as an effective reference for standards being developed. CLA+ is an American assessment instrument that provides users with the option to have a certificate with results (2015), to distribute to employers, which will highlight the skills that are important in the 21st century workforce.

The benefits of data are potentially wide reaching, but cannot be achieved broadly if individuals in Canadian society are not capable of working with data. For concepts like Open and Big Data to become real and meaningful, citizens must be educated not only in the practical uses of data, but what benefits it provides personally and communally. They must be capable of thinking critically about the data they are presented. Society is made up of many varieties of people, and it should be recognized that there is disparity between social, economic, and cultural backgrounds which must be addressed before Big Data and Open Data can fully realize their potential (Czerkawski & Lyman, 2015). Gurstein (2011) and Mitrovic (2015) go further, arguing that if the public education and access is not addressed there will be an ever widening 'data divide' in society; resources and skills are being distributed to the people who already have access to them. These authors argue that there should be special attention provided to communities in poverty.

Another challenge to data in society is access. Data is published in so many places and formats, which makes it difficult to find and use (Cowan, Alencar & McGarry, 2014). People are less likely to take advantage of the benefits data has to offer if they must check several different sources regularly to produce accurate results. Additionally, data is often published in areas that the every person would never think to look, such as dense academic journals or within obscure websites that are not

user friendly (Miller, 2015). There is also the risk that volumes of Open Data will be used to obscure the most important data. Addressing these challenges is necessary to allow a prosperous and successful knowledge economy to develop.

### **Canadian Employers and Economy**

Teaching data literacy at the postsecondary level has far-reaching implications for Canadian employers, companies, and the economy as a whole. At its core, data has always been a crucial driver of decision-making in the business world and the larger economic system. However, with the continuing shift toward “knowledge, service, and information based activities” in our 21st century economy, the ability of firms to “create and commercialize knowledge has become tantamount to its ability to generate sustainable returns” (Boyles, 2012, p. 41). Moreover, society now demands twenty-four hour a day “100 percent reliability, 100 percent connectivity, instantaneous response, the ability to store anything and everything forever, and the ability for anyone to access anything from anywhere at any time” (Wing, 2008, p. 3723).

#### *Right Tools, Right Job: Data Literacy Skills for the Workforce*

It is critical for Canadian companies to possess the capability to skillfully collect, aggregate, search, manage, and analyze massive data sets in order to stay competitive (Koltay, 2014; McKendrick, 2015). This same data, if harnessed properly, can be used to improve workflow efficiency, foster innovative thinking and creativity, and improve problem-solving by empowering employees with actionable, rich information (Boyles, 2012; Gunther, 2007; McKendrick, 2015; Pentland, 2013). It is clear that data is no longer the niche of economists or statisticians. All professionals must have the competencies to work with data, and decision-makers must have the knowledge to effectively understand it, and utilize it (Davenport & Kim, 2013; McKendrick, 2015; Pryor & Donnelly, 2009). In short, we must have a data literate work-force.

Conceptually, data literacy requires critical thinking, gaining knowledge from abstraction, and application of results (Gunter, 2007; Qin & D’Ignazio, 2010<sup>b</sup>). This critical, and often abstract reasoning is similar to computational thinking (Wing, 2008). Computational thinking involves “defining abstractions, working with multiple layers of abstraction and understanding the relationships among the different layers” (Wing, 2008, p. 3718). Data in itself is by nature an abstraction, and is only useable if an individual can understand how it relates to other information in the wider world. This ability to understand complex relationships and connections further enhances capacity for curiosity and deep thinking (Davenport & Patil, 2012). For example, analysis of Big Data can provide insight into the functioning of society involving the flow of ideas and information (Pentland, 2013). For business, this insight into information flows can translate into timely information gathering, more efficient system monitoring, and facilitation of the spread of ideas that “form the basis of innovation” (Pentland, 2013, p. 2).

In order to build up the cognitive abilities to effectively work with data, data literacy instruction must facilitate opportunities to solve complex problems critically, using higher-order thinking (Gunter, 2007). This requires building an information/data culture within the given post-secondary setting. An organization with an information/data culture is one that values data, utilizes it in its own operations, fosters innovation, and provides the necessary tools and atmosphere for students to engage with material in a meaningful manner (Johnson et al, 2015). In this type of environment, formal and informal learning methods should be utilized, which enforces the idea that students must be able to adapt to new challenges and innovate on-the-fly (Wing, 2008).

Data literacy courses or programs should also be forward-looking, and identify trends, challenges, and technological developments that are likely to occur within the near future (Johnson et al, 2015). This type of learning environment not only facilitates data literacy education, but instills in students intangible skills and work-styles that can be further applied in their careers, so-called ‘meta-skills’ (Liquete, 2012). Meta-skills include adaptability, deep thinking, and being able to critically assess problems (as opposed to simply regurgitating information), and are essential to engaging in the 21st century knowledge economy (Koltay, 2015; Liquete, 2012). These skills are highly-transferable, and applicable to multiple fields and professions (Ontario, 2008).

### *Go Fish: Matching Skills with Requirements*

Data literacy competencies and meta-skills must match up with the positions that Canadian employers need not only now, but in five, and ten years time. In terms of structure, content must go beyond one-off workshops, and instead be focused on ‘data-skilling’, or building up competencies incrementally through courses or modules (Pryor & Donnelly, 2009; Schneider, 2013; Stephenson & Caravello, 2007; Wright et al, 2012). This allows students to focus on specific aspects of data literacy (e.g. finding data, evaluation of data, visualization, manipulation tools, data storage, data ethics, data curation, etc.), and build competencies successively to accomplish a goal (e.g. a data management plan for research) (MIT, 2014; Qin & D’Ignazio, 2010<sup>b</sup>; Stephenson & Caravello, 2007; Wright et al, 2012). It also instills in students a continuous approach to learning. This is integral for professional development in one’s career. Technology and methods for working with data will continue to evolve, and so too must professionals.

In terms of content, post-secondary institutions must cooperate with industry and the private sector to re-assess and re-adjust curriculum in order to maintain relevancy. Skills assessment frameworks in-line with global standards can be also used to ensure that graduates are ready to enter the workforce (Chinien & Boutin, 2011). In the current business and economic climate, the ability to effectively manage, and make decisions based on data is paramount (Giles, 2013; Harris, 2012; Koltay, 2014; McKendrick, 2015).

### *Making the Call: Data Management and Decision Making*

Data management is therefore a foundational aspect of data literacy that is essential for employers and businesses (Pryor & Donnelly, 2009; Qin & D’Ignazio, 2010<sup>a</sup>). Indeed, before one can start effectively utilizing data for decision-making or otherwise, one must be able to actually understand and manage the data they possess. Both the private and public sectors are increasingly turning to digital and Cloud-based information systems, as opposed to traditional paper-based systems. This has led to the creation of large amounts of internal data (e.g. human resources, finances, briefing notes, reports, etc.) that must be managed properly. As Wright et al. posit, data management as a tenant of data literacy recognizes individuals as both consumers and producers of data (2012). Individuals must have core competencies relating to data organization, metadata creation and utilization, and data continuity and reuse (Ibid). Metadata and data continuity are especially important, and entails packaging and ‘labelling’ data in such a way that it can be easily transferred to or used by co-researchers/workers (Ibid). This is critical to increasing cooperating between different branches in an organization, and can reduce silos, and increase efficiency. Decision-making based on data is also critical for the Canadian economy.

Data Driven Decision Making (DDDM) is the ability to effectively transform information into actionable knowledge and practices by collecting, analyzing, and interpreting all types of data



(Koltay, 2014; Mandinach, Parton, Gummer, & Anderson, 2015). At a basic level, it requires the skills to identify problems, frame questions, collect required data, transform data into information, transform information into decisions, and evaluate said decisions (Mandinach et al, 2015). Evaluation, and making necessary changes to chosen courses of action is a key component. Data is not a static entity, and thus neither are decisions based on data. Data usage and evaluation should be continuous, and integrated into existing decision-making structures (Ibid).

As stated, the increased flow of information and financial importance attached to actionable knowledge has made DDDM a key component of modern business (Liquete, 2012). A survey by *The Economist* of 530 senior executives found that 43% believed that data are “extremely important” for strategic decision-making (Giles, 2013, p. 4). However, it can be difficult for companies to engage in DDDM due to data being siloed in different units, and at different levels (Giles, 2013). Therefore, skills and competencies for effective DDDM should be present at all levels within an organization, and across all positions and disciplines (Giles, 2013; Harris, 2012; Vahey et al, 2012). Teaching the skills required for DDDM at the undergraduate level should focus on practical learning.

The Thinking With Data (TWD) project focused on using data to determine whether the water distribution for the Euphrates river among the surrounding countries could be considered equitable (Vahey et al, 2012). Students were tasked to collect data from multiple sources, form coherent arguments, and put forth data-based conclusions. A key component of DDDM, and indeed data literacy, is the ability to recognize faulty information and create “valid, data-based arguments” (Vahey et al, 2012, p. 183). Connecting data to real-world issues (e.g. sustainability) allows students to connect procedure to practice, and encourages curiosity and independent thinking (Wyner, 2013). The same data literacy abilities required to effectively carry out DDDM can also improve entrepreneurial innovation, creativity, and efficiency (Boyles, 2012). DDDM can also be applied to data-driven marketing. A survey of 300 top-level executives by *Forbes* identified increased customer engagement and growth from companies who were considered to be ‘leaders’ in data driven marketing and decision-making (McKendrick, 2015).

### *Barriers and Challenges*

Many of these barriers to data literacy in employers and the economy relate to data literacy education within post-secondary institutions, which will be covered at a later point in the knowledge synthesis. At a high-level, there are ethical considerations that both educators and the private sector must be aware of in terms of data collection and usage (Manyika et al, 2011). The technology and applications for data are continuing to evolve at an incredible rate, and legislation concerning privacy has already fallen behind. Individuals and companies must be aware not only how they are using data, but why.

Speaking to technology, to “be merely in possession of technical infrastructure (hardware and software) is by no means sufficient to provide a comparative advantage and to become competitive and succeed in the digital economy.” (Chinien & Boutin, 2011, p. 12). During their post-secondary studies, students must be taught effectively how to use technological tools, and this is even more important for private sector organizations. Moreover, employers themselves must believe in the value and worth that technology, and indeed, data literacy brings to the table. Without the proper buy-in, the full potential of data usage in the workplace will be unfulfilled, and any skills brought forth by employees effectively wasted.

## Canadian Universities and Graduates

Canadian universities are being challenged to produce globally competitive graduates, due to the growing demand for a more skilled workforce able to work efficiently and effectively in a knowledge-based economy. However, many graduates are not at the level of proficiency required to work with data and data-based applications (Boyles, 2012). Being competitive on a global scale is a challenge for all countries, due to societal and economic factors, but there are steps that can be taken that will encourage development. An effective avenue to encourage such development is to include data literacy education as a key component of national standards, as the United States has done (Zalles, 2005). This would incentivize the education system to develop these skills more diversely. Human Resources and Skills Development Canada (now Employment and Social Development Canada) conducted a study in 2011 in response to the G20 Summit based on international and national standards, of which they identified important skills, including data literacy, and skills related to this, such as ICT, digital, and computational literacies (Chinien & Boutin, 2011). The Government of Canada recognizes data literacy as a growing issue, but has not yet acted.

This unfulfilled need on the governmental side may be due to a perpetual struggle to identify objectives and strategies to develop globally competitive skills. International standards are helpful in identifying how competitive Canadian university graduates are. The Organization for Economic Co-Operation and Development (OECD) frequently influences policy internationally, providing a benchmark for competency. The Skills Outlook that they produced in 2013 identified only a third of the population as having basic data analysis skills, with only 12.5% above basic skillset. In response to this study, Romani addresses this mismatch of skills being taught and in-demand, recommending that initiatives should be addressed regularly to ensure goals and practices align with the global knowledge-based economy (2009). Although the inaction of the government is problematic, it does give Canadian academia the opportunity to act and create programs and data literacy standards that can be implemented.

### *Team Effort: Collaborating to Deliver Data Literacy Education*

Another way to ensure globally competitive graduates includes collaboration between several groups such as educators, organizations, and stakeholders. These groups are integral in providing a targeted education to students with usable skills in the workforce. Educators collaborating with others may seem obvious. However, communication is key to a well rounded education. If assumptions are made, it can lead to gaps or unnecessary overlaps in education. Educators keeping communication open between subjects and levels are crucial to a systematic and comprehensive data literacy education, and will strengthen understanding though being taught across curriculum effectively (Zalles, 2005; and Ontario Ministry of Education, 2008).

Organizational collaboration also provides academia the opportunity to diversify learning. The different perspectives and resources that non-academic organizations can bring to education are valuable for the institution as well as the students. These can include partnerships in curriculum planning, student professional development, as well as extracurricular programs (Gold, 2007). As mentioned above, academic institutions also have the ability to consult with national and international standards such as OECD and HRSD. The Association of College Research Libraries (ACRL) has worked with industry leaders to determine what requirements must be met by standards to ensure accurate and relevant standards (Schwieder, Fielder, & LLIC, 2014).

Additionally, the Oceans of Data Institute works in partnership with organizations to provide users with targeted skills depending on level and need (2014).

Lastly, collaboration with stakeholders is essential for successful graduates. There must be as little misunderstanding as possible between what is being taught and what is needed in the workplace. In the field of data literacy there is a wide range of stakeholders: schools, practitioners, professional development providers, provincial education departments, government, professional organizations, and more. These groups should be involved in the curriculum process planning and implementation to ensure that education is comprehensive (Mandinach & Gummer, 2013). Collaboration with stakeholders provides perspective on the realities of the workplace such as disciplinary culture and local practices, as well as initiatives and tools (Romani, 2009; Giles, 2013; Benito, 2009; and Mooney & Carlson, 2014).

### *Barriers and Challenges*

Teaching data literacy at the undergraduate level is often left out of curricula for social sciences completely (Scheitle, 2006). This creates large gaps between educational experiences for students entering the postgraduate level of study and the workforce (Swan & Brown, 2008).

Data literacy being taught at the commencement of post-secondary education would benefit students to more easily integrate into their remaining discipline-specific education (Shorish, 2015; and Sapp Nelson, Zilinski, & Van Epps, 2014). This early engagement, before they have specialized, will allow peers to work together on common problems, at similar skill levels (Swan, et al., 2009).

Technical skills are difficult to learn, if the student is inserted into the middle of the lesson; starting from the beginning is especially important in learning technical skills. Some students may flourish, but most are likely to feel defeated without prior knowledge to guide them. This is why building a foundational knowledge of a skill is a very important part of the process (Littlejohn Shinder, 2013). This review has identified different or synthesized levels of data skills, which can be found in Appendix 1. This can assist in targeting learning to appropriate levels, or developing standards to ensure consistent education opportunities.

It is difficult to begin with the basics with a professional audience with varying skill levels, and as argued above, the earlier the education the better, as cross-disciplinary education maximizes the impact and applicability to various situations (Gunter, 2007; Erwin, 2015; and Johnson, & Jeffryes, 2014). Data literacy is increasingly necessary in throughout all levels of society and industry, and a natural extension of this is providing it to social sciences, humanities, and arts and culture educational curriculum as well as natural science and business (Koltay, 2014; and Maycotte, 2014). There are barriers to teaching data literacy at an interdisciplinary level, such as lack of depth (Johnson & Jeffryes, 2014), but the essential skills of data literacy are similar to computational thinking in some respects, and are more focused on the mental process to solve a problem, which is a generally useful skill, rather than real technical skills (Czerkawski & Lyman, 2015).

### **Best Practices for Teaching Data Literacy**

In this section, we synthesize documented best practices for data literacy education and instructions, including the timing and mechanisms of delivering data literacy content. These are all presented in light of the context and strategies for data literacy education described previously.

### **Appropriate Timing of Data Literacy Education**

Delivery of data literacy education has been recommended at several educational levels/grades in the research, and range from elementary school to Master programs. Erwin (2015) and Vahey et al., (2012) each recommend beginning with elementary aged students, which offers opportunities to standardize across grades within existing well-defined curricula. Romani (2009) and the Ontario Ministry of Education (2008) suggest that secondary education should include data literacy in complementary subjects and using technology to help students recognize the transferable nature of this skill. Any approach to data literacy at the post-secondary level will require awareness and adaptation as standards evolve at earlier levels of education. This synthesis deliberately focuses on post-secondary education, and does not assume the status quo will change at the elementary and secondary levels.

Data literacy education is increasingly being offered at the post-graduate levels, providing students with the skills and tools to deal with the 'Big Data question' (ACRL, 2014). However, teaching data literacy at this level only affects a small number of students, as opposed to the undergraduate level. Moreover, it is often more difficult to instill foundational knowledge at the post-graduate level. Qin and D'Ignazio found in their feedback from masters level students that lack of background knowledge makes data literacy jargon and exercises difficult to master; especially when there are varying skill levels in the group (2010<sup>b</sup>). In terms of usability, Womack further argues that the skills learned at a post-graduate level are very specific and discipline-focused. Additionally, teaching data literacy education early in undergraduate programs could instill good practices and improve their general study (2014). The effort institutions are making to include data into the curriculum is in the right direction, but the mark is too far. Students must be educated in a meaningful way, not simply to fill a void.

### **Delivery: From Courses to Workshops**

Many authors believe that data literacy education should be effected through a stand-alone class (Burdette & McLoughlin, 2010; Martin & Leger-Hornby, 2012; Qin & D'Ignazio, 2010<sup>b</sup>; and Swan, et al., 2009), generally due to the essential nature of the skill. This solution is not necessarily ideal, as current curricula are already full with required courses (Teal, et al., 2015; and Swan & Brown, 2008), and that a one-size-fits-all class may not account for the varying backgrounds and skills students have in their first few years of university education. Some researchers recommend in-class delivery be supplemented with workshops to bridge the gaps and provide targeted help (Carlson, Johnson, Westra & Nichols, 2013; and Swan & Brown, 2008). Others suggest supplementing in-class education with online courses, providing specific assistance when needed (Gray, 2004, Hattwig, et al., 2013), or to prepare students before entering an in-class course (MacMillan, 2010), but this idea did not receive broad acknowledgement or validation.

Workshops are mentioned often in the literature in reference to bridging the existing skills gap, or bridging the gap between knowledge and practice through more hands-on, active learning opportunity (MacMillan, 2010). Workshops are often short and intense, and for this reason are usually targeted to a certain skill, level, and/or domain, but generally effective (Teal, Cranston, Lapp, White, Wilson, Ram, and Pawlik, (2015). The most generally accepted delivery method is the modular-based system. This can be integrated into any formal or informal delivery method (Martin & Leger-Hornby, 2012; Qin & D'Ignazio, 2010<sup>b</sup>; Prado & Marzal, 2013; and Schneider, 2013), and provides a solid basis for targeted learning for individuals (Shorish, 2015; and MacMillan, 2014).

## Integration of Data Literacy into Curricula

As mentioned above, a barrier to data literacy courses includes a full curriculum. This has been recognized by many researchers, and countered with recommendations to integrate it into complementary subjects. Shorish recommends a generally useful class such as research methods, allowing for a fluid environment to apply skills (2015). Many other authors recommend integrating data literacy into other literacy education, because of the overlap in competencies. The most popular match is data information literacy (Carlson, Fosmire, Miller, & Sapp Nelson, 2011; MacMillan, 2010; Stephenson & Caravello, 2007; Hunt, 2004; Bresnahan & Johnson, 2014; Schneider, 2013). Changes and targeted design to incorporate explicit data literacy education could allow these two subjects to meet this growing need.

Other literacy education that could include data literacy includes visual literacy, due to the importance of critically analyzing other's interpretation of data as accurate, and representing conclusions correctly (Womack, 2014; Hattwig, Bussert, Medaille, & Burgess, 2013). Statistical information literacy also has strong connections with data literacy (Prado & Mazal, 2013; Twindale, 2013; and Schield, 2004). Statistical literacy is especially relevant due to the focus on the ability to judge quality and utility of data, and should at least serve as the foundation of teaching data literacy (Gray, 2004). Science literacy also shares similarities such as methods, approaches, attitudes and skills relating to critical thinking that are necessary for data literacy (Koltay, 2014; Shorish, 2015; Qin & D'Ignazio, 2010<sup>a</sup>).

Wanner reviews several research studies investigating candidate subjects into which data literacy could be integrated, including Information and Statistical Literacy, Data Information Literacy, Science Data Literacy, Visual Literacy and Geospatial Data (2015). The integration of these literacies, and the various competencies they share throughout, has been dubbed "Transliteracy" or "Metaliteracy" (Koltay, 2015; Dean, 2015). Metaliteracy is the ability to use multiple literacies in a multimedia layout and navigate through multiple domains (Frau-Meigs, 2012; and Ipri, 2010). This type of integration and understanding promotes critical thinking and collaboration (Mackey & Jacobson, 2011). The challenge is integrating a new set of competencies into already full courses could easily shortchange students on both forms of literacy, and as discussed in the next paragraph, the people well-suited to teach information literacy may not be good candidates for teaching data literacy. We thus view this approach as a temporary measure designed to mitigate the data literacy education gap in the short term. However, incorporation of data literacy into domain-specific courses may be effective. Data literacy competencies have overlapping skills with other practices, and can be integrated into complementary subjects allowing students to build on strengths and experiences of different disciplines (Swan & Brown, 2008).

Another barrier involved in data literacy integration is that instructors do not feel sufficiently confident in their own ability to teach data literacy competencies (ACRL, 2014; Carlson, Fosmire, Mandinach & Gummer, 2013; Johnson, Adams Becker, Estrada, and Freeman, 2015; Miller, & Nelson, 2011; Romani, 2009; Wanner, 2015). Integrating data literacy with other literacy training addresses the issue of finding space in the curriculum, but those instructors may not be prepared to teach data literacy competencies. In this scenario, librarians (particularly data librarians) are useful resources to bridge the knowledge gap for students and faculty alike (Koltay, 2014; and Koltay, 2015). There are many ways this can be achieved, including collaboration with subject specialists inside and outside of class (Schield, 2004; Shorish, 2015; Hogenboom, Holler Phillips, & Hensley,

2011; and Swan & Brown, 2008). Many of these literacies are within the scope of academic librarians, who are already on campus and can help design and carry out labs (MacMillen, 2014; and Hunt, 2004).

### **Emerging Teaching Approaches and Learning Environments**

Educators must provide students with an environment conducive to learning. This includes practices that are non-traditional but already employed at post-secondary institutions, like 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). One way to encourage ongoing learning is including digital devices and technologies, such as social media technologies, consumer technologies, learning technologies, and so on. Adopting these new teaching and learning paradigms is proven to improve the learning process and workforce preparedness (Romani, 2009; Frau-Meigs, 2012; Mackey & Jacobson, 2011; and Johnson, Adams Becker, Estrada, & Freeman, 2015). The flipped-classroom approach to content delivery (where instead of lecture in class and apply at home, you learn theory at home and apply it in class) allows students to recognize the relevance of the theory they are learning and identifying their own knowledge deficiencies, making the learning practical, not simply theoretical (Swan et al., 2009).

### **Engaging content with real world data to foster innovation**

Practical, hands-on learning is an important part of engaging students; universities must foster conditions for innovation to happen (Johnson, et al., 2015; Pentland, 2013; and McAuley, Rahemtulla, Goulding, & Souch, 2010). Real-world data provides students with the opportunity with diverse experiences and caters to a wide range of skills (Carlson, Johnson, Westra, & Nichols, 2013; Romani, 2009; and Davenport & Kim, 2013). This type of learning experience encourages students to find solutions, because they are contributing to the larger community, not simply a grade, promoting task commitment, which is essential for learning technical skills successfully (Erwin, 2015; Vahey, et al., 2012; Burdette & McLoughlin, 2010; and Scheitle, 2006). Real-world data also allows students to make connections to their own impact on society, enhancing the opportunity to provide solutions that are targeted to specific needs (Wyner, 2013; and Hu, 2012).

### **Successive/Iterative, Practical, Hands-on Learning**

Hunt goes further in explaining that students learn best when the data literacy curriculum is relevant and builds on previously learned skills and knowledge (2004). This is especially effective when imparting technical skills (Burdette & McLoughlin, 2010; and Yeh, Xie, & Ke, 2011). 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.

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). Pairing this with using real-world data has proven to be an effective way of integrating many of the previously mentioned approaches: complementary skills, iterative, practical, and engages students through the use of real-world data. PBL can connect multiple areas of curriculum and relate to personal experiences (Erwin, 2015), and provides students with the knowledge of processes that occur in the workplace, as well as seeing it to completion. PBL is useful when teaching students critical thinking and problem solving (P21, 2012), which is useful because data manipulation requires creative

thinking. Lastly, an important aspect of PBL is choice (Burdette & McLoughlin, 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.

In addition to the iterative learning process, it is important to connect procedure and practice. Striking a balance between these two concepts can be difficult. Students' ability to effectively use data in diverse situations, and having the ability to distinguish how to proceed appropriately is increasingly important in the competitiveness of Canadian graduates (Hu, 2012; Liquete, 2012; and Wyner, 2013). Practicing these skills is key to the learning experience. Using real data and tools to analyze it helps bridge the gap between learning facts and acquiring inquiry skills (Swan, et al., 2009). 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.

### Assessment and Evaluation

Assessment and Evaluation (A&E) occurs in two dimensions: A&E of data literacy education itself, and A&E of students engaged in data literacy learning.

As established earlier, students' levels of skill at commencement of courses and programs today are varied and inconsistent. For this reason it is important for pre tests/surveys to be conducted (Swan, et al., 2009; Jones, Ramanau, Cross, & Healing, 2009), to ensure that the education is understandable and appropriate for everyone, and allows for individual targeted help, or changes to the pace of the instruction (Reeves & Honig, 2015; and Qin & D'Ignazio, 2010<sup>b</sup>). Post test/surveys can then provide feedback for improvements for future design or application (Qin & D'Ignazio, 2010<sup>b</sup>; Reeves & Honig, 2015), ranging from tools, to instruction, to subject matter. A data literacy self assessment tool, informed by a need-driven competencies matrix like the one in Appendix 1, will help track success at imparting necessary knowledge and skills to students.

Assessment and evaluation of students is essential in the education process. Without validation, the instructor cannot know whether the instruction was effective or useful. The challenge arises in the method of conducting these for data literacy and twenty-first century skills. Swan and Brown argue that formal skills assessment is not as favored as practical assessment (2008), probably due to data literacy being a skill, not simply knowledge. Liquete recognizes that assessment must encompass more than just information, content, and results, but evaluation of the entire process (2012). Hattwig, Bussert, Medaille, and Burgess argue iterative assessment is the best way to evaluate data literacy (2013), ensuring that all aspects of the skill are understood, and not just parts of the whole. Chinien and Boutin recommend qualitative and quantitative analysis be joined through scenario-based testing, which measures cognitive and technical skills (2011).

There are several resources available for consultation that can be integrated into data literacy assessment. As mentioned above, Information and Visual literacies have many overlapping competencies; the ALA's *ACRL Information Literacy Competency Standards for Higher Education* (2000) and *ACRL Visual Literacy Competency Standards for Higher Education* (2011) include performance indicators that can be incorporated in data literacy assessment and evaluation. The OECD provides educators with a description of proficiency levels for task centred assessment, as well as highlights

the usefulness of the Programme for International Student Assessment (PISA) as an assessment resource (2013). The Council for Aid to Education is an American service that provides assessment for a range of education levels. Their CLA+ program assesses 21st century skill use such as critical thinking, problem solving, and reasoning using performance tasks (2015).

Overall, post-secondary education has the opportunity to enact change in 21<sup>st</sup> century society. This will help ensure graduates are prepared to be productive citizens and employees. Standards for education are always changing, and it is difficult to decide what is a priority in any given year. We assert that data skills and knowledge are vital, and should be addressed in the near term to ensure Canada continues to produce students prepared to be global citizens and leaders.

## Additional Resources

**Institute:** Web Science Conference

**Name:** First Data Literacy Workshop

**Type:** One-off workshop

**Description:** The First Data Literacy Workshop occurred on June 30, 2015, at the WebScience conference held at Oxford University, in Oxford, United Kingdom. The workshop investigated the potential for multidisciplinary research into what data literacy means for society as a whole, why it matters, and how it might be facilitated. The workshop included a panel of speakers, the presentation of four papers on topics such as data visualization, urban data schools, and others, and a full workshop discussion focused on creating a data agenda for future cooperation and work.

**Link:** Data Literacy. (2015). 1st DL Workshop. Retrieved from

<http://www.dataliteracy.eita.org.br/1st-dl-workshop/>

**Institute:** University of Toronto, iSchool

**Name:** INF 2115H Data Librarianship

**Type:** Masters-level Course

**Description:** The INF 2115H Data Librarianship course taught at the University of Toronto's iSchool (as part of its MLIS program) focuses on topics related to the acquisition, management, and retrieval of numerical statistics and data. Topics covered include public, private and academic sector data gathering, statistical production and dissemination, warehousing and management, repositories and consortia, data extraction and manipulation, and privacy issues. The course includes conceptual (e.g. readings) and practical elements (e.g. hands-on assignments), and requires that student become competent with data analysis tools Beyond 20/20 and SPSS.

**Link:** <http://mccaffrey.ischool.utoronto.ca/2115/syllabus.pdf>

**Institute:** Massachusetts Institute of Technology (MIT)

**Name:** Tackling the Challenges of Big Data

**Type:** Online Course

**Description:** This six week online continuing education course offered at MIT is aimed at educating professionals on major technologies and applications that are driving the "Big Data revolution", with the intent of imparting practical skills and knowledge. The course consists of five modules covering 18 topic areas, with 20 hours of pre-recorded lecture sessions. Each module has a corresponding assignment and related case studies. Topics covered by the course include data collection (e.g. smartphones, sensors, the Web), data storage and processing (e.g. relational databases, Hadoop, etc.), extracting structured data from unstructured data, systems issues,



analytics (machine learning, data compression, efficient algorithms), and visualization. The course is taught by The course is taught by a team from MIT, and the Computer Science and Artificial Intelligence Laboratory (CSAIL). The price of the course is \$545, and students are awarded a certificate upon successful completion of the program.

**Link:** <https://mitprofessionalx.mit.edu/courses/course-v1:MITProfessionalX+6.BDx+5T2015/about>

**Institute:** Open Data Institute

**Name:** Various

**Type:** In-person Workshops

**Description:** The Open Data Institute in London, United Kingdom offers a variety of practical courses on open data (e.g. 'Open Data for Smart Cities', 'Open Data in a Day for Government', 'Managing Risk with Open Data', etc.). Courses range from one to six days in length, and require varying levels of foundational expertise (i.e. beginner, or experienced information professional, analysts, intelligence personnel, etc.). Courses offer a mix of theoretical class-based learning, as well as hands-on exercises.

**Link:** <http://opendatainstitute.org/courses>

**Institute:** School of Data

**Name:** Various

**Type:** Online Course

**Description:** The School of Data offers free online independent courses. Courses include 'Data Fundamentals', 'A Gentle Introduction to Cleaning Data', 'Working with Budgets and Spending Data', and others. Courses are structured into modules, and include theoretical knowledge as well as hands-on exercises. The view of the School is that data should be open and useable to everyone, and thus "works to empower civil society organizations, journalists and citizens with the skills they need to use data effectively in their efforts to create more equitable and effective societies".

**Link:** <http://schoolofdata.org/>

**Institute:** University of Delaware

**Name:** Analytics: Optimizing Big Data Certificate

**Type:** In-person Course

**Description:** This continuing education course offered by the University of Delaware aims to teach students how to gather, organize, and effectively analyze large datasets in order to make informed business decisions. Students are also taught how to communicate their analyses in a clear and concise manner. The course is taught over a half-semester, and consists of four main modules: Analytics Basics, Big Data Tools, Process Control and Capability, and an Individual Case Study Project. Modules are taught at the BA level, but open to all-comers. The course costs \$2,795 and students are awarded a certificate upon completion.

**Link:** <http://www.pcs.udel.edu/data/>

**Institute:** Council for Aid to Education

**Name:** CLA+

**Type:** In-class or online assessment

**Description:** The College Learning Assessment (CLA+) evaluates skills that are essential for the workplace and life outside of the classroom, and measures critical thinking skills, problem-solving, scientific and quantitative reasoning, critical reasoning and evaluation, and critiquing an argument.

The assessment consists of two parts. The first is a Performance Task that involves a multi-step problem requiring students to analyze a real-world scenario and design a viable, yet creative solution. The second component consists of questions aimed at measuring qualitative and scientific reasoning, critical reading and evaluation, and the ability to recognize logical fallacies. The assessment can be carried out in-class at a designated university (if it is a requirement of said university), or online. Upon completion, students receive CLA+ percentile ranking and mastery level, and if they scored high enough, a digital badge to be used on LinkedIn.

**Link:** <http://cae.org/>

**Institute:** Data Science Central

**Name:** Data Science Webinars

**Type:** Online Webinar

**Description:** Data Science Central acts as an online community hub for data scientists and practitioners. The site hosts a number of video webinars created by community members that can be viewed for free. Length of webinars range from a few minutes to multiple hours, with topics such as data mining, deriving analytic insights, data visualization, predictive data modelling, and training for data analysis tools (e.g. Hadoop). Webinars are aimed at both basic and advanced users.

**Link:** <http://www.datasciencecentral.com/>

**Institute:** NASA Earth Data

**Name:** Data Discovery Tools

**Type:** Data Analysis/Visualization Tools

**Description:** As part of NASA's Earth Science Data Systems Program, the Earth Observing System Data and Information System (EOSDIS) provides a number of advanced data discovery tools that can be used by users to carry out search, analysis, and visualization of NASA earth science data. Tools are loosely classified into the following categories: Search and Order Tools, Data Handling (Read/Ingest, Format Conversion, Data Manipulation), Subsetting and Filtering Tools (Temporal, Spatial, Parameter, Channel), Geolocation, Reprojection, and Mapping Tools, and Data Visualization & Analysis Tools.

**Link:** <https://earthdata.nasa.gov/earth-observation-data/tools>

**Institute(s):** University of Massachusetts Medical School, George C. Gordon Library, Worcester Polytechnic Institute

**Name:** Frameworks for Data Management Curriculum

**Type:** Curriculum Document

**Description:** Consists of a framework for a data management curriculum and corresponding course plan data management instruction aimed at undergraduate and graduate students in the science, health science, and engineering disciplines. The proposed course consists of seven modules, and four case studies in order to put theory into practice.

**Link:** [http://library.umassmed.edu/data\\_management\\_frameworks.pdf](http://library.umassmed.edu/data_management_frameworks.pdf)

**Institute(s):** SRI International, National Science Foundation, Research Center for Educational Technology

**Name:** Thinking with Data Project

**Type:** In-person course/workshop

**Description:** Thinking With Data (TWD) consists of four, two week modules, respectively in social

studies, mathematics, science, and English language arts. They “address issues of data representation, common measure, and proportional reasoning, using real data accessed from real world media sources in discipline-specific, problem-solving contexts and align with relevant subject area standards” (2). This requires students to formulate and answer data-based questions; use appropriate data, tools, and representations; and develop and evaluate data based inferences based on world water issues. This relates to the approach *Preparation for Future Learning* that highlights structure, internalizing key dimensions and applying it in a variety of contexts.

Link: <https://www.sri.com/work/projects/thinking-with-data>

## Further Research/Research Gaps

We have identified several areas where we hoped to find literature, either formal or informal, but could not find anything that sufficiently addressed the questions we had.

### Geospatial/Temporal Data Management

Data management is a recurring competency throughout the existing literature. However, there is little (general) data literacy-related material that focuses on geospatial/temporal data management and operability (e.g. managing real-time tracking data using GIS and other software systems). This type of data management currently constitutes a ‘niche’ aspect of data literacy. However, the continuing proliferation and importance of open data and GIS applications is likely to raise the importance of knowledge associated with working with this type of data.

Many universities possess active GIS programs or centres. The bases of knowledge that these centres and programs provide could serve as foundation for joint partnerships and further research into how geospatial and temporal data management should be incorporated into data literacy learning.

### Data Literacy Requirements informed by Employer Need

The reviewed literature lists many data literacy competencies and skills that could be taught at the postsecondary level. However, with the exception of a few sources, these competencies are exclusively written and compiled by academics and/or educational organizations. There are few sources from specific industries (e.g. business, public sector, etc.) that list data literacy-related skills required for prospective jobs and positions.

Further research into this area (e.g. a survey of industry professionals from different sectors) may reveal connections (or possible disconnects) between what academia believe should be taught, and what industry employers are actually looking for when hiring new graduates.

### Lack of Agreed Upon Standards and Best Practices

There are currently no agreed upon universal standards or best practices for teaching data literacy at the postsecondary level. Methods of assessment also vary based on delivery type (e.g. course, workshop, etc.), and content. Authors stress the importance of the creation of a universal standard for data literacy competencies akin to the Association of College and Research Libraries (ACRL)’s Information Literacy Competency Standards for Higher Education.

Cooperation with other universities, associations, and government could potentially lead to the creation of agreed upon standards and best practices that could facilitate effective and widespread data literacy education.

## Data Security

One need only look at the Edward Snowden case or any such number of news stories pertaining to lost hard-drives containing sensitive information to recognize the importance of data security. The same technology and tools that make data so much easier to connect, analyze, and share can also be used to breach systems to steal or corrupt data. Data security extends not only to the digital and online world, but also how to handle sensitive data in physical form (e.g. external hard drives, USB keys, etc.). Current focus on data security leans toward the computer science discipline, with little thought given to how this should be taught to more general data literacy audiences (i.e. undergraduate students).

## Data Ethics

Connected to the issue of data security, is that of data ethics. The technology surrounding data continues to advance at an incredible rate, and there will be uses, issues, and implications for data that we cannot currently imagine or envision (and that will likely outpace current legislation). Moreover, as Pentland puts it when describing Big Data, “the ability to track, predict and even control the behavior of individuals and groups of people is a classic example of Promethean fire: it can be used for good or ill” (Pentland, 2013, p. 1). Although some authors put emphasis on data ethics, the majority either gloss over or fail to make note of it. However, in order for students to understand and critically think about the larger issues regarding data literacy, thus must have an understanding and awareness of the ethics surrounding data.

Further investigation on how to effectively integrate and teach data ethics will ensure that students are aware and able to think critically on current, and yet to emerge issues and challenges related to data literacy.

## Data Literacy for the Existing Workforce

Data literacy is a desperately needed skill in both society and economy today, but teaching the mass amount of people who require it is a challenge, and merits further study. What is the best method of facilitating a large group of learners? Massive Online Open Courses (MOOCs) are popular right now, but how can this most effectively be provided to the public? Who should be responsible for providing this training, and for paying for it?

## Knowledge Mobilization

The importance of data literacy education, and best practices and strategies for data literacy teaching and learning, should and will be communicated to key audiences. We have launched, and in the coming months will populate, [www.dataliteracy.ca](http://www.dataliteracy.ca), a website aimed at multiple audiences and sharing the results of this knowledge synthesis. We are in the process of authoring an op-ed piece for submission to the Globe and Mail, presenting these results to the public and advocating for national engagement in data literacy teaching. We will disseminate the results to government and NGO stakeholders at the Imagining Canada's Future Forum in November 2015 in Ottawa, and to academic audiences via journal articles and conference presentations over the next several weeks. Preliminary results have already been presented at the Atlantic Universities Teaching Showcase.

We also plan to showcase this work at the 19<sup>th</sup> Annual Dalhousie Conference on University Teaching and Learning (DCUTL), to be held in April 2016. It includes as many as 200 participants from across Atlantic Canada and beyond. The 2016 DCUTL conference theme will explore the key outcomes of a 21<sup>st</sup> Century Curriculum. We will, with Centre for Teaching and Learning (CLT) support, organize a

DCUTL conference stream dedicated to data literacy across the curriculum, featuring research papers, panel presentations, and workshops from members of the project team, as well as other leaders in the area of data literacy in higher education.

We have applied for academic innovation funds at Dalhousie University to support a project to develop modular curriculum materials based on what we learned when preparing this synthesis, and to empirically validate some of the theories presented in the literature. This course would be offered to first- or second-year students across Dalhousie University, and would follow the best practices identified here. The teaching resources created will be made available on [dataliteracy.ca](http://dataliteracy.ca). Additionally, three of the investigators are working with Dalhousie's Executive Education group to address the data skills gap at the management level of local businesses.

Finally, we intend to collaboratively author a Green Guide for the general audience of higher education educators interested in teaching data literacy in higher education. This book series, published by the Society for Teaching and Learning in Higher Education, features introductory guides to key teaching and learning topics that are accessible to faculty and instructors across the disciplines (see <http://www.stlhe.ca/publications/green-guides/>).

## References

- ALA. (2000). *ACRL Information Literacy Competency Standards for Higher Education*. Retrieved from <http://www.ala.org/acrl/standards/informationliteracycompetency>
- LA. (2011). *ACRL Visual Literacy Competency Standards for Higher Education*. Retrieved from <http://www.ala.org/acrl/standards/visualliteracy>
- Association of College and Research Libraries. (2013). Working Group on Intersections of Scholarly Communication and Information Literacy. *Intersections of Scholarly Communication and Information Literacy: Creating Strategic Collaborations for a Changing Academic Environment*. Chicago, IL: Association of College and Research Libraries. Retrieved from <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Intersections.pdf>
- Benito, E.M. (2009). Learning.com joins the Partnership for 21st century skills. *Business Wire*. Retrieved from [https://global.factiva.com/ha/default.aspx#!?&\\_suid=143231383276103258974098134786](https://global.factiva.com/ha/default.aspx#!?&_suid=143231383276103258974098134786)
- Boyles, T. (2012). 21st century knowledge, skills, and abilities and entrepreneurial competencies: A model for undergraduate entrepreneurship education. *Journal of Entrepreneurship Education*, 15, 41-55.
- Bresnahan, M., and Johnson, A. (2014). Building a Unified Data and Information Literacy Program: A Collaborative Approach to Instruction. [PowerPoint]. Portland, OR: University of Colorado Boulder. [doi.org/10.6084/M9.FIGSHARE.1119760](https://doi.org/10.6084/M9.FIGSHARE.1119760)
- Burdette, A.M., & McLoughlin, K. (2010). Using census data in the classroom to increase quantitative literacy and promote critical sociological thinking. *Teaching Sociology*, 38(3), 247-257.
- Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an approach for data management education: A report from the data information literacy project. *The International Journal of Digital Curation*, 8(1), 204-217. doi:10.2218/ijdc.v8i1.254
- Carlson, J., Fosmire, M., Miller, C.C., and Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *Libraries and the Academy*, 11(2), 629-657. DOI: 10.1353/pla.2011.0022
- Chinien, C., & Boutin, F. (2011). Defining essential digital skills in the Canadian workplace: Final report. Retrieved from [http://www.nald.ca/library/research/digi\\_es\\_can\\_workplace/digi\\_es\\_can\\_workplace.pdf](http://www.nald.ca/library/research/digi_es_can_workplace/digi_es_can_workplace.pdf)
- Council for Aid to Education. (2015). CLA+ Assessment. CAE. Retrieved from <http://cae.org/>
- Cowan, D., Alencar, P., & McGarry, F. (2014). Perspectives on open data: Issues and opportunities. *IEEE*, IEEE Conference on Software Science, Technology and Engineering. DOI: 10.1109/SWSTE.2014.18
- Czerkawski, B., & Lyman III, E. (2015). Exploring issues about computational thinking in higher education. *TechTrends*, 59(2), 57-65.
- Davenport, T., & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. Retrieved from <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1>

- Dean. (2015). Data Literacy. *HLWiki International*. [Wiki]. Retrieved from [http://hlwiki.slais.ubc.ca/index.php/Data\\_literacy](http://hlwiki.slais.ubc.ca/index.php/Data_literacy)
- Doucette, L., & Fyfe, B. (2013). Drowning in research data: Addressing data management literacy of graduate students. *ACRL*. Retrieved from [http://www.ala.org/acrl/sites/ala.org/acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe\\_Drowning.pdf](http://www.ala.org/acrl/sites/ala.org/acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf)
- Erwin, R. (2015). Data literacy: Real-world learning through problem-solving with data sets. *American Secondary Education*, 43(2), 18-26.
- Frau-Meigs, D. (2012). Transliteracy as the new research horizon for media and information literacy. *Media Studies*, 3(6), 14-27
- Giles, J. (2013). Fostering a data-driven culture. *The Economist*, 1-22. Retrieved from [https://www.tableau.com/sites/default/files/whitepapers/tableau\\_dataculture\\_130219.pdf](https://www.tableau.com/sites/default/files/whitepapers/tableau_dataculture_130219.pdf)
- Government of Canada. (2013). "Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada" [consultation document]. Available: [http://www.sshrc-crsh.gc.ca/about-au\\_sujet/publications/digital\\_scholarship\\_consultation\\_e.pdf](http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf)
- Gray, A.S. (2004). Data and statistical literacy for librarians. *IASSIST Quarterly*, 28(2), 24-29. Retrieved from <http://www.iassistdata.org/content/data-and-statistical-literacy-librarians>
- Green Button. (2015). *Ontario's Green Button Initiative*. Retrieved from <http://greenbuttondata.ca/>
- Gunter, G. A. (2007). Building student data literacy: An essential critical-thinking skill for the 21st century. *MultiMedia & Internet@Schools [H.W.Wilson - EDUC]*, 14(3), 24-28.
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/3316/2764>
- Hattwig, D., Bussert, K., Medaille, A., and Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *Libraries and Academy*, 13(1), 61-89. doi:10.1353/pla.2013.0008
- Harris, J. (2012). Data is useless without the skills to analyze it. *Harvard Business Review*. Retrieved from <https://hbr.org/2012/09/data-is-useless-without-the-skills>
- Hogenboom, K., Holler Phillips, C.M., and Hensley, M. (2011). Show me the data! Partnering with instructors to teach data literacy. *ACRL 2011*, Philadelphia, Pennsylvania. 410-417.
- Hu, X. (2012). Session summary: The RDAP12 panel on training data management practitioners. *Bulletin of the American Society for Information Science and Technology*, 38(5), 29-30. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/bult.2012.1720380509/epdf>
- Hunt, K. (2004). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly* 28(2), 12 -16. Retrieved from [http://www.iassistdata.org/sites/default/files/iq/iqvol282\\_3hunt.pdf](http://www.iassistdata.org/sites/default/files/iq/iqvol282_3hunt.pdf)

- Ikemoto, G., and Marsh, J. (2008). Chapter 5: Cutting through the data-driven mantra. *Different conceptions of data-driven decision making*, 106(1), 105-131. RAND Corporation. Retrieved from [http://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND\\_RP1372.pdf](http://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND_RP1372.pdf)
- Ipri, T. (2010). Introducing Transliteracy: What does it mean to academic libraries? *College and Research Libraries News*, 71(10), 532-567. Retrieved from <http://crln.acrl.org/content/71/10/532.full>
- Johnson, C. (2012). *The Information Diet*. Sebastopol, CA: O'Reilly Media, Inc.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). NMC Horizon Report: 2015 Higher Education Edition. Austin, Texas: The New Media Consortium. Retrieved from <http://www.nmc.org/publication/nmc-horizon-report-2015-higher-education-edition/>
- Johnston, L., and Jeffryes, J. (2014). Steal this idea: A library instructor's' guide to educating students in data management skills. *C&RL News*, 75(8), 431-434. Retrieved from <http://crln.acrl.org/content/75/8/431.full>
- Jones, C., Ramanau, R., Cross, S., and Healing, G. (2009). Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers & Education*, 54, 722-732. doi: 10.1016/j.compedu.2009.09.022
- Koltay, T. (2014). Big data, big literacies? *TEMA*, 24, 3-8.
- Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation*, 71(2), 401 - 415. <http://dx.doi.org/10.1108/JD-02-2014-0026>
- Liquete, V. (2012). Can one speak of an "Information Transliteracy"? *International Conference: Media and Information Literacy for Knowledge Societies*. Moscow, Russia. Retrieved from <https://hal.archives-ouvertes.fr/hal-00841948>
- Littlejohn Shinder, D. (2013). The Big Data Dilemma. [Blog]. *GFI Blog*. Retrieved from <http://www.gfi.com/blog/the-big-data-dilemma/>
- Mackey, T.P., & Jacobson, T.E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, 72(1), 62-78. doi: 10.5860/crl-76r1
- MacMillan, D. (2010). Sequencing genetics information: Integrating data into information literacy for undergraduate biology students. *Issues in Science and Technology Librarianship*, 61. doi: 10.5062/F44F1NNK
- MacMillen, D. (2014). Using open access resources in data literacy instructions: Renewing the IL curriculum by aligning it with changing needs. *Library Instruction West 2014*. Paper 24. [http://pdxscholar.library.pdx.edu/liw\\_portland/Presentations/Material/24](http://pdxscholar.library.pdx.edu/liw_portland/Presentations/Material/24)
- Mandinach, E.B, Parton, B.M., Gummer, E.S., & Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, 96(5), 25-28. Retrieved from <http://web.b.ebscohost.com.ezproxy.library.dal.ca/ehost/pdfviewer/pdfviewer?sid=c0f93557-337c-4e0f-96eb-24b20658a586%40sessionmgr110&vid=1&hid=106>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*.



Retrieved from

[http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)

Martin, E., and Leger-Hornby, T. (2012). Framework for a data management curriculum: course plans for data management instruction to undergraduate and graduate students in science, health science, and engineering programs. Retrieved from

[http://library.umassmed.edu/data\\_management\\_frameworks.pdf](http://library.umassmed.edu/data_management_frameworks.pdf)

Maycotte, H.O. (2014). Data Literacy: What It Is and Why None of Us Have It. [Blog]. *Forbes*.

Retrieved from <http://www.forbes.com/sites/homaycotte/2014/10/28/data-literacy-what-it-is-and-why-none-of-us-have-it/>

McKendrick, J. (2015). Data driven and digitally savvy: The rise of the new marketing organization. *Forbes Insights*. Turn: New York, N.Y. Retrieved from

[https://www.turn.com/livingbreathing/assets/089259\\_Data-Driven\\_and\\_Digitally\\_Savvy\\_The\\_Rise\\_of\\_the\\_New\\_Marketing\\_Organization.pdf](https://www.turn.com/livingbreathing/assets/089259_Data-Driven_and_Digitally_Savvy_The_Rise_of_the_New_Marketing_Organization.pdf)

Miller, G., Environmental Commissioner of Ontario. (2015). Connecting the Dots on Climate Data in Ontario: ECO Climate Data Roundtable Executive Summary. Retrieved from

<http://www.eco.on.ca/uploads/Reports-GHG/2015/2015CDRoundtableReport.pdf>

MIT. (2014). [Pamphlet] *Tackling the Challenges of Big Data*. Retrieved from

<https://mitprofessionalx.mit.edu/courses/course-v1:MITProfessionalX+6.BDx+5T2015/about>

Mitrovic, Z. (2015). *Strengthening the Evidence-Base for Open Government in Developing Countries: Building Open Data Capacity Through e-Skills Acquisition*. Mitrovic Development and Research Institute. Retrieved from

<http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper3.pdf>

Mooney, H., and Carlson, J. (2014). Developing Data Literacies for Graduate Students in the Social Sciences. Libraries Faculty and Staff Presentations. [PowerPoint]. Paper 51. Retrieved from

[http://docs.lib.purdue.edu/lib\\_fspres/51](http://docs.lib.purdue.edu/lib_fspres/51)

Oceans of Data Institute. (2014). Profile of a big-data-enabled specialist. Waltham, MA: Education

Development Center, Inc. **OR** Ippolito, J., and Malyn-Smith, J. (2014). Profile of a big-data-enabled specialist. *Ocean's of Data Institute*. Retrieved from <http://oceansofdata.org/our-work/profile-big-data-enabled-specialist>

OECD Skills Outlook 2013. *Organization For Economic Co-Operation and Development*. Retrieved from

[http://skills.oecd.org/OECD\\_Skills\\_Outlook\\_2013.pdf](http://skills.oecd.org/OECD_Skills_Outlook_2013.pdf)

Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 10 to 12: Computer Studies*.

Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/computer.html>

Ontario Minister of Education. (2008). *The Ontario Curriculum Grades 9 and 10: Technological Education*.

Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/teched.html>

Ontario Minister of Education. (2008). *The Ontario Curriculum Grades 11 and 12: Technological Education*.

Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/teched.html>

Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 9 and 10: Mathematics*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/math.html>

Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 11 and 12: Mathematics*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/math.html>

Ontario Ministry of Education, (2008). *The Ontario Curriculum Grades 11 and 12: Interdisciplinary Studies*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/interdisciplinary.html>

P21: Partnership for 21st Century Learning. (2012). Teaching critical thinking skills through project based learning [blog]. Retrieved from <http://www.p21.org/news-events/p21blog/1097-teaching-critical-thinking-skills-through-project-based-learning>

Pentland, A.S. (2013). The data-driven society. *Scientific American*, 309(4), 78-83. Retrieved from <http://www.nature.com/scientificamerican/journal/v309/n4/pdf/scientificamerican1013-78.pdf>

Prado, J.C., and Marzal, M.A. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123-134. DOI: [10.1515/libri-2013-0010](https://doi.org/10.1515/libri-2013-0010)

Pryor, G., & Donnelly, M. (2009). Skilling up to do data: Whose role, whose responsibility, whose career? *The International Journal of Digital Curation*, 2(4), 158-170.

Qin, J., and D'Ignazio, J.<sup>a</sup> (2010). The central role of metadata in a science data literacy course. *Journal of Library Metadata*, 10(2-3), 188-201. DOI: [10.1080/19386389.2010.506379](https://doi.org/10.1080/19386389.2010.506379)

Qin, J., and D'Ignazio, J.<sup>b</sup> (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Librarians, 31st Annual Conference*, Paper 5. Retrieved from <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>

Romani, J.C.C. (2009). Strategies to promote the development of e-competencies in the next generation of professionals: European and international trends. Publisher's version. Retrieved from <http://ora.ox.ac.uk/objects/uuid:da0007a3-b504-4c20-858b-21dd359e3cae>

Sapp Nelson, M., Zilinski, L., and Van Epps, A. (2014). Developing Professional Skills in STEM Students: Data Information Literacy. *Libraries Faculty and Staff Scholarship and Research*. Paper 85. DOI:10.5062/F42V2D2Z  
[http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1087&context=lib\\_fsdocs](http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1087&context=lib_fsdocs)

Scheitle, C.P. (2006). Web-based data analysis: creating a sociological laboratory. *Teaching Sociology*, 34(1), 80-86

Schild, M. (2004). Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*, 28(2/3), 6-11. Retrieved from [http://www.iassistdata.org/downloads/iqvol282\\_3shields.pdf](http://www.iassistdata.org/downloads/iqvol282_3shields.pdf)

Swan, K., Vahey, P., Kratcoski, A., van 't Hooft, M., Rafanan, K., and Stanford, T. (2009). Challenges to cross-disciplinary curricula: Data literacy and divergent disciplinary perspectives. Presented at the Annual Conference of the American Educational Research Association, San Diego, CA. Retrieved from <http://www.sri.com/work/publications/challenges-cross-disciplinary-curricula-data-literacy-and-divergent-disciplinary-p>

Schneider, R. (2013). Research data literacy. *Communications in Computer and Information Science*, 397, 134-140.

Schwieder, D., Fielder, R., and LPSS Library Instruction Committee. (2014). Report: LPSS political science research competency Guidelines Survey. ALA. Retrieved from <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/aboutacrl/directoryofleadership/sections/lpss/lpsswebsite/publications/LPSS%20Research%20Competency%20Guidelines%20Survey%20Report.pdf>

Shorish, Y. (2015). Data information literacy and undergraduates: A critical competency. *College & Undergraduate Libraries*, 22(1), 97-106. doi: 10.1080/10691316.2015.100124

Stephenson, E., and Caravello, P.S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences. *Reference Services Review*, 35(4), 525-540.

Swan, K., Vahey, P., Kratcoski, A., van 't Hooft, M., Rafanan, K., & Stanford, T. (2009). Challenges to cross-disciplinary curricula: Data literacy and divergent disciplinary perspectives. Presented at the Annual Conference of the American Educational Research Association, San Diego, CA. Retrieved from <http://www.sri.com/work/publications/challenges-cross-disciplinary-curricula-data-literacy-and-divergent-disciplinary-p>

Teal, T., Cranston, K., Lapp, H., White, E., Wilson, G., Ram, K., & Pawlik, A. (2015). Data carpentry: Workshops to increase data literacy for researchers. *International Journal of Digital Curation*, 10(1), 135-143.

Thompson, P. (2012). The digital natives as learners: Technology use patterns and approaches to learning. *Computers & Education*, 65(2013), 12-33.

Twidale, M.B., Blake, C. & Grant, J. (2013). Towards a data literate citizenry. In Schamber, L. (Ed.). *iConference 2013 proceedings*, 247-257. Retrieved from <https://www.ideals.illinois.edu/bitstream/handle/2142/38385/189.pdf?sequence=4>

Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educ Stud Math*, 81, 179-205. doi: 10.1007/s10649-012-9392-z

Vesset, D., Olofson, C. W., Schubmehl, D., McDonough, B., Woodward, A., Stires, C., Fleming, M., Nad-karni, A., Zaidi, A., and Dialani, M. (2014). *IDC FutureScape: Worldwide Big Data and Analytics 2015 Predictions* [report]. International Data Corporation.

Wanner, A. (2015). Data literacy instruction in academic libraries: Best practices for librarians. *Archival and Information Studies Student Journal*, 1, 1-17. University of British Columbia. Retrieved from <http://ojs.library.ubc.ca/index.php/seealso/article/view/186335>

Wing, J. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of The Royal Society*, 366, 3717-3725. doi:10.1098/rsta.2008.0118

Womack, R. (2014). Data Visualization and information literacy. *IASSIST Quarterly*, 38(1), 12-17. Retrieved from <http://dx.doi.org/doi:10.7282/T3X92CZF>

Wright, S., Fosmire, M., Jeffryes, J., Bracke, M., & Westra, B. (2012). A multi-institutional project to develop discipline-specific data literacy instruction for graduate students. *Libraries Faculty and Staff Presentations*. Retrieved from [http://docs.lib.purdue.edu/lib\\_fspress/10](http://docs.lib.purdue.edu/lib_fspress/10)

Wyner, Y. (2013). A case study: Using authentic science data for teaching and learning of ecology. *Journal of College Science Teaching*, 42(5), 54-60.

Yeh, K., Xie, Y., & Ke, F. (2011). Teaching computational thinking to non-computing majors using spreadsheet functions. 41st ASEE/IEEE Frontiers in Education Conference: Session F3J. Rapid City, SD.

Zalles, D.R. (2005). Designs for assessing foundational data literacy. *Center for Technology in Learning, SRI International*. Retrieved from <http://serc.carleton.edu/files/NAGTWorkshops/assess/ZallesEssay3.pdf>

## Appendices

Appendix 1: Data Literacy Competencies Matrix

Appendix 2: Data Literacy Definitions Word Cloud

Appendix 3: Key Themes in Data Literacy Literature

Appendix 4: Annotated Bibliography

## Appendix 1 - Data Literacy Competencies Matrix

This matrix consists of key data literacy ability/knowledge areas, and the corresponding competencies and tasks required for each.

Definition: Data literacy is the ability to **collect, manage, evaluate, and apply** data; in a critical manner.

Key Ability/Knowledge Area	Competency	Knowledge/Tasks	LEGEND:			
			Conceptual Competencies	Core Competencies	Advanced Competencies	
<b>Conceptual Framework</b>	<b>Introduction to Data</b>	Knowledge and understanding of data	Knowledge and understanding of the uses and applications of data			
<b>Data Collection</b>	<b>Data Discovery and Collection</b>	Performs data exploration	Identifies useful data	Collects data		
	<b>Evaluating and Ensuring Quality of Data and Sources</b>	Critically assesses sources of data for trustworthiness	Critically evaluates quality of datasets for errors or problems			
<b>Data Management</b>	<b>Data Organization</b>	Knowledge of basic data organization methods and tools	Assesses data organization requirements	Organizes data		
	<b>Data Manipulation</b>	Assesses methods to clean data	Identifies outliers and anomalies	Cleans data		
	<b>Data Conversion (from format to format)</b>	Knowledge of different data types and conversion methods	Converts data from one format or file type to another			
	<b>Metadata Creation and Use</b>	Creates metadata descriptors	Assigns appropriate metadata descriptors to original data sets			
	<b>Data Curation, Security, and Re-Use</b>	Assesses data curation requirements (e.g. retention schedule, storage, accessibility, sharing requirements, etc.)	Assess data security requirements (e.g. restricted access, protected drives, etc.)	Curates data		
	<b>Data Preservation</b>	Assesses requirements for preservation	Assesses methods and tools for data preservation	Preserves data		
<b>Data Evaluation</b>	<b>Data Tools</b>	Knowledge of data analysis tools and techniques	Selects appropriate data analysis tool or technique	Applies data analysis tools and techniques		
	<b>Basic Data Analysis</b>	Develops analysis plans	Applies analysis methods and tools	Conducts exploratory analysis	Evaluates results of analysis	Compares results of analysis with other findings
	<b>Data Interpretation (Understanding Data)</b>	Reads and understands charts, tables, and graphs	Identifies key take-away points, and integrates this with other important information	Identifies discrepancies within the data		
	<b>Identifying Problems Using Data</b>	Uses data to identify problems in practical situations (e.g. workplace efficiency)	Uses data to identify higher level problems (e.g. policy, environment, scientific experimentation, marketing, economics, etc.)			
	<b>Data Visualization</b>	Creates meaningful tables to organize and visually present data	Creates meaningful graphical representations of data	Evaluates effectiveness of graphical representations	Critically assesses graphical representations for accuracy and misrepresentation of data	
	<b>Presenting Data (Verbally)</b>	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 visualizationsto communicate data	Presents arguments and/or outcomes clearly and coherently
	<b>Data Driven Decisions Making (DDDM) (Making decisions based on data)</b>	Prioritizes information garnered from data	Converts data into actionable information	Weighs the merit and impacts of possible solutions/decisions	Implements decisions/solutions	
<b>Data Application</b>	<b>Critical Thinking</b>	Aware of high level issues and challenges associated with data	Thinks critically when working with data			
	<b>Data Culture</b>	Recognizes the importance of data	Supports an environment that fosters critical use of data for learning, research, and decision-making			
	<b>Data Ethics</b>	Aware of legal and ethical issues associated with data	Applies and works with data in an ethical manner			
	<b>Data Citation</b>	Knowledge of widely-accepted data citation methods	Creates correct citations for secondary data sets			
	<b>Data Sharing</b>	Assesses methods and platforms for sharing data	Shares data legally, and ethically			
	<b>Evaluating Decisions Based on Data</b>	Collects follow-up data to assess effectiveness of decisions or solutions based upon data	Conducts analysis of follow-up data	Compares results of analysis with other findings	Evaluates decisions or solutions based on data	Retains original conclusions or decisions, or implements new decisions/solutions

	Competency			LEGEND:	Conceptual Competencies	Core Competencies	Advanced Competencies			
Authors/Source	Introduction to Data	Critical Thinking	Data Culture	Data Ethics	Data Tools	Data Discovery and Collection	Data Management and Organization	Data Manipulation	Evaluating and Ensuring Quality of Data and Sources	
Carlson, Fosmire, Miller, and Sapp Nelson (2011; Carlson, Johnston, Westra, and Nichols, 2012)	*		*	*		*	*		*	
Koltay (2014)						*	*		*	
Prado and Marzal (2013)	*					*	*		*	
Gummer and Mandinach (2015)										
Qin and D'Ignazio (2010)				*		*	*		*	
Gray (2004)	*								*	
Haendel, Vasilevsky, and Wirz (2012)						*	*			
Reeves and Honig (2015)						*	*			
Schild (2004)					*		*			
Ikemoto and Marsh (2008)						*	*			
MacMillan (2010)					*			*		
Shorish (2015)	*	*				*	*		*	
Stephenson and Caravello (2007)				*	*	*		*	*	
Teal, Cranston, Lapp, White, Wilson, Ram, and Pawlik (2015)					*		*			
Wright, Fosmire, Jeffryes, Bracke, and Westra (2012)				*	*		*			
Cowan, Alencar, and McGarry (2014)		*			*	*				
Gunter (2007)		*			*	*	*	*		
MacMillen (2014)		*					*			
McAuley, Rahemtulla, Goulding, and Souch (2010)						*	*		*	
Mooney and Carlson (2014)	*		*	*	*	*	*		*	
Sapp Nelson, Zilinski, and Van Epps (2014)						*			*	
Scheitle (2006)							*	*		
Stout and Graham (2007)				*	*		*			
Swan and Brown (2008)	*				*		*			
Twidale, Blake, and Grant (2013)					*				*	
Womack (2014)		*							*	
Association of College and Research Libraries (2013)				*		*	*		*	
Association of College and Research Libraries (2014)						*		*	*	
Doucette and Fyfe (2013)							*			
McKendrick (2015)	*				*	*				
Swan, Vahey, Kratcoski, van t'Hoof, Rafanan, Stanford (2009)		*			*	*			*	
Zalles (2005)									*	
<b>Totals (across)</b>	7	6	2	7	14	18	20	5	16	

Authors/Source	Data Citation	Basic Data Analysis	Data Visualization	Presenting Data (Verbally)	Data Interpretation (Understanding Data)	Identifying Problems Using Data	Data Driven Decision Making (DDDM)	Evaluating Decisions/Conclusions Based on Data	Metadata Creation and Use
Carlson, Fosmire, Miller, and Sapp Nelson (2011; Carlson, Johnston, Westra, and Nichols, 2012)		*	*						*
Koltay (2014)									*
Prado and Marzal (2013)					*		*		
Gummer and Mandinach (2015)					*	*	*	*	
Qin and D'Ignazio (2010)					*		*		
Gray (2004)					*				
Haendel, Vasilevsky, and Wirz (2012)									
Reeves and Honig (2015)		*			*		*	*	
Schild (2004)		*							
Ikemoto and Marsh (2008)		*			*		*	*	
MacMillan (2010)		*							
Shorish (2015)		*							
Stephenson and Caravello (2007)	*								
Teal, Cranston, Lapp, White, Wilson, Ram, and Pawlik (2015)		*							
Wright, Fosmire, Jeffryes, Bracke, and Westra (2012)									*
Cowan, Alencar, and McGarry (2014)									
Gunter (2007)			*				*		
MacMillan (2014)				*			*		
McAuley, Rahemtulla, Goulding, and Souch (2010)					*		*		
Mooney and Carlson (2014)	*		*						*
Sapp Nelson, Zilinski, and Van Epps (2014)	*	*	*		*	*			
Scheitle (2006)									
Stout and Graham (2007)									*
Swan and Brown (2008)									*
Twidale, Blake, and Grant (2013)			*		*		*		
Womack (2014)			*				*		
Association of College and Research Libraries (2013)	*		*						
Association of College and Research Libraries (2014)									
Doucette and Fyfe (2013)					*				
McKendrick (2015)		*	*	*	*		*	*	
Swan, Vahey, Kratcoski, van t'Hoof, Rafanan, Stanford (2009)				*	*		*		
Zalles (2005)		*	*	*	*		*		
<b>Totals (across)</b>	4	10	9	4	13	2	14	4	6



Authors/Source	Data Curation and Re-Use	Data Sharing	Data Preservation	Data Conversion (from format to format)
Carlson, Fosmire, Miller, and Sapp Nelson (2011; Carlson, Johnston, Westra, and Nichols, 2012)	*		*	*
Koltay (2014)	*		*	*
Prado and Marzal (2013)				
Gummer and Mandinach (2015)				
Qin and D'Ignazio (2010)				
Gray (2004)				
Haendel, Vasilevsky, and Wirz (2012)	*			
Reeves and Honig (2015)				
Schild (2004)				*
Ikemoto and Marsh (2008)				*
MacMillan (2010)				
Shorish (2015)	*	*		
Stephenson and Caravello (2007)				
Teal, Cranston, Lapp, White, Wilson, Ram, and Pawlik (2015)				
Wright, Fosmire, Jeffryes, Bracke, and Westra (2012)	*	*		
Cowan, Alencar, and McGarry (2014)			*	
Gunter (2007)				
MacMillan (2014)				
McAuley, Rahemtulla, Goulding, and Souch (2010)				
Mooney and Carlson (2014)	*		*	*
Sapp Nelson, Zilinski, and Van Epps (2014)				*
Scheitle (2006)				
Stout and Graham (2007)		*	*	
Swan and Brown (2008)	*	*	*	
Twidale, Blake, and Grant (2013)				
Womack (2014)				
Association of College and Research Libraries (2013)	*		*	
Association of College and Research Libraries (2014)	*		*	*
Doucette and Fyfe (2013)		*	*	
McKendrick (2015)				
Swan, Vahey, Kratcoski, van t'Hooft, Rafanan, Stanford (2009)				
Zalles (2005)				*
<b>Totals (across)</b>	9	5	9	8



## Appendix 3 – Key Themes

The following are key themes derived from the literature reviewed (i.e. emergent themes). Although the implications of the knowledge synthesis are based on practical, actionable implications of data literacy, these themes are present through the results of the knowledge synthesis.

### 1. Delivery and Assessment

In terms of top-level delivery, the literature has identified three main learning environments for teaching data literacy at the postsecondary level: in-class, through workshops, and online courses. Classes and workshops often take a **module-based** approach. This allows instructors to focus on specific aspects of data literacy as needed, and for students to build competencies successively (MIT, 2014 ;Qin & D'Ignazio, 2010; Schneider, 2013; Shorish, 2015; Stephenson & Caravello, 2007; Wright, Fosmire, Jeffryes, Bracke & Westra, 2012). There are multiple techniques and tools for evaluating data literacy competencies, and most courses or workshops include a pre and post-assessment.

- **Teaching Environments**
  - **In-class courses**
    - Courses embedded in the curriculum can be an effective way to teach data literacy to students (Martin & Leger-Hornby, 2012),
    - Due to institutions struggling to provide a wide variety of courses, integrating data literacy into existing information literacy courses could be the most feasible option (Carlson, Johnston, Westra, & Nichols, 2013; Hattwig, Bussert, Medaille, & Burgess, 2013; Schneider, 2013; Teal, Cranston, Lapp, White, Wilson, Ram, & Pawlik, 2015; Zilinski, Scherer, & Maybee, 2013)
    - Data literacy could also be embedded into subject-specific courses as targeted training (Hunt, 2004; Schneider, 2013; Teal, Cranston, Lapp, White, Wilson, Ram, & Pawlik, 2015; Wright, Fosmire, Jeffryes, Bracke, & Westra, 2012)
    - Developing foundational skills in-class is important, but encouraging students to improve these skills through experimentation at home is essential to the level of skill they develop (Romani, 2009).
  - **Workshops**
    - Workshops can be used to fill the inevitable gaps in knowledge for students (Carlson, Johnston, Westra, & Nichols, 2013; Hattwig, Bussert, Medaille, & Burgess, 2013; MacMillen, 2014).
    - Academic libraries and librarians (e.g. data librarians) have a key role to play. Can provide workshops, in-class support, and/or act as liaisons to faculty members (Prado & Marzal, 2013; Schield, 2004; Stout & Graham, 2007).
  - **Online courses**
    - Online courses are helpful in developing skills, but students require interaction and targeted teaching methods, according to skill development (Gray, 2004; MacMillen, 2014).

- MIT (2014) offers a comprehensive online course entitled *Tackling the Challenges of Big Data* which has been maximized into a versatile, focused, skill developing class.
- **Assessment**
  - A pre-course inventory of technological and information skills is useful to determine pre-existing skillsets, as well as gaps that require specific attention (Carlson, Fosmire, Miller, & Sapp Nelson, 2011; Chinien & Boutin, 2011; Teal, Cranston, Lapp, White, Wilson, Ram, Pawlik, 2015).
  - Ongoing, iterative evaluation through informal assessment (e.g. conversations, office hours), and critical assessment (e.g. tests, assignments) is important in order to build confidence and mastery (Carlson, Fosmire, Miller, & Sapp Nelson 2011; Hattwig, Bussert, Medaille, & Burgess, 2013; Koltay, 2014).
  - Post-course assessment of students for perceived relevancy and mastery is also important (Carlson, Fosmire, Miller, & Sapp Nelson, 2011).
  - Assessment tools should encompass both process, as well as end results (Liquete, 2012).
  - Methods include pre-screening and testing, self-assessment, demonstration of data tools, online assessment tools, practical case study assignments, performance assessment as an indicator, test of workplace essential skills (TOWES), and academic-style tests (Carlson, Fosmire, Miller, & Sapp Nelson, 2011; Chinien & Boutin, 2011; Romani, 2009; Teal, Cranston, Lapp, White, Wilson, Ram, Pawlik, 2015)
  - Assessment of the course and professors from students is also crucial, so as to improve delivery of content and instruction (Johnson & Jeffryes, 2014)

## 2. Barriers to Effective Data Literacy Instruction

Barriers to teaching data literacy at the postsecondary level can be both tangible and intangible. These barriers can be further broken into three areas: cultural, operational, and technical.

- **Cultural:**
  - Lack of support/conditions for innovative learning, e.g. creativity, risk-taking (Johnson, Adams Becker, Estrada, & Freeman, 2015).
  - Lack of university faculty trust in benefits of data literacy (Ikemoto & Marsh, 2008; Johnson, et al 2015).
  - Mixed support for different data literacy competencies, e.g. teaching some aspects of data literacy, but not others (Carlson, Johnston, Westra, & Nichols, 2013)
  - Abstract nature and perceived complexity of data literacy and related concepts (Qin & D'Ignazio, 2010; Twidale, Blake, & Grant, 2013).
  - Misconception that the Net Generation/Digital Natives are inherently more knowledgeable technically than past students (Thompson, 2012).
- **Operational:**
  - Modifying existing curriculum to include room for course(s) on data literacy (Hunt, 2004; Teal, Cranston, Lapp, White, Wilson, Ram, & Pawik, 2015)
  - Not enough skilled professors, or not enough time to learn the required skills to effectively teach a data literacy course (Boyles, 2012; Carlson, et al, 2013; Teal, et al, 2015)
  - Poor communication regarding student learning goals, as well as a lack of communication/collaboration between professors relating to what has been

- covered in courses, and what should be developed for future courses (Cowan, Alencar, & McGarry, 2014 ; Teal, et al., 2015)
- Differing levels of comfort and experience with technology and/or data, differing levels of education (graduate vs. undergraduate), and age/generational differences between students (Jones, Ramanau, Cross, & Healing, 2009; Shorish, 2015)
- **Technical**
  - Lack of proper storage and organizational tools for data, lack of lab space, lack of access to databases, and lack of computers/other hardware. Cost is also a factor (Stout & Graham, 2007)

### **3. Data Literacy Best Taught At The Commencement of Post-Secondary Studies**

It has been found that teaching data literacy to new (i.e. first year) students is more effective, as students have not yet learned advanced research methodologies.

- This makes it is easier to ingrain good data literacy practices and competencies into workflows and study habits (Hunt, 2004; Sapp Nelson, Zilinski, & Van Epps, 2014; Shorish, 2015; Stephenson & Caravello, 2007; Womack, 2014).
- Although still useful, teaching data literacy and related competencies late in a student's educational career limits their skills to very specific, discipline-focused areas (as opposed to broader, more transferable areas) (Womack, 2014).
- It was also found that workshops and courses open to all levels (e.g. undergraduate and post-graduate) were more challenging to teach due to the inequality of students' technical skills and educational backgrounds (Qin & D'Ignazio, 2010; Scheitle, 2006; Womack, 2014).
- Case studies and practical projects can help bridge the knowledge gaps if courses are taught at a higher level (Twidale, Blake, & Grant, 2013).

### **4. Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions**

Data literacy can sometimes be viewed as the ability to transform information into actionable knowledge and practices by collecting, analyzing, and interpreting all types of data (Koltay, 2014; Mandinach, Parton, Gummer, & Anderson, 2015).

- Focus on data driven decision making (DDDM) has been increased due to recent technological advances and perceived benefits of open and Big Data (e.g. enhanced transparency, enhanced service delivery, citizen engagement, and creation of economic and social value) (Koltay, 2014; Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Hung Byers 2011).
- Effective DDDM is dependent on individuals being data literate (Cowan, Alencar, & McGarry, 2014).
- DDDM's core competencies are centered on interpretation of data, data analysis, and judgement.
- Individuals must be familiar with collecting and analyzing raw data, converting said data/information into actionable knowledge (by using their judgment to prioritize information and weigh the merit of possible solutions), and collecting new data in order to ascertain effectiveness of decisions made (Ikemoto & Marsh, 2008; Mandinach & Gummer, 2012).

### **5. Teach The Teachers**

Teachers cannot help students reach the required data proficiency if they themselves are not proficient, or confident in their data skills (ACRL, 2014; Carlson, Fosmire, Mandinach & Gummer,

2013; Johnson, Adams Becker, Estrada, and Freeman, 2015; Miller, & Nelson, 2011; Romani, 2009; Wanner, 2015).

- Teachers don't know what they don't know, and this can create an "ignorance loop" that can skew their assessment of student performance and feedback (Carlson, Fosmire, Miller, & Nelson, 2011, p. 644).
- Educators should receive systematic training in how to use data, preferably beginning in their preservice years but continuing throughout their academic career (Mandinach & Gummer, 2013).
- New technology being adopted by educational institutions is not enough, teachers must also adopt new teaching and learning paradigms, especially those related to information communication technologies (ICTs) (Romani, 2009).
- Universal data literacy teaching standards are also important (Mandinach & Gummer, 2013)

## **6. 21st Century Skills and Literacies**

Data literacy is considered to be a critical aspect and foundation for the skills (e.g. computational thinking) required in order to be successful in 21st century business, academic, social, and political contexts. It is also considered one of a number of critical literacies that have overlapping competencies and build upon each other (referred to as 'transliteracy' or 'metaliteracy') (Frau-Meigs, 2012; Liquete, 2012; Hattwig, Bussert, Medaille, & Burgess, 2013; Liquete, 2012). Due to their wide breadth of knowledge on these literacies, academic libraries can provide useful expertise and knowledge to help design and deliver data literacy teaching content.

- Data literacy could be considered one of the most relevant E-Skills or "essential survival skills for the 21st century" (Chinien & Boutin, 2011, p.8) (14), because it provides the foundation for interacting in an innovative knowledge-based economy (Mitrovic, 2015; Wanner, 2015)
- Data literacy allows for critical thinking, and for processing more complex cognitive problems, including the ability to analyze problems, create abstractions, and solve said problems (Chinien & Boutin, 2011; Cowan, Alencar, & McGarry, 2014; Yeh, Xie, Ke, 2011)
- Uniquely 21st century problems require 21st century thinking. Data analytics and literacy can provide solutions and explanations to track, predict, and control behaviour, which can be used for good or ill (Pentland, 2013)
- For this reason, teaching data literacy can be very flexible through incorporating these skills into established courses and settings and can be recognized as transferrable (Prado, & Marzal, 2013; Stephenson & Caravello, 2007; Wanner, 2015; Womack, 2014; Zilinski, Scherer, & Maybee, 2013).
- Academic libraries already engage in information literacy training (and other outreach activities) in order to equip students with skills to "locate, evaluate, and effectively use information for any given need" (Shorish, 2015, p. 99).
- Data literacy-related work could be considered a natural progression for the liaison, service delivery, and teaching activities of the modern academic library (Association of College and Research Libraries 2014; Hunt, 2004; Shorish, 2015; Stout & Graham, 2007).
- Key data literacy competency areas for librarians to teach/advise on include data ethics, data organization, data preservation, and data citation (Shorish, 2015; Wright, Fosmire, Jeffries, Bracke, & Westra, 2012).

- Avenues for collaboration include joint workshops, guest lectures, joint course creation and/or teaching (e.g. University of Winnipeg), and training (ACRL, 2014; Hattwig, Bussert, Medaille, & Burgess, 2013; Hunt, 2004; Schield, 2004; Wright, Fosmire, Jeffryes, Bracke, & Westra, 2012).

In order for collaboration to be effective, library staff must be data literate, or at least have a data practitioner toolkit, with the core skills necessary to advise on data-related issues (Hunt, 2004; Pryor & Donnelly, 2009).

## **Appendix 4 – Annotated Bibliography**

The following is an annotated bibliography of sources reviewed in the knowledge synthesis. It includes all items included in the synthesis up to the end of July 2015; additional resources have been included since then, and are not represented here.

# **ANNOTATED BIBLIOGRAPHY: DATA LITERACY**

***DALHOUSIE SSHRC DATA LITERACY  
KNOWLEDGE SYNTHESIS***

**COMPLETED ON: July 27, 2015**



## Table of Contents

<b>Preamble:</b> .....	<b>2</b>
<b>Annotated Bibliography</b> .....	<b>2</b>
Books.....	2
Peer Reviewed Sources .....	3
Grey Literature.....	30
White Papers.....	52
Websites.....	64
Policies.....	65
Courses and Workshops .....	72
Associations and Organizations.....	73

## Preamble:

This document provides an annotated bibliography of sources reviewed and deemed relevant to for the Dalhousie SSHRC Data Literacy Knowledge Synthesis. The information sweep was focused on primarily data literacy related articles, but includes a wide variety of material types in order to cover a wide breadth of knowledge.

## Annotated Bibliography

The following annotated sources contain salient information and points regarding data literacy education at the postsecondary level. The length of entries is varied based on relevance, and actual length of material reviewed. Each entry contains a citation, and related observed themes (via a coloured tag system). Entries are organized based on type of material:

- Books
- Peer-Reviewed Sources
- Grey Literature
- White Papers
- Websites
- Policies
- Courses and Workshops
- Associations and Organizations

Observed themes are colour-coded as follows:

- 21st Century Skills and Literacies
- Barriers to Effective Data Literacy Instruction
- Data Literacy Competencies and Skills
- Data Literacy Best Taught At The Commencement of Post-Secondary Studies
- Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions
- Delivery and Assessment
- Teach The Teachers

### Books

**Citation:** Johnson, C. (2012). *The Information Diet*. Sebastopol, CA: O'Reilly Media, Inc.

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

This book by Clay Johnson includes a chapter (7) devoted to data literacy. It includes points relating to data literacy competencies as well as delivery and educational instruction.

Key data literacy skills include knowledge of how to search for, filter, process, produce, and synthesis data. Searching is especially useful, as knowing how to navigate around barriers to information is crucial. The ability to find data outside of a search engine is integral to success. Moreover, recognizing what is a reliable source of accurate information is important, and requires thinking critically. Individuals require at least basic statistical literacy and fluency in tools to in order to maximize data usage. These skills are best learned through practice and refinement over experiences.

When working with data, it is also important to consider the intent (e.g. to inform, make a point, make a decision, confirm belief or find truth, etc.) and objectivity. John S and James L Knight Foundation “describe this skill as the ability to determine ‘message quality, veracity, credibility, and point of view, while considering potential effects or consequences of messages” (p. 83). Creators are not one dimensional anymore, they engage others to strengthen and clarify their arguments.

### Peer Reviewed Sources

**Citation:** Boyles, T. (2012). 21st century knowledge, skills, and abilities and entrepreneurial competencies: A model for undergraduate entrepreneurship education. *Journal of Entrepreneurship Education*, 15, 41-55.

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

Boyles’ peer reviewed article posits that to be competitive in this economy businesses and especially entrepreneurs must have the ability to create and commercialize knowledge, with emphasis on knowledge, service, and information. The demand for a highly skilled workforce is growing, but graduates are not ready for this level of application upon graduation. Additionally, the American Society for Training and Development revealed that companies value leadership, critical thinking and creativity highest in skillsets, which can be helpful when creating a curriculum that is relevant to today’s industry requirements.

Core competencies of 21st century skills include: capabilities in analytical problem solving, innovation and creativity, self-direction and initiative, flexibility and adaptability, critical thinking, and communication and collaboration skills. Figure one provides a simple breakdown of knowledge, skills, and abilities. The focus is on entrepreneurship, but these skills can be applied to data literacy and 21st century skills and literacies in general, such as the ability to think and reason logically in an effort to solve complex problems, open-ended problems, and goes hand in hand with critical analysis ensuring useful and relevant results; as well as application of analysis, inference and interpretation, evaluation, and synthesis to develop new solutions to complex problems

Entrepreneurial Competencies	21st Century KSAs
Cognitive: Opportunity recognition, alertness; ability to apply systematic search.; creativity	Information, media, and technology literacy: The ability to reason logically to solve complex open-ended problems; to generate meaning and knowledge from information; to critically evaluate information and distill it down to what is useful and relevant, recognize patterns and engage in divergent thinking
	Inventive Thinking: the act of bringing something new and original into existence; the application of analysis, comparison, inference and interpretation, evaluation and synthesis to develop new solutions to complex problems
Social: entrepreneurship as social process; human and social skills; access to resources	Communication and Collaboration: Cooperative interaction to solve problems and create innovations, the ability to read and manage emotions of self and others, to communicate and create meaning through a range of tools and processes

Action-Oriented: Initiative, self-efficacy; self managed process of planning and evaluating; proactiveness; focus on controlling outcomes and personal responsibility for outcomes

Productivity and Resu: the ability to utilize time and resources efficiently and effectively, to develop a plan and monitor progress effectively through the implementation of a plan, self-evaluate, flexibility and adaptability, initiative, and self-direction, and accountability

Table 2 provides a comprehensive overview of evaluation for problem solving abilities: emerging, developing, and mastering categories.

**Citation:** Burdette, A.M., & McLoughlin, K. (2010). Using census data in the classroom to increase quantitative literacy and promote critical sociological thinking. *Teaching Sociology*, 38(3), 247-257.

**Theme(s):** [Delivery and Assessment](#)

**Contribution:**

This paper by Amy M. Burdette and Kerry McLoughlin is centered on a project/assignment carried out at a community college that involved students using United States (US) Census data to compare two counties, with the objective of increasing quantitative literacy and fostering critical (sociological) thinking about communities. The paper includes the assignment and associated grading rubric as appendices. The assignment and other elements of the project have meaningful take-away regarding data literacy.

The authors define quantitative literacy as the ability to understand and manage statistical information. Critical thinking is crucial to being quantitatively literate, i.e. the ability to apply, analyze, and evaluate information. Hands-on exercises and assignments are key to developing data interpretation skills. Students are often exposed to generalizations in introductory courses at the postsecondary level, rather than critically engaging on content. Thus, the authors designed a simple project for students to practice quantitative literacy skills. Students were tasked with choosing two counties from North Carolina, and using US Census Data to compare and contrast similarities and differences based on such factors as: age composition, poverty level, industry, education levels, foreign born residents, percentage of married people, etc.

The assignment consisted of five main sections/steps::

- Data collection
- Critical examination of data
- Formulation of two research questions
- Literature review
  - Could be a mix of academic articles, but also newspapers, online govt. sources, social media (although this may fall under data collection rather than 'literature') etc.
- Summary

The project allowed students to choose which counties to focus on. The authors state that having some element of choice is important for assignment design, as freedom of choice for students (even choosing between two options) promotes more active learning. The assignment was evaluated based on three areas: accuracy in interpretation and discussion of census data, quality of answers to research questions, and writing style. Instructors used a pre-test to measure quantitative data literacy skills before the project, and a post-test after completion of the assignment.

The assignment documented in this paper is very similar to a more advanced type offered in the **Dalhousie Master of Public Administration (MPA)** program course PUAD 6235 Issues in Applied Economics (taught by Thomas Storrington). The assignments for this course consisted of four briefing notes on economic issues in Canada, in which students were required to:

- Collect statistical data (from govt. or other trusted NGOs or institutions, eg. the UN)
  - Create data visualizations from said data
- Conduct a literature review
- Conduct a jurisdictional review/environmental scan
- Analyze said data and literature in order to identify considerations
- Propose policy recommendations

The assignments built on skills through a successive approach:

- Briefing Note #1 (Only collect statistical data and create visualizations; max length six pages)
- Briefing Note #2 (Collect data, conduct a literature review; max length five pages)
- Briefing Note #3 (Collect data, conduct literature review, conduct jurisdictional scan; max length four pages)
- Briefing Note #4 (Collect data, conduct literature review, conduct jurisdictional scan, identify considerations; max length three pages)
- Briefing Note #5 (Collect data, conduct a literature review, conduct jurisdictional scan, identify considerations, propose recommendations, present note as a cabinet submission to class; max two pages)

**Citation:** Carlson, J., Johnston, L., Westra, B., and Nichols, M. (2013). Developing an approach for data management education: A report from the data information literacy project. *The International Journal of Digital Curation*, 8(1), 204-217. doi:10.2218/ijdc.v8i1.254

**Theme(s):** Barriers to Effective Data Literacy Instruction, Delivery and Assessment, Data Literacy Competencies and Skills, Teach the Teacher

**Contribution:**

This article written by Carlson et al., presents the initial results of the Data Information Literacy (DIL) project (approximately half complete). The project consisted of five teams including data librarians, information literacy librarians, and faculty of science representatives from Purdue University, Cornell University, the University of Minnesota, and the University of Oregon, and focused on science research. The team defines essential competencies and investigate the importance of providing graduate students with a diverse experience, while catering to the wide range of skill levels. This article posits that the embedded, standalone course in the postgraduate curricula is important, but not enough; workshops are helpful in supplementing knowledge and bridging this gap, but a balance between the two methods of teaching are ideal to ensure a consistent skill level development for students.

The five teams were assigned to case studies, each defining learning outcomes and developed targeted pedagogies for teaching and evaluating outcomes based on interview feedback. Each team explored a variety of training options, and tested approaches while remaining grounded in disciplinary and local needs. These projects use disciplinary data, relevant to the real world, and integrated into current research practices

Team	Disciplinary Focus	Needs of Project Partner	Educational Approach
Corell	Natural Resources	Data Sharing, Databases and Stewardship	Mini-course (for credit)
Purdue Team 1	Electrical and Computer Engineering	Documenting Code, Organizing Code, Transfer of Responsibility	Embedded Librarianship

Purdue Team 2	Agricultural and Biological Engineering	Standard Operating Procedures Metadata	Workshops
Minnesota	Civil Engineering	Data Ownership, Long-Term Access	Online Course
Oregon	Ecology	Cultures of Practice, Data Sharing & Metadata, Closing Out a Grant	Readings and Team Meetings

\* There is more detailed information about case studies, but not overly useful

The methodology used in the first part of this study was interviews with faculty, and recent graduates, to identify gaps, define most important competencies, and highlight the differing interpretations based on disciplinary experiences. Results confirmed: lack of training in data management; absence of formal policies governing the data in the lab; learning is largely self-directed through trial and error, which focuses on mechanics rather than concepts; and education is too shallow, i.e. know how to use sensors, but not how sensors work.

12 competencies were identified:

1. Data processing and analysis
2. Data management
3. Data preservation
4. Database and data formats
5. Ethics and attribution
6. Data quality and documentation
7. Data curation and Reuse
8. Data conversion and interoperability
9. Data visualization and representation
10. Discovery and acquisition
11. Metadata and data description
12. Cultures of practice

Participants were asked to rank these competencies on a likert scale, identifying data processing and analysis, data visualization and representation, and data management and organization as highest ranked due to their direct importance to research. Although faculty believed these to be essential to data literacy, they did not believe they were knowledgeable enough to teach the skills appropriately. Librarians are an important resource for faculty that lack skill/knowledge, and integrating them into teaching can improve success rates.

**Citation:** Czerkawski, B., and Lyman III, E. (2015). Exploring issues about computational thinking in higher education. *TechTrends*, 59(2), 57-65.

**Theme(s):** 21st Century Skills and Literacies, [Delivery and Assessment](#)

**Contribution:**

This peer-reviewed article, written by Czerkawski and Lyman, argues that computational thinking (CT) is a useful skill across all disciplines, because it is a mental process to solve problems and discover solutions, using logic, algorithmic thinking, recursive thinking, abstraction, parallel thinking, pattern-matching and related processes

This paper focuses on STEM faculties; it can be incorporated into humanities learning, but this would be more difficult, due to the primarily text based analysis. The authors review the status of CT education, and propose potential directions for future application through analysis of survey results. Soh (2009) found that

there is a false dichotomy between “propositional knowledge” and “procedural knowledge”, and an emphasis on emerging models of computation and interdisciplinary training may encourage the development of computational thinking methods suitable to the “open-ended” issues studied in the humanities and fine arts. Soh proposes a framework for multiple pathways of CT learning specialized according to discipline, but students would complete projects collaboratively with students from other disciplines to allow an adaptive learning environment, but reminds the reader that there is a lot of disparity between social, economic, and cultural backgrounds that must be address to be successful

**Citation:** Erwin, R. (2015). Data literacy: Real-world learning through problem-solving with data sets. *American Secondary Education*, 43(2), 18-26.

**Theme(s):** [Delivery and Assessment](#), [21st Century Skills and Literacies](#)

**Contribution:**

Erwin’s peer-reviewed article focuses on the benefits of project-based learning (PBL) in middle school, regarding teaching data literacy. This is the type of skill that must be practiced and not simply taught. Although it is targeted for younger students, it may be adapted to fulfill post-secondary goals, because the learning experience connects multiple curriculum areas and relates to personal experiences, which cultivates higher order thinking.

Projects/lessons are listed for 6-8 graders

1. consider the “story” or guiding questions for the data-centric unit
2. learn about the power and utility of basic descriptive statistics for assigning meaning to large sets of data
3. learn some basic skills for how to use spreadsheet software to examine large data sets
4. examine a relevant data set
5. clean the data
6. analyze the data using basic statistical functions in a spreadsheet
7. interpret the importance of the findings
8. report the findings to an external audience. (19-20)

The audience for PBL goes beyond the teacher; the data is relevant to the community and promotes motivation for students to find solutions, because they are contributing to the community not only a grade. This type of learning encourages task commitment and problem-solving, which are essential to learning complex skills to the fullest. Ocean’s of Data Institute aims to provide more useable, appropriate data for K-16 education.

When developing a PBL course, some resources that can be helpful in guaranteeing that the project is appropriate include Bloom’s Taxonomy, Webb’s Depth of Knowledge Criteria, and the Common Core State Standards (2010). Resources to find data sets to incorporate into projects include: [gapminder.org](http://gapminder.org); [tuvalabs.com](http://tuvalabs.com); *InspireData*; [inspiration.com/inspiredata](http://inspiration.com/inspiredata); [statlit.org](http://statlit.org).

**Citation:** Frau-Meigs, D. (2012). Transliteracy as the new research horizon for media and information literacy. *Media Studies*, 3(6), 14-27.

**Theme(s):** 21st Century Skills and Literacies,

**Contribution:**

Frau-Meigs’ peer reviewed article defines transliteracy as:

1. the ability to embrace the full layout of multimedia which encompasses skills for reading, writing and calculating with all the available tools (from paper to image, from book to wiki);
2. the capacity to navigate through multiple domains, which entails the ability to search, evaluate, test,

validate, and modify information according to its relevant contexts of use (as code, news, and document) (15-16)

- Study of facebook and education concluded that undergraduate students were more positive than graduate students about this tool, and it should be used in conjunction with other LMS - private groups allow for interaction between students and teacher alike.
- Twitter is seen as useful, but limiting - it could replace posting in discussion boards, but the 140 character max is feared to solicit subpar writing - good tool for creating group knowledge, but does not encourage deep thought
- Some suggest that YouTube could exceed the educational uses of the previous two. It was most actively used social media in class to better illustrate certain topics and have 'guest speakers' to clarify points

**Citation:** Gray, A.S. (2004). Data and statistical literacy for librarians. *IASSIST Quarterly*, 28(2), 24-29. Retrieved from <http://www.iassistdata.org/content/data-and-statistical-literacy-librarians>

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment, Teach The Teachers

**Contribution:**

This peer-reviewed article by Ann S. Gray looks at data and statistical literacy for librarians. Specifically, the paper covers how data and statistical resources are evaluated, types of information about data and statistics that are required in order to provide assistance in the use of statistical resources, and the possibility that training in statistics would be useful to providing these services.

The author posits that the growth of online and open data sources has made the ability to understand and represent data crucial for everyday life, as well as studies and work related to economic and social development. Academic librarians have a role in providing support for users, so that they can find "useful information that reflects the nature of the real world" as well as "help users avoid the possible misuse of data and statistics" (p. 24). Librarians should be able to assist/teach students a number of competencies relating to data literacy. Evaluation of data sources and statistics is key, and the ability to judge **quality and utility of data** should serve as a bedrock for teaching data and statistical literacy.

Interpretability, i.e. knowing how to write and communicate the supporting information necessary to interpret and utilize statistical information appropriately is a key skill that should be taught to users and students. Data without description is useless (or at least very hard to use). Data coherence is also important, e.g. standardized terms, classification, and concepts among data products. Students should be aware of the standards surrounding data, and its usage. First year undergraduate students should have the knowledge and skills necessary to:

- Interpret data presented in tables and graphs
- Know the basics of probability theory and the concept of a sample (statistics-focused)
- Have basic knowledge of statistical software packages, and/or analytics tools

Knowing which method to use to gather or analyze statistical information or data is also important, i.e. the right tool for the job. This requires both practical understanding and theoretical grounding. In terms of content delivery, although online courses and texts on data and statistical literacy are useful, user-targeted specific assistance and teaching is often needed (e.g. for users with differing skill levels, or users studying a specific discipline).

**Citation:** Haendel, M.A., Vasilevsky, N.A, and Wirz, J.A. (2012). Dealing with data: A case study on information and data management. *PLOS Biology*, 10(5), 1-4.



**Theme(s):** [Barriers to Effective Data Literacy Instruction](#), [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

This paper by Melissa Haendel, Nicole Vasilevsky, and Jacqueline Wirz provides a case study of the eagle-i Network, a \$15 million NIH-funded pilot project with the aim of facilitating biomedical research by creating a network of research resources repositories. A main theme of the paper is that the scientific community must embrace an information culture, and have the competencies to manage, navigate, and curate huge amount of data.

The current NIH mandate includes the requirement that all peer-reviewed publications funded by the NIH must be accessible by the public. Many other agencies also require data-sharing plans from researchers. Emerging technologies have enabled improved data presentation, and data-citations. The authors argue that being able to create and evaluate these types of data and publications is more important than ever.

Moreover, new and emerging issues concerning data volume, storage, sharing, and cataloguing have created problems for researchers and publishers - 85% surveyed are interested in using other researchers' data, but only 36% report their own data is easily accessible. Information collected on laboratories concerning protocols, instruments, services, software, etc., is lacking. Unique identification and semantic linking are therefore becoming essential to the scientific success of labs in this regard.

To overcome this challenge, the authors propose that researchers should be trained to tag data throughout the research process using universally agreed upon standards. This, as well as linking data, could generate new insight and advance scientific discovery. The authors further state that "statistics, ethics, data and information literacy should accompany scientific training to establish a new cultural standard" (p. 3). It is essential that skills and tools for sharing, organizing, and accessing information and data are accessible, and the authors also put forward the notion that academic libraries can assist in this regard.

**Citation:** Hattwig, D., Bussert, K., Medaille, A., and Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *Libraries and Academy*, 13(1), 61-89. doi:10.1353/pla.2013.0008

**Theme(s):** [21st Century Skills and Literacies](#), [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

This article by Denise Hattwig et al. focuses on visual literacy standards in the post-secondary system, specifically the Association of College and Research Libraries (ACRL)'s *Visual Literacy Competency Standards for Higher Education*, and how to implement teaching said standards. Many of the standards and competencies related to visual literacy are directly related to data literacy.

The authors put forward the notion that visual literacy is associated with with a broader set of literacies that are perceived to be critical for contemporary students/researchers, i.e. transliteracy, metaliteracy. Transliteracy entails working across multiple literacies to construct meaning, while metaliteracy focuses on similarities and connections between different literacies in order to emphasize higher order thinking and collaborative knowledge production. Across all disciplines, students must have the skills to find, interpret, evaluate, use, and produce visual materials in a scholarly context (this also applies directly to data). Visual literacy itself can be defined as the ability to decode, interpret, and create visual messages, as well as encode and compose meaningful visual communications.

The *Visual Literacy Competency Standards for Higher Education* are based upon (and are meant to

complement) the ACRL's *Information Literacy Competency Standards for Higher Education*. Many of the points relating to the competencies can be directly applied to data literacy (and indeed, data visualization). According to the standard, a "visually (data) literate individual is both a critical consumer of visual media (data), and a competent contributor to a body of shared knowledge and culture (brackets inserted by author).

The Visual Literacy Standards consist of seven skill areas:

- Defining the need
- Finding and accessing
- Interpreting and analyzing
- Evaluating
- Using
- Creating; and
- Understanding ethical and legal issues

In terms of instructional method, visual literacy can be merged with information literacy in terms of content (or data literacy), or can be taught as stand-alone courses or workshops. In either setting, librarians can be brought in to work with faculty to help teach students to create visual materials. Effective instruction must be supported by detailed learning outcomes, and must be shaped by an iterative assessment and instruction process

Each standard/skill area has learning performance indicators/outcomes that can be applied to semester-long courses, stand-alone class activities, one-on-one consultations, distance learning situations, and online instructional resources. **The standards are listed as follows, but with 'visual' or 'visually' replaced with 'data', in order to demonstrate similarity:**

1. The 'data' literate student determines the nature and extent of the 'data' materials needed
  - Students must be able to build a 'research context' for their use of visuals and data
  - Students must consider what the required data will look like, the information it might contain, the subject, and the concepts and terms to describe said data
2. The 'data' literate students finds and accesses 'data materials' and media effectively and efficiently
  - Students must be able to make the connections between how they plan to use data, and the most appropriate sources for that use
  - Discovery strategies such as browsing, using search engines, and social linking should be part of a student's data research skills
3. The 'data' literate student interprets and analyzes the meaning of data, and visual data representation
  - Students should be able to carefully assess a visualization or group of data and observe details that may not be noticeable on first glance
  - Choices relating to data-sets and visualization technique are important, and the meaning and influence behind said choices should be explored and understood
4. The 'data' literate student evaluates data and their sources
  - This is a standard that librarians can take the lead in teaching/communicating. e.g. how to differentiate between scholarly/trusted sources of data, and where to find said sources
5. The 'data' literate student uses data and data visualizations effectively
  - This includes being able to use data and visualizations as part of academic projects, or to inform research or decision-making
  - Also includes the knowledge of how to use technology and tools to manipulate and organize data and create visualizations
  - Students must also be able to not only communicate using data, but to communicate about data used in papers, presentations, etc.
  - Students must have the technological know-how to use analytics and visualization tools, and must be aware of and follow design standards
  - Students in essence must be comfortable being content creators and curators
7. The 'data' literate student understands many of the ethical, legal, social, and economic issues

surrounding the creation and use of 'data', and accesses and uses data materials ethically

- Students should explore intellectual property, copyright, and fair use concepts that apply to data, and build a level of knowledge that allows them to use data ethically and responsibly
- Also relates to data presentation, and avoiding misleading charts or visual representations of data
- 'Data' citation is also crucial. Need consistent citation practices, including what should be included in a citation, how to format citations, and how to adapt citations to a variety of end products

**Citation:** Ikemoto, G., and Marsh, J. (2008). Cutting through the data-driven mantra: Different conceptions of data-driven decision making. *Evidence and Decision Making: Yearbook of the National Society for the Study of Education*, 106(1), 105-131.

**Theme(s):** Barriers to Effective Data Literacy Instruction, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Data Literacy Competencies and Skills

**Contribution:**

This article by Gina Schulyer and Julie Marsh focuses on data driven decision-making (DDDM) in education at the k-12 level. Specifically, it looks at how teachers, principals, and administrators systematically collect and analyze data to guide decisions in order to improve the academic success of schools and students. To this end, the authors put forward four types of DDDM, ranging from simple to highly complex. The article does not focus on teaching data literacy, but does have some useful take-away points related to the competencies that are required to carry out DDDM.

According to the authors, there are conflicting interpretations on what constitutes DDDM. Based on a survey of school teachers and administrators collected prior to writing the paper, examples of DDDM included teachers using print-out test scores to target weak areas for further resources, or on the opposite end, numerous stakeholders triangulating multiple forms of data to uncover underlying causes of patterns observed. The authors classified these examples into four types of DDDM:

- **Basic:** Simple analysis decision-making, simple use of data
- **Analysis-focused:** Complex analysis and decision-making, simple use of data
- **Data-focused:** Simple analysis and decision-making, complex use of data
- **Inquiry-focused:** Complex analysis and decision-making, complex use of data

In terms of core competencies for DDDM, the most important are **interpretation** of data, **analysis**, and **judgement**. Individuals must first be able to effectively collect and organize different types of raw data. They must then be able to convert information into actionable knowledge through analysis. Lastly, they must use their judgment to prioritize information, weigh the relative merit and weaknesses of possible solutions, and then make a decision based on this information. Once a decision has been made, individuals must be able to collect new data in order to ascertain whether their actions have been effective.

The above process entails that individuals must know how to continually collect, organize, and synthesize data in support of decision-making. Therefore, individuals must be familiar working with both qualitative and quantitative data. Culture and leadership within a school can influence patterns of data usage, and the importance of these skill-sets. It follows then, that in order to teach data literacy effectively, there exists a need for a supporting data and/or information culture. This type of culture is integral for building trust in using data to help drive decision-making.

**Citation:** Johnson, L., and Jeffryes, J. (2014). Steal this idea: A library instructors' guide to educating students in data management skills. *C&RL News*, 75(8), 431-434. Retrieved from <http://crln.acrl.org/content/75/8/431.full>

**Theme(s):** Delivery and Assessment

**Contribution:**

This peer-reviewed article, written by Johnson and Jeffryes, is part of the Data Information Literacy (DIL) project in association with Purdue, Cornell, Oregon, and Minnesota. This part focuses on research management for graduate students at the University of Minnesota, where faculty implemented an online self-paced course, and then built on this experience to create a comprehensive (non-academic credit) workshop series of seven videos (3-9 minutes long) and five in-person classroom, module-based sessions (1 hour):

1. How to inventory, store, and backup your data
2. How to create data that you (and others) can understand
3. How to navigate rights and ownership of your research data
4. How to share your data and ethically reuse data created by others
5. How to digitally preserve your data for the future
  - a. Optional Homework: complete a data management plan for instructor to review for feedback

Each session included:

- A 'concept check-in' consisting of three to four questions on concepts covered in the videos. Questions embedded in Powerpoint slides, and answered by students with Clickers
- \*\*Hands-on exercises and data scenarios, designed so that students without their own datasets could still engage with content
- Worksheets to apply concepts for students who did have datasets
- One-minute assessment paper with a mix of qualitative and quantitative questions

Johnson and Jeffryes recommend incentivized learning for workshops. This would involve receiving a certificate, or some sort of reward for completing non-credit courses, as well as having the introductory course open to all disciplines would be a positive learning experience, but this would mean a lack of depth for topics. More advanced workshops would benefit from being targeted to specific disciplines to increase relevance and depth of content

**Citation:** Jones, C., Ramanau, R., Cross, S., and Healing, G. (2009). Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers & Education* 54, 722-732. doi: 10.1016/j.compedu.2009.09.022

**Theme(s):** 21st Century Skills and Literacies, [Barriers to Effective Data Literacy Instruction](#), [Delivery and Assessment](#)

**Contribution:**

This peer-reviewed article written by Chris Jones et al., focuses on the debate over whether the Net Generation (i.e. Millennial generation, those born after 1983) have a naturally high aptitude and skills with Web 2.0 and other technologies. Stemming from this, post-secondary students of this generation have been assumed to have very high level proficiency when using technology in relation to their studies and academic research. This has implications for how data literacy (which often utilizes high varying levels of technology) is taught at the postsecondary level.

The conclusions of the authors are based on a two-year study which took place across five universities in 2008 in the United Kingdom. The team distributed a questionnaire across the universities to incoming first-year students in fourteen across a range of pure and applied subject areas.

Based on the results of their survey, the authors argue that in reality, there is a much more complex and diverse range of technological skills within students born post-1983. The majority of students in this age group own a laptop computer, and do use the Internet, but there are significant minorities of students who

do not use certain technologies (e.g. email, social networking sites, etc.). Moreover, the authors' research that although students born post-1983 are comfortable using basic technologies (e.g. office computing software, instant messaging, browsing the Internet), many are not comfortable using more specialized technologies (e.g. data manipulation software, graphics and design software). Moreover, students being of a younger age is not necessarily a key indicator for technological prowess.

In terms of actual study habits, although students do use Web 2.0 technologies to supplement their studies, there has not been a radical shift in student study patterns, and most study patterns still conform to traditional lecturing methods. In terms of data literacy, instruction in a post-secondary setting should take into account that any given student base will likely have a very diverse range of skills and competencies in terms of technology.

**Citation:** Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation*, 71(2), 401 - 415. <http://dx.doi.org/10.1108/JD-02-2014-0026>

**Theme(s):** 21st Century Skills and Literacies, Barriers to Effective Data Literacy Instruction, Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

Koltay's focus is primarily on research data. It is argued that data allows researchers to ask questions in new ways, in all disciplines, posing concern for access, management, sharing, and preservation. Also, the "lack of tools, infrastructure, standardized processes and properly skilled personnel may impede the continued development of e-research" (402), limiting our advancement of knowledge in a knowledge economy.

Koltay argues that one of the most important goals of data literacy should be fostering critical thinking to keep people realistic and asking questions, not only accepting information at face value, as well as the skills for understanding data's underlying meaning. Methods proposed focus on the role of big data; data literacy in relation to information literacy and other literacies; content analysis of data literacy education; and the role of the library in it. The audience is reminded that library professionals have the experience, skills, familiarity of needs, and ability to instruct and consult in research data to help academic institutions succeed, and should be used for support frequently. This type of support is recognized in relation to the dual nature of data literacy education: target appropriate audience and provide able professional support and training; re-skilling may be essential to the success of programs, and librarians are a great resource in helping with this

Koltay lists Carlson, Qin, Schneider, Calzada, and Mandinach and Gummer's varying definitions and competencies of data literacy, and argues that there should be only one definition; many definitions overlap with different terms, but the meaning is the same, and most recognize the difference between producer and creator, and the importance of including each. Additionally, competency mobilizations for researchers today are on three levels: 1. conceptual competencies that include among others innovative thinking, problem solving, and critical thinking; 2. human competencies, like social networking skills, self-management and cross-cultural interaction skills; and 3. practical competencies that include media literacy and information literacy (Lee, 2013) (408)

Koltay also lists several academic authors that argue data literacy education should borrow heavily from information literacy, and others that include *scientific literacy* into the necessary component of education in data. additionally, data literacy education is encouraged to be collaborative throughout faculties and disciplines; this places the student at the centre of the process, and broaches the theory-practice divide.

It is argued that many different literacies have overlapping concepts:

- *Scientific literacy* is not only for scientists, "it is a complex set of knowledge of methods, approaches, attitudes and skills, related to a set of questions on how to do scientific research"

(410), which is widely applicable in 21st century society.

- *Academic literacy* has several similar aspects to data literacy for researchers - for one, the vocabulary used is helpful in identifying which strategy to use in analyzing the data according to discipline
- Some competencies that are listed in *information literacy* are more directly connected to data literacy: "sensitivity for meaning and for the intended audience, interpreting, using and producing information presented in graphic and visual format, as well as making distinctions between essential and nonessential information, fact and opinion, proposition and arguments; distinguishing between cause and effect are examples of this. The requirements to classify, categorize and handle data and to do simple numerical estimations and computations are more directly connected to data literacy" (Weideman, 2013) (409)
- *Statistical literacy* is tied closely with data literacy with a common set of problems and a similar approach to them, dealing with interdisciplinary study and fundamentals - information and data literacies evaluate sources and are needed to access, manipulate and summarize, but statistical literacy guides this process (Shield, 2004)
- *Visual literacy* shares methodological similarities with data and information literacies with a critical approach
- *Media literacy* shares characteristics with data literacy regarding use and reuse of content by third parties - there is a convergence of media, information, and communications technologies; and Web 2.0 has invited everyone to be content creators, who would benefit from knowing these skills - more of a general literacy, less influential than other literacies, but worth mentioning
- *Digital literacy* is also a general literacy, but most information is digital today, and this emphasizes technology to accentuate skills, especially communication of results
- *Metaliteracy* provides a foundations for media, digital, and other literacies, emphasizing content; the lines become blurred between them

**Citation:** MacMillan, D. (2010). Sequencing genetics information: Integrating data into information literacy for undergraduate biology students. *Issues in Science and Technology Librarianship*, 61. doi: 10.5062/F44F1NNK

**Theme(s):** [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

This paper by Don MacMillan, looks at a case study of an information literacy lab for undergraduate biology students at the University of Calgary. The lab was aimed at building competencies for working through a range of resources to discover different aspects of genetic information. Although the content of the lab is specifically focused on information literacy, and learning how to use specific databases, the structure and execution of the lab sessions holds lessons that could be applied when developing data literacy labs and workshops.

The information literacy lab in question was developed specifically for students within the course Biology 311 - Principles of Genetics course, which covers such topics as molecular genetics, sex determination, and structure and function of genetic material. The lab took place during a regularly scheduled three-hour lecture in week 11 of the autumn semester, with over 560 students taking part. As such, the lab was in actuality three lab sessions running parallel, with students divided into groups of 72. This is an unusually large-size for a workshop-style lab, and is evidenced by the fact that the 72 students per session were required to work out of a computer lab with a capacity of 50. As a result, students had to work in pairs. Despite this, the lab sessions were considered a success.

Key to this success was strong collaboration between university librarians, who assisted the course instructors in designing and carrying out the lab (as well as providing the physical lab space). The lab followed a pattern of demonstration, practice, and discussion among students. Moreover, examples and resources (e.g. NCBI Genes and Disease database, PubMed, etc.) were targeted to the discipline of the

students. This made demonstrating more abstract theories easier, as there were readily available and understandable examples where said theories could be applied. Three other factors were critical for the successful implementation of the lab:

1. Labs consisted of a structured demonstration of resources/tools, progressing from simple to more complex material
2. Hands-on exercises requiring the utilization/manipulation of data through the covered resources/tools
3. Inclusion of a follow-up assignment that involves deeper critical thinking, further independent exploration of resources, and synthesis and analysis of information in order to consolidate learning

The lab was also supported by a workbook containing detailed information, and an online resource with links to follow ongoing demonstrations, as well as other relevant resources. The author concludes that workshop style labs benefit from having a clear purpose and connection to material specific to students; discipline, opportunities for hands-on practice and active learning, and adequate preparation before the lab takes place (e.g. pre-homework or online introductory material).

**Citation:** Mandinach, E.B., & Gummer, E.S. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, 42(1), 30-37. DOI: 10.3102/0013189X12459803

**Theme(s):** Barriers to Effective Data Literacy Instruction, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Data Literacy Competencies and Skills, Delivery and Assessment, Teach the Teacher

**Contribution:**

Mandinach and Gummer present the context and trends surrounding data literacy from the perspective of educating pre-service teachers, including supportive technologies, standards, and accreditation. Exploration or requisite skills and knowledge, current practices in schools of education, and the role of various stakeholder groups are considered. The authors go on to explore the systematic nature of components that can contribute to developing data literacy, and not gaps in research, postulating a framework to move the field forward. This framework would allow for the ability to understand and use data effectively to make decisions. It is recommended to consult *The Interstate Teacher Assessment and Support Consortium* in developing a framework and accreditation.

Mandinach and Gummer highlight four trends that have emerged that are pushing data literacy: (a) the increased emphasis on data in (US) federal policy, (b) the development of the statewide longitudinal data systems (SLDS), (c) the growth of local data systems, and (d) additions to standards and accreditation processes that address data literacy. Considering the wide range of stakeholders in this field, schools, practitioners, professional development providers, provincial educators departments, the federal government, professional organizations, etc., there should be involvement from these groups to ensure that data literacy education is comprehensive in implementation.

It is recommended that data literacy be taught in a targeted environment, such as schools of education for preservice teachers, to develop the appropriate foundation for applying the skills.

**Citation:** Pentland, A.S. (2013). The data-driven society. *Scientific American*, 309(4), 78-83. Retrieved from <http://www.nature.com/scientificamerican/journal/v309/n4/pdf/scientificamerican1013-78.pdf>

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills

**Contribution:**

This article by Alex 'Sandy' Pentland looks at the data-driven society of the 21st century, and the data analysis skills required in order to interact and solve problems within this context.

The author posits that 21st century challenges (e.g. global warming, etc.) demand 21st century thinking. However, many professionals still think about social systems using Enlightenment-era concepts (e.g. markets) that reduce societal interactions to rules or algorithms while ignoring human behaviour. There is a need to go deeper, to take into account the fine-grained details of societal interactions. Data analytics can provide solutions and explanations through the ability to track, predict, potentially control behaviour for good or ill.

Engagement and exploration are key factors in data analysis, often resulting in innovation and creative output. Establishing new connections among people is also a driver of innovation and creative output; humans are social by nature and productivity tends to rise when this is encouraged. Moreover, the scientific method is not enough anymore, as data collected is often messy, complex, and large-scale with thousands of reasonable hypotheses. Individuals must be able to carry out more frequent testing, ongoing testing and analysis.

**Citation:** Prado, J.C., and Marzal, M.A. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123-134. DOI: [10.1515/libri-2013-0010](https://doi.org/10.1515/libri-2013-0010)

**Theme(s):** [21st Century Skills and Literacies](#), [Data Literacy Competencies and Skills](#)

**Contribution:**

This paper by Javier Calzada Prado and Miguel Angel Marzal looks at the issue of library standards regarding data literacy. They propose of a set of core competencies that can serve as a framework of reference for its inclusion in the information literacy programs of academic libraries. The paper includes these core competencies (and includes competencies put forth by other scholars), and outlines some data literacy programs/courses taught by universities (at the time of publication).

Due to the advent of the so-called Information Society and events such as the Open Data movement, there have been massive changes in the public sector, scientific, and academic spheres in terms of the availability of useful data. Therefore, the authors argue that there exists a need for greater sensitization and training in data literacy, a suite of data acquisition, evaluation, handling, analysis and interpretation-related competencies that lie outside the scope of traditional literacies. That being said, data literacy does have strong connections with information and statistical literacies.

Data literacy is the part of statistical literacy that involves training individuals to access, assess, manipulate, summarize, and present data. In terms of information literacy, data literacy is the component that enables individuals to access, interpret, critically assess, manage, handle, and ethically use data. In all aspects of data literacy, critical thinking on the part of the user is crucial. Academic libraries must continue to update and provide training for data literacy to students and faculty along these lines, and are well-position to lend support due to the inherent expertise of librarians regarding information literacy.

The data literacy competencies framework for teaching put forth by the authors is below. Under each module is the recommended to be covered.

1. Understanding data
  - What is data?
    - Data definition, types of data
  - Data in society



- Data producers and consumers, data lifecycle, data applications and their impact on society and science, copyright and data reuse
- 2. Finding/obtaining data
  - Data sources
    - Data source examples, criteria for assessing data sources
  - Obtaining data
    - Main research methods for obtaining original data
- 3. Reading, interpreting, and evaluating data
  - Reading and interpreting data
    - Ways to present, and represent data
  - Evaluating data
    - Data evaluation criteria
      - Authorship
      - Method of obtaining data
      - Method of analyzing data
      - Comparing data
      - Inference and data summary
- 4. Managing data
  - Data and metadata collection and management
    - Metadata, reference management tools, databases, data management repositories, policies and practices
- 5. Using data
  - Data handling
    - Data conversion, handling data analysis tools locally and online (e.g. Excel, SPSS, Stata, etc.)
  - Producing elements for data synthesis
    - Choosing suitable data representation methods, visualization tools
  - Ethical use of data
    - What is the ethical use of data, how to cite data sources

**Citation:** Pryor, G., and Donnelly, M. (2009). Skilling up to do data: Whose role, whose responsibility, whose career? *The International Journal of Digital Curation*, 2(4), 158-170.

**Theme(s):** 21st Century Skills and Literacies, [Barriers to Effective Data Literacy Instruction](#), [Data Literacy Best Taught At The Commencement of Post-Secondary Studies](#), [Delivery and Assessment](#)

**Contribution:**

This peer-reviewed article focuses on data management, and provides the reader with The Data Practitioners "Toolkit", which is applicable mainly to four roles: data creator, data scientist, data manager, and data librarian. These roles would benefit from support from IT and library communities, because the skills and knowledge associated with success are instilled in these professionals. The authors defined IT experts as conduit specialists, library or information scientists as content specialists, and academics or professionals as context specialists. This clarification allows for a more targeted education, because it highlights what skills are being used most.

Sheila Corral's provides a potential shape for a training regime, which addresses "breadth and depth of competency requirements, combining technical expertise with contextual understanding (ie. significant domain knowledge) and interpersonal skills" (166) relating primarily to research data management; Chris Rusbridge and Martin Donnelly also provide Core Skills for Data Management. Corral also recommends that data skills become a core academic competency, so the appropriate level of knowledge is instilled in students entering the professional environment.

**Citation:** Qin, J., and D'Ignazio, J. (2010). The central role of metadata in a science data literacy course. *Journal of Library Metadata*, 10(2-3), 188-201. DOI: 10.1080/19386389.2010.506379

**Theme(s):** Data Literacy Competencies and Skills, Barriers to Effective Data Literacy Instruction

**Contribution:**

Qin and D'Ignazio conduct a study at Syracuse University, recognizing the rapid development of ICTs and internet, leading to requiring improvements to data literacy and management in science research and promote contributions to the field and for future use. Knowledge and skills in data life-cycle, metadata standards and practice, data tools, and communication and collaboration mechanisms are increasingly valuable qualifications in the workforce today.

With the growing amount of data being produced and used there is a demand that students graduate with skills in data management and data use, including understanding a wide variety of tool for accessing, converting, and manipulating data. Science data literacy core competencies and skills support collecting, processing, manipulating, evaluating, and using data, which are widely transferable skills. Hunt 2005, Shrimplin and Yu 2004, and Whitmer, Blanchette, and Caron 2004 all discuss integrating data literacy into college classrooms, using collaboration, tools, and active learning pedagogies in teaching, "pushing" the data to students. Another approach is "pulling" data (processing and managing data), which is an outcome-based data literacy education that provides more hands-on opportunities and promote learning

The Science Data Literacy (SDL) project and Science Data Management (SDM) course at Syracuse University's iSchool are studied in this paper. The majority of students enrolled were IM students, and included graduate and undergraduate students alike in STEM disciplines (Qin and D'Ignazio) geospatial, climate, and biological data. They distributed a survey to faculty, in two interactions, and it was clear that their methods differed drastically, and there is a need for standards in STEM disciplines for metadata.

These responses helped develop the course into three aspects:

- data life-cycle: different formats, records, and resources, a stage environment for who, what, where, why, when, and how which is useful internally and externally;
- the technical aspect of data management: description, indexing, storing, and managing data objects and repositories, important management tool providing consistent access, includes format, types, and sizes
- Social and policy issues of data: use and management, but also privacy, ethics, security, and intellectual property, as well as separate disciplinary issues

The authors also conducted pre and post surveys to students evaluating outcomes; these results are reported in more depth in the other paper by these authors in this framework, but some were listed, including challenges arose in the inequality of students' technical skills and educational backgrounds. Case studies and authentic project involvement are recommended to help bridge these gaps in understanding.

Furthermore, Qin and D'Ignazio searched on the Web for other courses offered in data-related topics in any disciplines, and searched for the three aspects being focused on for the development. Table 2 lists categories of relevant courses, level, and focus; the most relevant ones were project oriented with practical implications, and case study based providing flexibility to both data and metadata oriented skills

Through developing the course, the authors' goals included (a) the informational issues of managing data resources, (b) the technologies currently being used in data and metadata management, and the social and policy issues related to data's role in the work practice of individual science disciplines and research domains (197). Four strategies were used in this development, and reflect previous projects and is a culmination of all projects:

1. topic introduction and development in three modules: 1. fundamentals of science data and

- management; 2. managing sets in aggregation; 3. broader issues in science data management - Table 3 presents subtopics of modules
2. course content concerning literacy and skills in collecting, processing, managing and using data in scientific enquiry with the central role of metadata. Course readings were diverse and incorporated data management and metadata
  3. attempted to provide students with a sense of comfort with multiple-level, distributed information systems representing and containing an unfamiliar collection of digital data files or values (200) - Case studies are the most useful way to teach this
  4. enabled was to overcome the barriers of understanding abstraction and complexity. Attempted to allow students to make connections to existing knowledge through their disciplines, done in groups and interviewing researchers in their field, and recommended tools, technology, schemes, complete with some degree of executed solution with examples.

**Citation:** Reeves, T., and Honig, S. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, 50, 90-101. <http://dx.doi.org/10.1016/j.tate.2015.05.007>

**Theme(s):** Barriers to Effective Data Literacy Instruction, Delivery and Assessment, Teach the Teacher

**Contribution:**

Reeves and Honig’s peer reviewed article focuses on pre-service teachers learning how to use data to evaluate student outcomes. They begin by placing utmost importance in including data literacy training in core courses for student, and referring to Marsh (2012), who outlines competencies for data literacy as:

1. accessing or collecting data;
2. filtering, organizing, or analyzing data into information
3. combining information with expertise and understanding to build knowledge
4. knowing how to respond and taking action or adjusting one’s practice
5. assessing the effectiveness of these actions or outcomes that result (91)

Additionally, Wayman and Jimerson (2013) are referred to for their list of areas that are perceived to need more instruction: “identification of questions to ask; analysis and interpretation of data; linking data to instructional practice; and collaboration about data” (92). Other studies recognize alternative ways to address this lack of instruction, such as: classroom-contextualized intervention (targeted), and active learning; it can also be useful to connect curricular and institutional goals to interventions, as well as mentoring and feedback can be essential in developing these skills.

Reeves and Honig identify numerous dimensions to data education: data types, level of education, activities, use of technology for analysis, interpretation and use, and report that positive outcomes came from collaboration between teachers, the presence of an expert facilitator, a clear and specific process for data protocol and tools to guide the protocol, and balancing attention to general assessment with classroom instruction

Their study included undergraduate pre-service elementary teachers (k-8) and enrolled in an assessment course in either the spring and fall terms of 2014. They were provided with instruction in scoring student work, summarizing, and disaggregating data

- Intervention schedule for Spring 2014:
  - Day 1- using Excel to aggregate the data into spreadsheet format from scoring and assessments keys, and examined qualitative data for themes and commonalities between incorrect answers, and examined quality and reliability of assessments
  - Day 2- analysis and interpretation of traditional and performance assessment data and making decisions based on these from spreadsheet from day 1. investigated frequency distribution through graphs, and articulated results in a Word document

The methodology used consisted of a pretest and posttest, measuring answers to questions asked above. Pre-intervention surveys were sent out one week before commencement, and post-intervention surveys sent one week following completion. The feedback collected allowed for changes to provide more targeted feedback between semesters. The second iteration of the test included 13 items focused on participants' data literacy knowledge. This included questions about interpreting z-scores and dichotomous item difficulties, read and interpret tables and graphs, compare scores across groups and over time, compare group and individual scores represented in tables, and identify evidence that would best support particular instructional decisions (95-96). Participants identified the most useful aspect of the experience as 1. use of Excel to manage and analyze data; 2. feedback from faculty during process; 3. use of graphs to interpret data; 4. opportunity to learn how to interpret data; and 5. step-by-step guidance; and identified the least useful aspects as: too repetitive, working with peers was not helpful, and paced too slowly

Fall 2014 small iterations were made, but most substantial were 5 hours opposed to 3 hours spent on feedback on drafts pertaining to validity and quality, and roundtable discussions concerning this; more focus was placed on intervention protocols and tools: repetitiveness of step-by-step protocol, teaching traditional and performance assessment separately, rather than jumping back and forth; and the data-based decision making worksheet and the qualitative worksheet were modified to increase specificity and scaffolding (specific, relevant evidence), as well as provided with a worksheet from students from the first implementation

Overall, the findings from this study are consistent with the value of contextualized, instructionally relevant interventions. The pre and post tests allow for targeted training, as well as changes to enhance effectiveness.

**Citation:** Schield, M. (2004). Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*, 28(2/3), 6-11. Retrieved from [http://www.iassistdata.org/downloads/iqvol282\\_3shields.pdf](http://www.iassistdata.org/downloads/iqvol282_3shields.pdf)

**Theme(s):** 21st Century Skills and Literacies, [Delivery and Assessment](#), [Teach The Teachers](#)

**Contribution:**

This peer-reviewed article by Milo Schield posits that the evaluation of information is a critical element of information literacy, statistical literacy, and data literacy. Due to this, all three literacies are interconnected, and it is thus difficult to promote one without involving elements of the others. The author argues that academic librarians (specifically data librarians and information literacy specialists) should engage in teaching these literacies. The article puts emphasis on statistical literacy, but does have some useful information regarding data literacy instruction.

The author argues that the aforementioned literacies are often all involved in the problems that university students face. Students must be able to think critically about concepts, claims and arguments: to read, interpret and evaluate information (i.e. information literacy). Students must be statistically literate: they must be able to think critically about basic descriptive statistics (i.e. statistical literacy).. Analyzing, interpreting and evaluating statistics as evidence is a special skill. And students must also be data literate: they must be able to access, assess, manipulate, summarize, and present data. Therefore, integrated teaching of these literacies can better equip students with the tools they need to tackle challenges both within university and post-graduation.

According to the author, students within the social sciences and business/management disciplines in particular must be able to work with data. Related to this, specific data literacy competencies include:

- Understanding SQL
- How to build and use relational databases
- Data manipulation techniques
- Statistical software, e.g. SPSS, STATA, Minilab, and Microsoft Excel

In terms of instruction, students can be taught to use skills from each of the above three 'literacy' types in different ways. e.g. Critical Thinking Perspective, Discipline Perspective. Moreover, librarians can serve as teachers due to the fact that "they are eminently qualified to teach students how to think critically, how to become information literate, how to become statistically literate and how to become data literate" (p. 9). However, in order to teach students these literacies, librarians (and university instructors) must be data, information, and statistically literate.

**Citation:** Schneider, R. (2013). Research data literacy. *Communications in Computer and Information Science*, 397, 134-140.

**Theme(s):** 21st Century Skills and Literacies, [Delivery and Assessment](#)

**Contribution:**

This peer-reviewed article, written by Schneider, posits that the integration of research data literacy education being incorporated into information literacy established curricula, may be the best way to ensure that these skills are being taught, in a curriculum that is already stretched too thin.

The course consisted of processing of all types of data, raw or primary; creation, management, the data lifecycle and reuse of research data from two dimensions: different student populations and various teaching modules. Schneider highlights the importance of recognizing the varying degrees of education, and personal context based on experience and need. Having a broad focus allows for the level of experience to even out throughout the course.

A methodology for teaching the varying levels of skills is outlined:

1. two-hour unit providing basic principles and methods;
2. full course or workshop providing a broad theoretical overview and introduction to methods and tools used
3. full module - teaching unit made up of several courses providing a complete overview of theory and practice, between 6 months and a year
4. specialization of all techniques in preparation for the field after graduation
5. full study program is comprehensive over two years based on foundations and new competencies
6. certificate is similar to full study, but geared towards people already in the field with a need to improve skills and competencies

Schneider based the course on three studies: Shapiro and Hughes provides basis for data literacy curriculum as tool literacy, resource literacy, socio-cultural literacy, emerging technology literacy, and critical literacy; Eisenberg's Information Literacy (2008) defines essential skills that relate to data literacy as ability to clarify, locate, select/analyze, organize/synthesize, create/present, and evaluate information; and the Working Group on Information Literacy defines seven pillars that align with data literacy as identify, scope, plan, gather, evaluate, manage, and present/provide

Research Data Literacy	Data Management Competencies
Identify	Documentation (research environmental, temporal) / Context / From Information Management to Knowledge Management
Scope	Monitoring Process / Extracting Information from Data Models (and People)
Plan	Data Modeling / Metadata / Standards Development
Store	Data Analysis and Manipulation / Merging, Mashing, Integration

Protect	Data Preservation / Data Security / Access Authentication / Conditions of Use / Data Legislation
Evaluate	Data Appraisal and Retention / Value of Data / Economic Issues
Manage	Complaints and Expectation Management / Coordination of Practice across Institution / Negotiation Skills / Risk & Disaster Management / Contingency / Advocacy, Promotion, Marketing
Provide	Facilitation, Communication / Raising Awareness

**Citation:** Shorish, Y. (2015). Data information literacy and undergraduates: A critical competency. *College & Undergraduate Libraries*, 22(1), 97-106. doi: 10.1080/10691316.2015.100124

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills, Data Literacy Best Taught At The Commencement of Post-Secondary Studies

**Contribution:**

Shorish uses the DIL project as a framework for incorporating data literacy into established science curricula, with importance placed on beginning, optimally, in first year of the undergraduate degree to more easily integrated into workflows. This can be done through building foundations through courses such as research methods, which would help in creating a fluid environment to learning skills, which will encourage lifelong learning through adaptability and relativity. The overlap in data, statistical, and information literacies can also be seen in this teaching method.

Librarians are a great resource to use for data research skills development. A collaborative approach to teach can also improve the outcomes for students. Their skill-set encourage adaptability and teaching at different levels, and they are required to have a knowledge of these competencies to some degree already.

Examples of undergraduate instruction in DIL are listed from Syracuse University, University of Massachusetts, University of Minnesota, New England Collaborative Data Management Curriculum, and Edinburgh's MANTRA online course. Take aways according to these include teaching mixed audience classes are difficult, due to varying skill levels; teaching through modules allows the teacher to target training to specific subjects and skill levels, outside of a class setting

**Citation:** Stephenson, E., and Caravello, P.S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences. *Reference Services Review*, 35(4), 525-540.

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

This paper by Elizabeth Stephenson and Patti Schifter Caravello describes and analyzes the confluence of data literacy with information literacy. It does so by focusing on a case study (the UCLA Sociology Information Literacy Project) of an undergraduate Sociology information literacy course that incorporated data literacy into its curriculum through the use of different sessions and modules. The paper also includes a literature review of material relating to data literacy.

The project was put together as a joint initiative by the UCLA Sociology Department and the University library. The course Sociology 105 was selected to serve as the project case study. The data literacy modules were designed to engage students in activities to help them effectively use statistical resources in course assignments and papers, as well as critically evaluate graphical representations of data.

Students were required to demonstrate that they could find, evaluate, and use information and data effectively and ethically for sociological inquiries. The course also built upon more basic information literacy instruction by delving into evaluation and research strategies, sociological resources, and concepts in using statistical information. Students were also encouraged to use the data literacy concepts learned to conduct research in other courses.

Relevant learning outcomes/competencies of the data literacy modules included:

- Ability to read and critically evaluate simple tables;
- Produce accurate bibliographic citations for data;
- Utilize American Factfinder to create a table; and
- Evaluate graphical representations of data, and discuss the content in relation to an accompanying written material.

**Citation:** Teal, T., Cranston, K., Lapp, H., White, E., Wilson, G., Ram, K., and Pawlik, A. (2015). Data carpentry: Workshops to increase data literacy for researchers. *International Journal of Digital Curation*, 10(1), 135-143.

**Theme(s):** [Barriers to Effective Data Literacy Instruction](#), [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

- Two-day, hands-on bootcamp style online domain-specific workshop teaching foundational skills and best practices of in Software Carpentry, enabling researchers to retrieve, view, manipulate, analyze and store their and other's data
- Many researchers lack the computational and statistical knowledge to communicate analysis, many are unfamiliar with best practices and what they have learned has been piecemeal or not learned at all - 2013 survey conveyed an overwhelming demand for training in the research sector, lack of skills and confidence is limiting research progress
- Barriers include: full curriculum, no room or time to add additional subject matter; not enough time to learn the required skills to teach them due to existing commitments; no good model for community lesson development, resulting in disjointed and overlapping - communication is key in offering an effective and efficient learning goals - professors or community of practices offering guidance to related courses on what has been covered and what should be developed more in future classes
- Guidelines for the initial Data Carpentry core content:
  - Workshops are domain specific: data type vary between fields, analysis and standard problems allowing achievement of two goals: more immediate understanding of questions and approaches, and applying these to their work, using "real world" examples
  - Workshops are a narrative that show the data lifecycle for a given dataset or problem: fundamental in quality of final analysis from beginning to end, and enabling ability for results to be reproducible. Models a user workflow using their own datasets
  - Workshops are designed for people with no prior computational experience: there is a clear expectations for the pace of instruction, no pressure to prepare, allows them to learn at own pace and build on existing practices and knowledge
  - Workshops can be focused on any research domain: the principles of the data lifecycle can be applied to any domain and materials adapted to meet specific domain needs
- First workshops used datasets to teach: how to organize data in spreadsheet programs (such as Excel), use spreadsheets more effectively and the limitations of such programs; how to get data out of spreadsheets and into more powerful tools; how to use databases, including managing and querying data in SQL; and how to create workflows and automate repetitive tasks, in particular using the command line shell and shell scripts
- Initial focus as introductory workshops, but working to develop more advanced topics such as

Natural Language Processing, more advanced statistical topics, using cloud resources and using APIs for data access and sharing

- Hackathons to assess the workshops for meeting learning objectives, and more being planned to develop and to improve existing materials. Assessment is done formally at the beginning and end of the workshop through surveys and post-workshop interviews. Informal assessment is done throughout the workshops as instructors gauge the progress of their group, and have conversations through breaks and activities
- Working with foundations, industry, and communities to develop content, respond to user input, and solicit lesson contributions
- Two day workshops are not enough to educate researchers with all of the skills they need, but it is a way to lay the groundwork and begin the process of further learning

**Citation:** Thompson, P. (2012). The digital natives as learners: Technology use patterns and approaches to learning. *Computers & Education*, 65(2013), 12-33.

**Theme(s):** 21st Century Skills and Literacies, [Barriers to Effective Data Literacy Instruction](#), [Delivery and Assessment](#)

**Contribution:**

This paper by Penny Thompson focuses in on digital natives (those born post-1983). Specifically, the study investigated the claims made in the popular press about the “digital native” generation as learners, and that this generation of students thinks and learns differently due to their high interaction with digital media and technology. However, the research carried out by the author debunks this notion, and the evidence to support these claims is scarce. This article is relevant in terms of delivery of content, and provides lessons surrounding assumptions made over the level of technical skills that students may have going into any given data literacy course/workshop.

Digital natives have a distinct set of characteristics and preferences for speed, nonlinear processing, multitasking, and social learning due to immersion in digital technology during childhood when neural plasticity is high. While most students between 20-30 were born after 1980, a study on Australian students found only 14% could be classified as “power users” and most had restricted range of technological experience, using it primarily for basic web functions (e.g. email, searching). Few used multimedia content creation or advanced smartphone capabilities. To further explore this issue, the looked at use of technology and frequency, characteristics associated with interest, personal interest and productivity of learning, and patterns of technology and productivity of habits within digital natives.

A survey was put out to first-year undergraduate students consisting of four sections: digital characteristics scale, the productive learning habits scale (each eight points), technology use (forty-one tools listed), and basic demographic information. Response rate to the survey was 13%. Relevant findings were:

- Proficiency varied widely concerning types of technology tools, and was more limited than the popular press suggests
- Tendency toward fast, expedient web search, rather than iterative style. Students are not taking full advantage of opportunities available on the web for deep learning, and students would benefit from explicit instruction on search term refinement and evaluating hyperlinks
- Students recognize the need to persist even when the information is not entertaining, contrary to popular theories. Students are still able to control multi-tasking, and listen attentively to lectures.
- Lectures are considered a useful learning technique, but should have attempts to engage students as well (e.g. flipped classroom learning)

The author’s findings challenge the popular press assumption that digital natives are a homogenous group of learners. Based on this, the author believes that technology is an influence that interacts with many other influences in digital native culture. The author further states that, “Students may be using a narrower range



of technology tools than the popular press authors claim, and they may not be exploiting the full benefits of these technology tools when using them in a learning context. Findings from this study also suggest that the influence of technology on the digital natives' approaches to learning is varied and complex rather than deterministic" (p. 23). The author puts forth the notion that teachers still play an integral role in the classroom and providing instruction for technologies to broaden student's abilities (e.g. data literacy competencies).

**Citation:** Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educ Stud Math*, 81, 179-205. doi: 10.1007/s10649-012-9392-z

**Theme(s):** [Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Delivery and Assessment](#)

**Contribution:**

This article by Philip Vahey et al. covers the Thinking With Data (TWD) project, carried out in two elementary schools in the United States. The goal of the project was to create a set of cross-disciplinary curricular materials designed to increase student data literacy (as well as an investigation of said material's effectiveness).

Data literacy here is defined by the authors as the ability to make sense of the mass amounts of quantitative data proliferating today's society. It involves students being able to "investigate authentic problems; use data as part of evidence-based thinking; develop and evaluate data-based inferences and explanation; and communicate solutions (p. 181). It also entails students being able to recognize faulty arguments based on data, and create their own valid, data-based arguments.

The research team found that the following data competencies aligned across the disciplines of social studies, mathematics, science, and language arts:

- Formulating and answering data-based questions
- Using appropriate data, tools, and representations; and
- Developing and evaluating data-based inferences and explanations

The team puts forth the notion that the creation of common measures (e.g. compound measures used for comparison, prediction, and argumentation), can form a foundation for more complex learning involving data literacy, specifically proportionality. To this end, the team grounded the TWD project in the teaching of proportional reasoning, utilizing data literacy concepts such as data-based argumentation. The subject of fair/equitable water distribution in the Tigris/Euphrates watershed shared by Syria, Iraq, and Turkey, was used as a case study. By the end of the project, students were expected to determine whether the current allocation of water is equitable among the three states (using the UN Convention on the Law of Non-navigational Uses of International Watercourses as a base guideline).

The project used the Preparation for Future Learning (PFL) framework, wherein students prepare to learn an important concept by investigating a set of problems designed to highlight the structure of the target concept. This reverses the traditional lecture-and-apply process, wherein the students work to find the solution themselves, and then receive formal instruction after the fact. This PFL framework was applied across four successive modules (dubbed PFL+):

- The **social studies** module provides background and preparation for the mathematics module
- The **mathematics** module teaches students the content of proportional reasoning
- The **science** module engages students to apply proportional reasoning to the understanding of the science of water distribution and quality
- The TWD culminates with **English language arts**, wherein students consolidate their learning across the modules by communicating solutions to water allocation problems using data-based argumentation

In terms of actual delivery, the TWD project required a team teaching approach, and collaboration across the four aforementioned disciplines. The project was carried out in two schools, with one using the PFL+ approach, and the other a more traditional curriculum. Students were assigned a pre, and post-test designed to assess data literacy skills in the following areas:

- Creating a data-based argument
- Analyzing a data-based argument
- Calculating a per capita measure and using it effectively in an argument
- Identifying data needed to create an argument

The research team found that students using the PFL+ approach outperformed those in the regular curriculum in all four of the above areas. It was found that students could apply mathematical reasoning to help them answer difficult social studies projects, and use this knowledge to analyze more general data literacy questions. An accumulation of benefits effect was also observed, in that the data literate skills built in the modules had a positive impact on later work, e.g. teachers reported a discernible improvement in quality of student's essays in the last module, wherein essays utilized effective evidence (data) based argumentation.

**Citation:** Wanner, A. (2015). Data literacy instruction in academic libraries: Best practices for librarians. *Archival and Information Studies Student Journal*, 1, 1-17. University of British Columbia. Retrieved from <http://ojs.library.ubc.ca/index.php/seealso/article/view/186335>

**Theme(s):** 21st Century Skills and Literacies, [Barriers to Effective Data Literacy Instruction](#), [Delivery and Assessment](#), Teach the Teacher

**Contribution:**

Wanner's peer-reviewed article examines existing literature related to data literacy from both producer (research data management, and consumer (functional use) perspectives, and focuses on two questions while reviewing publications: What is data literacy and how does it differ from its counterpart, information literacy? and what might a data literacy curriculum look like in post-secondary institutions?

Wanner recognizes that many competencies listed in ACRL's *Framework for Information Literacy for Higher Education* (2015) and *Characteristics of Programs of Information Literacy that Illustrate Best Practices: A Guideline* (2012), apply to the general data literacy concepts, although they do not account for newer responsibilities, tools, and skills when using data specifically. This would support the proposed idea that data literacy education could be incorporated into information literacy courses already established in the curriculum, with some changes and targeted integration.

5 studies examined:

- Information and Statistical Literacy in Sociology-Stephenson and Caravello course on centred around critical thinking and information evaluation of data literacy. Outcomes: develop the ability to read and critically evaluate simple 2x2 or 3 way tables; produce accurate bibliographic citations for data tables; use American Factfinder to create a table, which they could describe and cite correctly; and read an article containing a graphical representation of data and discuss the table in relation to the article content (530-531)
- Data literacy in geoinformatics-Carlson, Fosmire, Miller and Nelson course around needs assessment of faculty from science and engineering departments. Used ACRL standards to develop course through integration. Core competencies include: introduction to databases and data formats, discovery and acquisition of data, data management and organization, data conversion and interoperability, quality assurance, metadata, data curation and reuse, cultures of practice, data preservation, data analysis, data visualization, and ethics (including citations) (652-653)
- Science data literacy-Qin and D'Ignazio based course on findings in gaps in science faculty data, using a range of strategies based on faculty needs; metadata and data management; case studies

from many different disciplines; and “authentic activities” based on real world use in their field. Three modules: fundamentals of science data and data management; managing datasets in aggregation; and broader issues in science data management (197).

- Geospatial and data curation in geoinformatics- Fosmire and Miller course based in theory and experience with tools using a hands-on approach with case studies and lab modules. Course topics included: basic computing environments, geospatial data, geographic information systems (GIS), data gathering, independent statistical procedures, and scientific workflow tools.
- Data literacy instruction framework-Calzada and Marzal course based on data use and management. Modules: understanding data; finding and/or obtaining data; reading, interpreting, and evaluating data; managing data; and using data.

Best practices recommended based on common themes and initiatives within higher education:

- Critical thinking skills: ability to transfer classroom learning into practical experiences. Also includes problem-solving to produce well rounded 21st century citizens. Second most mentioned theme
- Collaboration between library and departments: these allow broad centralized development creating a feedback loop promoting refinement as needed
- Librarian involvement: studies found that many faculty members were not competent enough to teach these skills-resulting in the need to educate the educators, and librarians are the best suited for this task, as well as ease the pressure from faculty
- Ongoing data literacy instruction at all levels of schooling: repeated exposure is necessary to build on concepts and situations. Cross curriculum incorporation and varied activities are most effective

Overall, ongoing instruction is integral to student learning, and should include multiple opportunities to practice their skills - “repeated contact with material, over the period of several courses (to several years) with progressively more complex material that is presented in different ways.” (14)

**Citation:** Wing, J. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of The Royal Society*, 366, 3717-3725. doi:10.1098/rsta.2008.0118

**Theme(s):** 21st Century Skills and Literacies, [Data Literacy Best Taught At The Commencement of Post-Secondary Studies](#), [Delivery and Assessment](#)

**Contribution:**

Wing introduces the concept of Computational Thinking (CT) as requiring people be attuned to science, technology, and society, because the fundamental concepts are solving problems, designing systems, and understanding human behaviour. It is driven by scientific questions, technological innovation, and societal demands; combining these factors make the field unique from science, math, and engineering, with a special emphasis on analytical thinking. These three drivers produce a push and pull loop where “scientific discovery feeds technological innovation, which feeds new societal applications; in the reverse direction, new technology inspires new creative societal issues, which may demand new scientific discovery” (3722).

CT is increasingly relevant in today’s society, because it is the transformation of statistics, which are in abundance in every field. Abstraction is an essential part of CT; deciding what details are important enough to include in layers of inquiry, and what is not, underlies CT. This type of thinking will help us solve the Big Data Dilemma, allowing us to fine tune simulation models and asking new questions.

Wing discusses the optimal time to teach CT is in elementary and high school. These levels provide the best environment for instilling foundational skills in problem-solving. Universities are expecting these skills at commencement, and are equipped to hone them in post-secondary and graduate studies. Wing recommends that first year undergraduate student should be taught with focus on the principles of computing, not programming. Computers are simply tools, and can be manipulated efficiently with CT, but the tool can get in the way of learning the concepts if teachers are not careful. Furthermore, formal and

informal learning is encouraged when teaching CT, to provide a diverse learning environment, and individual motivation.

**Citation:** Wright, S., Fosmire, M., Jeffryes, J., Bracke, M., & Westra, B. (2012). A multi-institutional project to develop discipline-specific data literacy instruction for graduate students. *Libraries Faculty and Staff Presentations*. Retrieved from [http://docs.lib.purdue.edu/lib\\_fspress/10](http://docs.lib.purdue.edu/lib_fspress/10)

**Theme(s):** [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

This PowerPoint slide-deck generated by Sarah Wright et al. provides an overview of an Institute of Museum and Library Studies (IMLS)-funded project to develop discipline-specific data literacy instruction for graduate and undergraduate students. The end goal was to use the experience gained from the project to develop a model for other academic to create their own data information literacy programs. In this case, data information literacy was considered the recognition of researchers as producers of data, as well as consumers of data.

The project was divided into five teams from the participating universities (Cornell, Minnesota, Oregon, and two teams from Purdue University). The teams were composed of a data librarian, subject librarian, and faculty researcher.. Each university team was focused on delivering a data literacy program for a specific science, technology, engineering, or mathematics (STEM) discipline:

- Cornell - natural resources
- Minnesota - civil engineering
- Oregon - ecology
- Purdue 1 - agricultural and biological engineering
- Purdue 2 - electrical and computer engineering

The teams interviewed faculty, students, and staff from their respective universities to garner insights and key themes regarding data literacy that could then be used to design a suitable data information literacy program. Three main themes/core skills regarding data literacy emerged from interviews:

- **Data management and organization**
  - There was a noticeable lack of formal training in data management at the graduate student level in relation to research practices
  - Most learning on data management by graduate students done on an ad-hoc/word-of-mouth basis
- **Data continuity and re-use**
  - Data continuity defined as the skills and ability to package data so that it can continue to be used after a graduate student leaves a project
  - Was found to be a high priority among faculty researchers
- **Metadata and data description**
  - Metadata key, but many faculty and students were confused, or had difficulty grasping the concept
  - Intentional and formalized metadata is critical for proper data management

The participants of the interviews further identified other areas of data literacy that they required for their research/students:

- Best practices for sharing data
- Addressing access and ownership of data
- Documentation of data
- Understanding external, and developing internal metadata
- Utilization of data repositories

Each university team came up with different methods of teaching data literacy based in the need of their

universities

- **Cornell - Workshops + Half-Semester course**
  - Three Fall semester workshops offered by the library covering an introduction to data management, relational databases, and data documentation
  - Fall semester course on Special Topics in Data Information Literacy offered to natural resources graduate students
- **Minnesota - Online Course**
  - Seven practical modules, with the end outcome being the creation of a data management plan
- **Oregon - Readings + Workshop**
  - The Oregon team took an **embedded approach**, and worked directly with a research team from the ecology department
  - Three readings were given to the research team to read
  - One informal workshop was then carried out focused on aspects of data management, storage, data sharing, metadata, and data citation
- **Purdue 1 - Workshops**
  - Team one carried out three one-hour workshops covering data management, metadata, and data continuity
  - Workshops included pre, and post-workshop homework in order for students to apply their skills
- **Purdue 2 - Skills Sessions**
  - Purdue 2 also took an embedded approach, working with graduate research teams to assist in designing their data management plans
  - The team also developed skills sessions for undergraduate students focused on aspects of data information literacy (the presentation does not clarify the specific skills taught)

The project teams found that regardless of method used, assessment of student achievement and ability were crucial.

**Citation:** Wyner, Y. (2013). A case study: Using authentic science data for teaching and learning of ecology. *Journal of College Science Teaching*, 42(5), 54-60.

**Theme(s):** [Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Delivery and Assessment](#)

**Contribution:**

Wyner's peer-reviewed article focuses on Master level students in science education, with a goal to make connections between data and human impact on daily life and sustainability. This is widely integratable across science disciplines and more - it is geared towards teachers, but can also be modified to reflect the complexity, accessibility, and accuracy issues of data, especially in the media-conveying the version most appropriate for their story-what is omitted, changed, held constant, what are the strengths and weaknesses, etc.

The study consisted of 14 students, with 7 partners. There was an array of concepts: food webs, biodiversity, carbon cycle, communities, and predator-prey relationships. Students were to relate them to the everyday life of humans, and recommend ways to live more sustainably through initiatives such as reducing sewage, carbon footprint, over fishing/hunting, and offer alternatives. It is proposed that the distillation of data forced students to analyze how to make it meaningful while keeping it accurate. It is designed to help learners read and interpret data, as well as explore, explain, and present it in various formats; learning through real-life data helps students connect procedure to practice, and relating it in a broader way; this approach encourages students' curiosity and independent thinking, as well as hypothesize reasons and courses of action

Wyner highlights the American Museum of Natural History in New York, as a helpful resource in pursuing this type of learning. It provides resources from k-12, with lesson plans for teachers to integrate into classrooms, where students were directed to this website, and were to develop a lesson plan based on areas of interest, using one publication with data to support it. This was followed by an in-class exercise for pre-college students, and then complete a post presentation reflection on effectiveness of lesson and the role that data played - students also complete a pre and post course response to describe and then modify lessons. This website is especially useful, because it includes recent, general interest media pieces, and can be easily searched to find interesting, applicable articles that will engage students.

Overall, relevant, real-world, data-based learning is an increasingly important approach, due to the data-driven decisions in the 21st century; science is not a theoretical based discipline, it is based in evidence and data to support claims

### Grey Literature

**Citation:** Bresnahan, M., and Johnson, A. (2014). Building a Unified Data and Information Literacy Program: A Collaborative Approach to Instruction. [PowerPoint]. Portland, OR: University of Colorado Boulder. [doi.org/10.6084/M9.FIGSHARE.1119760](https://doi.org/10.6084/M9.FIGSHARE.1119760)

**Theme(s):** 21st Century Skills and Competencies, Data Literacy Competencies and Skills

#### Contribution:

Bresnahan and Johnson's presentation focuses on integrating data literacy into an established information literacy program through 5 key principles:

- Scholarship is a conversation: "recognize that they are often entering into the midst of a scholarly conversation, not a finished conversation" (20)
- Research as inquiry: "Engage in informed, self-directed learning that encourages a broader worldview through the global reach of today's information technology" (21)
- Authority is constructed and contextualized: "Identify markers of authority when engaging with information, understanding the elements that might temper that authority" (22)
- Searching as exploration: "Are inclined to discover citation management features, moving them from searching for information to information management strategies" (23)
- Format as a process: "Decide which format and most of transmission to use when disseminating their own creations of information" (24)

The authors notes that the ACRL *Framework* "threshold concepts" align with the education efforts surrounding data literacy, and offer examples of appropriate assignments that highlight this skill overlap:

- Retract Watch involves students in asking questions and investigating sources to judge why the article was good or bad, explaining why;
- students develop a research question based on literature, and investigate if supporting data exists in repositories, or if data from separate sources can be combined,
- and 3 more

**Citation:** Carlson, J., Fosmire, M., Miller, C.C., and Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *Libraries and the Academy*, 11(2), 629-657. DOI: 10.1353/pla.2011.0022

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment, Teach The Teachers

#### Contribution:

This article written by Carlson, et al., focuses on their study conducted by MIT and Syracuse University under a joint initiative entitled Data Information Literacy (DIL). They propose integrating data literacy with

information literacy to bridge the widening educational gap in science. This course would require a multi-faceted teaching approach, and targeted to the skill level of the students. The course was provided to graduate students

Methodology: interviewed faculty about important issues about research data to identify gaps, and asked for recommendations on programming. These responses resulted in a list of core competencies to cover in this course:

- *Introduction to Databases and Data Formats*: Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline. Understands which formats and data types are appropriate for different research questions.

- *Discovery and Acquisition of Data*: Locates and utilizes disciplinary data repositories. Not only identifies appropriate data sources, but also imports data and converts it when necessary, so it can be used by downstream processing tools

- *Data Management and Organization*: Understands the lifecycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets. Creates standard operating procedures for data management and documentation.

- *Data Conversion and Interoperability*: Becomes proficient in migrating data from one format to another. Understands the risks and potential loss or corruption of information caused by changing data formats. Understands the benefits of making data available in standard formats to facilitate downstream use.

- *Quality Assurance*: Recognizes and resolves any apparent artifacts, incompleteness, or corruption of data sets. Utilizes metadata to facilitate understanding of potential problems with data sets.

- *Metadata*: Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others. Develops the ability to read and interpret metadata from external disciplinary sources. Understands the structure and purpose of ontologies in facilitating better sharing of data.

- *Data Curation and Reuse*: Recognizes that data may have value beyond the original purpose, to validate research or for use by others. Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research. Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle. Articulates the planning and actions needed to enable data curation.

- *Cultures of Practice*: Recognizes the practices, values, and norms of his/her chosen field, discipline, or subdiscipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.

- *Data Preservation*: Recognizes the benefits and costs of data preservation. Understands the technology, resource, and organizational components of preserving data. Utilizes best practices in preservation appropriate to the value and reproducibility of data.

- *Data Analysis*: Becomes familiar with the basic analysis tools of the discipline. Uses appropriate workflow management tools to automate repetitive analysis of data.

- *Data Visualization*: Proficiently uses basic visualization tools of discipline. Avoids misleading or ambiguous representations when presenting data. Understands the advantages of different types of visualization, for example, maps, graphs, animations, or videos, when displaying data.

- *Ethics*: including citation of data. Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing data. Appropriately acknowledges data from external sources.

Offers a useful MIT data planning checklist covering topics: documentation and metadata, security and backups; directory structures and naming conventions; data sharing and citation; data integration; good file formats for long-term access; and best practices for data retention and archiving

Refers to a syllabus through Syracuse University, and provides a link to the document, which includes details of weekly topic, activities, and readings:

Module 1: Fundamentals of science data and data management

1. Introduction to course; activity: *pre-course assessment survey*; science data life-cycle-collect, store, retrieve, use, and present;

2. Introduction to database, database programs and database attributes
3. Data relationships; activity: identifying data sets, and in-class presentation: share your information analysis of data repository/dataset; Fundamentals about data forms, scales, types, levels: data structure and models-physical data and model data; activity: presentations continue, and quiz on data and database fundamentals
4. Data formats: standards, representation of data, communication of data markup languages and meta formats; Describing datasets: intro to metadata, aboutness, type, elements, schemes, standards; activity: in-class group work to investigate CSDGM profiles
5. Describing datasets: metadatas' role in resource management; Describing datasets: metadata elements in schemes and tools; Case study; activity: in class group work: form use for data description
6. Describing datasets: relationships-making and relational database lab; activity: in class work describe dataset and create an extension; Managing data: encoding, storing, import/export, cleaning, transformation; activity: quiz data formats and description
7. Managing data: data query and retrieval-SQL lab; activity: practice with SQL statements and data reports; Managing data: ownership and access, data quality

Module 2: Managing data in aggregation

8. Managing data: applying XML to data and metadata; activity: querying data and metadata; Data aggregation scenario: research collection
9. Managing aggregation scenario: reference collection; Managing aggregation scenario: resource collection; activity: XML for metadata and data records
10. Data and users: requirements interviewing faculty about data management need and practices analysis; Data and users case study; activity: in class group work
11. Organizational planning: goals and objectives, procedures, quality control; Organizational planning: Metadata issues, long-term preservation; activity: report to class findings from interviews

Module 3: Broader issues in science data management

12. Enabling technologies: organizing and managing data, storing and retrieving, using data; activity: report on interview result; Understanding data curation: metadata description, quality criteria, archival concepts
13. Data repositories and discovery: directory services, controlled vocabularies; activity: quiz using data; Data analysis: data mining and meshing
14. Data presentation: visualization, tools, formatting for publication; Sharing data: ethics, publishing, citations
15. Project presentations and discussion; Wrap up; activity: post-course assessment survey

Refers to ACRL and recommends using standard to develop the course, because of similar concepts:

- a. determine nature and extent of information need - research question and methodology for analysis
- b. access needed information effectively and efficiently - familiar with converting formats, including resolution and timescales
- c. evaluate information and its sources critically and incorporate selected information into his or her knowledge base and value system - reputable, quality control, relevancy and compatibility, authority, and appraise metadata
- d. use information effectively to accomplish a specific purpose - communicate and choose appropriate information technologies; planning and creation of a product
- e. understand many of the economic, legal, and social issues surrounding the use of information; access and use information ethically and legally - correct citations of datasets

**Citation:** Cowan, D., Alencar, P., and McGarry, F. (2014). Perspectives on open data: Issues and opportunities. *IEEE*, IEEE Conference on Software Science, Technology and Engineering. DOI: 10.1109/SWSTE.2014.18

**Theme(s):** 21st Century Skills and Literacies, Data Literacy as the Ability to Understand and Use Data Effectively to Inform Decisions, Data Literacy Competencies and Skills



**Contribution:**

This technical report by Donald Cowan et al. looks at issues regarding open data, and provides practical examples in an attempt to illustrate both issues and opportunities. It also puts forth a partial research agenda for open data based on these examples. Some of the issues covered in the paper are also transferable to data literacy.

Open data has the potential to unlock new opportunities across sectors. The benefits are wide reaching, and include:

- Enhanced transparency and accountability of the government and agencies that release data;
- Efficiency and improvements in Public Service delivery;
- Enhanced inspection and collection of data through increased citizen engagement; and
- Creation of economic and social value

However, open data is published in multiple areas (federal, provincial, municipal), which is costly and can make data difficult to find and compile. Aggregating and publishing data can be time consuming and costly as well. Problems exist today with access to and use of open data for ecological/geospatial issues, and require significant human and technical resources to address them. These problems will only become more widespread and difficult to address if something is not done now concerning environmental open data.

Therefore, competencies regarding open data should be developed within research and academic institutions. These include:

- How to find and access open data;
- How to provide and use the right tools to work with open data;
  - Software adaptation for big data analysis
- Ensuring privacy of individuals and property;
- Sustaining the cost of storage, delivery, and maintenance of open data;
- Interdisciplinary collaboration and critical thinking;

Creating an Open Data Registry would go a long way to mitigating perceived challenges regarding open data. Such a registry could provide supply location, related metadata/ontologies, gateways supporting access, and examples of how to use open data analytics.

**Citation:** Giles, J. (2013). Fostering a data-driven culture. *The Economist*, 1-22. Retrieved from [https://www.tableau.com/sites/default/files/whitepapers/tableau\\_dataculture\\_130219.pdf](https://www.tableau.com/sites/default/files/whitepapers/tableau_dataculture_130219.pdf)

**Theme(s):** Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions

**Contribution:**

Giles' article explores the challenges in nurturing a data-driven culture, delves into trends and highlights best practices, and what companies can do to meet them. Using two main sources: survey, 530 senior executives from diverse backgrounds; and a series of indepth interviews with four leading industry experts.

Results showed that forward thinking companies are integrating data into their everyday operations, and using it to make decisions. There is evidence that links financial performance and the successful exploitation of data. 43% of survey respondents ranked data as "extremely important" to strategic decision making, which is higher than any other unit. Companies are also recognizing the benefits of incorporating data into their accounting, marketing, communications, and recruiting strategies.

"Many of my clients are clearly aware of the importance of data, but they don't know where to start in terms of where they should focus to get the most value, as well as how to translate the data into actionable insight."

- Jerry O'Dwyer, Principal Deloitte Consulting

Giles posits that encouragement to collaborate and share data has helped top companies generate a data-driven culture in their organization. Increased availability to training is another factor; empower employees with data training to become more data literate, can increase engagement and buy-in. Although, data specialists will always be essential to successful data-driven culture, because predictive analytics requires an in depth knowledge of statistics. These individuals are difficult to recruit and retain because they are in high demand.

Key findings from this study include: data is a resource and sharing it is key; training is essential to full utilization; data collection must be a primary activity cross departments; and buy-in from the top is integral. Industry leaders must be open to counter-intuitive theories and unorthodox strategies, as long as they are supported by data, which empowers employees to be more engaged and contribute more fully

**Citation:** Gold, A. (2007). Cyberinfrastructure, data, and libraries, part 2. *D-Lib Magazine*, 13(9/10), 1-12.

**Theme(s):** [Delivery and Assessment](#)

**Contribution:**

Gold's article focuses on social science, geo-references (GIS), and bioinformatics; where all three areas have a commonalities in the way of dedicated data centers, most manipulation takes place using widely adopted and supported commercial or open source software, and extensive training is offered regularly by national organizations or enterprises. The growing reliance on data in these fields present an opportunity for libraries and librarians to be more involved in the data processes and education. Libraries are increasingly becoming more like labs than the traditional warehouses or repositories of materials, because reuse of data in research is becoming a fundamental priority. This requires understanding and documentation of data's provenance, the development of ontologies, expert annotations, and analysis - downstream it require visualization, simulation, data mining, and other forms of knowledge representation and extraction, which may spur a change in values and practices in this profession, and more collaboration with non-library organizations and professionals to develop programs in partnership.

Domain expertise is essential when working with researchers, because it enables the effective flow of information throughout different up and downstream phases. These skills can be developed through several ways, including professional conferences, reviewing key documents relating to data science, formal and informal training programs, and continually working to understand perspectives, practices, and culture (lifelong learning). Additionally, marketing data consultancy and referral services will help professionals and researchers understand the support available to them through knowledgeable representatives, understand and support technologies and systems of data publishing and reuse, and advise and advocate for systems, standards, and issues, as well as promoting interoperability.

**Citation:** Gunter, G. A. (2007). Building student data literacy: An essential critical-thinking skill for the 21st century. *MultiMedia & Internet@Schools [H.W.Wilson - EDUC]*, 14(3), 24-28.

**Theme(s):** [21st Century Skills and Literacies](#), [Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions](#), [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

Gunter argues that in today's society, there is a perpetual struggle to provide students with objectives and strategies to develop skills to be competitive in the new global economy. Education is too focused on skills for seeking information, not on evaluating and analyzing it. Students must be empowered with critical-thinking skills to be successful today. These skills should be taught across the curriculum to maximize the impact and applicability to various situations. Gunter refers to the *Framework for 21st Century Learning*, which identifies learning and thinking skills as one of the 6 core components of the framework, especially

noting critical-thinking and problem-solving as essential.

To prepare students for the workforce, there is an increasing necessity for data literacy, because of the value added to outcomes and decision making in most industries. The abilities to view, manipulate, analyze, and interpret data are expected of students entering the workforce, and critical thinking and problem-solving lend to these skills, as well as understanding how to use new technologies to maximize the opportunities in data. Gunter offers *InspireData* as a tool that allows students to question, explore, solve problems, draw conclusions and create visualizations using data in a variety of disciplines.

**Citation:** Hu, X. (2012). Session summary: The RDAP12 panel on training data management practitioners. *Bulletin of the American Society for Information Science and Technology*, 38(5), 29-30. Retrieved from [http://www.asis.org/Bulletin/Jun-12/JunJul12\\_MayernikSessionSummary.pdf](http://www.asis.org/Bulletin/Jun-12/JunJul12_MayernikSessionSummary.pdf)

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

This paper by Xiao Hu provides a session summary from the RDAP12 Panel on Training Data Management Practitioners from the the ASIS&T Research Data Access and Preservation (RDAP) Summit held in March 2012. The summit discussed education and training for data managers, and data literacy was covered as a critical skill.

In terms of curriculum, George Mason University, Rensselaer Polytechnic Institute, Cornell University, and the Syracuse University iSchool are among the leaders offering programs in data management and science.

- Key curriculum topics include:
  - Web science and information technology
  - Discipline oriented informatics
  - Data science
  - eScience project planning
  - IT competency
  - Librarianship (vague on what this exactly entails)
  - 'Soft skills', i.e. collaboration and communication

Teamwork was identified as a key component of data literacy training/education, as team-based projects are common in the workplace. Data literacy courses and workshops must also strike a balance between theory, and hands-on practice in order to be effective. Moreover, the closer the applications and exercises are to real world examples, the easier it is to conceptualize data literacy in practice.

**Citation:** Hunt, K. (2004). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly* 28(2), 12 -16. Retrieved from [http://www.iassistdata.org/sites/default/files/iq/iqvol282\\_3hunt.pdf](http://www.iassistdata.org/sites/default/files/iq/iqvol282_3hunt.pdf)

**Theme(s):** Delivery and Assessment, Teach The Teachers, Barriers to Effective Data Literacy Instruction

**Contribution:**

This article written by University of Winnipeg librarian Karen Hunt, discusses the results of a project carried out by the library to integrate data literacy into the curriculum of an existing course. The author puts forth recommendations based on her experience for the successful extension of data literacy programs in other post-secondary institutions.

According to the author, best practices teaching for data literacy can potentially be built from the ACRL's

best practices for information literacy. The author makes a point of this by substituting 'data literacy' into four points from the ACRL's best practices to illustrate the baseline competencies for data literacy::

- Articulate the integration of 'data literacy' across the curriculum;
- Accommodate student growth in skills and understanding throughout the college years;
- Apply to all learners, regardless of delivery system or location; and
- Reflect the desired outcomes of preparing students for their academic pursuits and for effective lifelong learning.

The actual project came about following a request from a faculty member. As a relatively small institution (12 librarians for approximately 8000 students in 2004), the author acted as both information literacy, and data librarian. Knowing that the author had this expertise, an instructor from the geography department approached her to help come up with a plan that would get "students into the data" (p. 13) The author argues that integrating data literacy directly into the research methodologies of specific subjects allows for stronger connections to be made, and applicable examples to be used, and thus took this approach (as opposed to a series of workshops or a stand-alone course open to all-comers).

The author and the instructor designed an assignment for the instructor's Human Geography course that required students to use United Nations Human Development indicators as a base for creating their own development indicators for provinces in Canada. The assignment also required students to construct their own population pyramids for two small Canadian towns. The assignment involved finding, retrieving, and manipulating data using Microsoft Excel. Involvement in designing and delivering the course was as follows:

- Librarian's involvement:
  - Developed the assignment; initially for a small class
  - Held labs for the class
  - Was available for tutoring and questions
- Instructor's involvement:
  - Created the shape of the assignment
  - Fielded questions from students
  - Lectured on development and population distribution
  - Ultimately responsible for assignment evaluation

The author took away a number of lessons based on the experience of carrying out this project. A main conclusion is that students learn best when the data literacy curriculum is relevant and builds on previously learned skills and knowledge. Opportunities for making connections and practicing is also important (e.g. homework). Moreover, instructors cannot assume students know how to use technology, spreadsheets, etc., and must design any course or workshop to take into account different levels of experience and skills using statistics. Staff training (both within libraries and faculty departments) is also important in order to make data literacy across curriculums scalable. On the more high-level, the author states that a unified terminology and definition on data literacy and its competencies is required for cross-university collaboration, and that leadership must come from organizations such as the ACRL to ensure that this happens.

The author puts forth a number of high-level recommendations, including:

1. Decide on one term and agree upon a definition of data literacy;
2. Codify data literacy learning outcomes
3. Endorse and promote the standards for data literacy; and
4. Articulate best practices for data literacy programs.

**Citation:** Kenney, A.R. (2014). Leveraging the liaison model: From defining 21st century research libraries to implementing 21st century research universities. *Ithaka S+R*. Retrieved from <http://www.sr.ithaka.org/blog-individual/leveraging-liaison-model-defining-21st-century-research-libraries-implementing-21st>

**Theme(s):** 21st Century Skills and Literacies, [Delivery and Assessment](#)

**Contribution:**

This article from *Ithaka S+R* by Cornell University Librarian Anne R. Kenney looks at the role of academic libraries within the context of 21st century universities, and what this means in terms of service delivery and new roles for librarians. Specifically, the author argues that academic libraries should focus on thinking about what kind of universities will succeed in the 21st century, and directing service delivery and development along this line. The academic library liaison model is an option put forth that is well-suited to requirements of students and faculty in this changing context.

Over the past decade the liaison model has developed as full-time reference and collection-development positions shifted to more expanded portfolios. These portfolios have increasingly focused on engagement and outreach to students and faculty. In this regard, liaison librarian roles include scholarly communication, digital tools education, data curation, managing research workflows, and promoting data driven-scholarship. In terms of data literacy education, liaison librarians are thus well-positioned to lend support and pair up with disciplinary experts and functional specialists, as well as regular faculty to teach data specific courses. Libraries could also provide technical support (e.g. providing systems to help researchers create data management plans), equipment, and/or lab space for data literacy courses (especially for departments which may not usually have access to said technology).

Besides assisting in the teaching of data literacy, academic libraries should also quantify their own service delivery, and partner with organizations on campus that work and manage data in order to improve their own operations (i.e. practice what you preach).

**Citation:** Koltay, T. (2014). Big data, big literacies? *TEMA*, 24, 3-8.

**Theme(s):** [21st Century Skills and Literacies](#), [Data Literacy Competencies and Skills](#), [Delivery and Assessment](#)

**Contribution:**

Koltay argues that Big Data is no longer an exclusive issue for the natural sciences, but it is present in the social sciences, the humanities, and arts and culture. It is becoming more relevant to the workplace in most industries, making data literacy more essential than ever before. Future students and workers must be able to translate vast amounts of data into abstract concepts, as well as understand data-based reasoning. Furthermore, knowledge and skills on how to verify the provenance of data is key, i.e. critical examination and assessment of data, and data sources. It is recommended that Information Professionals (e.g. librarians, information managers, etc.) take the lead in terms of scholarly communication and teaching regarding information and data literacy, showcases their relevance and skills to students and faculty alike.

Data literacy can be defined as a set of skills and abilities related to accessing research data, critically evaluating and using it. In other words, understanding, using, and managing data; it accommodates both the data producer and data consumer's viewpoints. Competencies can include:

- Discovery and acquisition of data
- Data management
- Data conversion and interoperability (i.e. dealing with risks, and potential loss or corruption of information by changing data formats)
- Quality assurance
- Metadata
- Data curation and re-use
- Data preservation
- Data analysis
- Data visualization
- Ethics, including citation of data

Koltay identifies data literacy as a possible sub-discipline of information literacy, but also brings attention to the similarities to science literacy, such as, methods, approaches, attitudes, and skills relating to critical thinking. Questions are provided for the reader to review in association to what skills and thought processes are required by graduate students in regards to answering key questions:

- Who owns the data?
- What requirements are imposed by others (e.g. funding agencies or publishers)?
- Which data should be retained?
- For how long should data be maintained?
- How should digital data be preserved?
- Are there ethical considerations?
- What sort of risk management is needed for research data?
- How are data accessed?
- How open should the data be?
- What alternatives to local data management exist?

These competencies can further enhance data literacy:

- Ability to collaborate and work in teams
- Familiarity with scientific data sources
- Familiarity with quantitative research methods
- Knowledge of general metadata standards
- Data structure of digital objects
- Ways to assess digital object's authenticity, integrity, and accuracy over time
- Storage and preservation policies, procedures, and practices
- Relevant quality assurance standards
- Risks associated with information loss or corruption of digital entities

**Citation:** Liqueste, V. (2012). Can one speak of an "Information Transliteracy"? *International Conference: Media and Information Literacy for Knowledge Societies*. Moscow, Russia. Retrieved from <https://hal.archives-ouvertes.fr/hal-00841948>

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

Liqueste's article focuses on the idea of transliteracy as essential in the 21st century, which is centred on a radically ecological position, consisting constantly in questioning one's own actions and the influence of environments (technical, organizational, informational) on oneself (self-analytic attitude). Four positions are presented for consideration:

1. "Assessment" encompasses more than the nature of information, content, and results, but also the entire *process* of content production and the chain of activities that lead to the content in question, and the evaluation of process to identify the real process at work
2. Evaluate the overall potential of the socio-economic information environment, particularly in learning and the workforce
3. Familiarity with operative procedures, going beyond the simple stage of reacting to a stimulus. This is dependent on digital objects and available tools, and adopting an analytical stance to technical features available
4. Maintain a "cognitive distance" from the immediate results offered by information systems such as engines. Evaluate the results and quality of "answers" to questions

These positions make it clear that critical thinking and a focus on process, not information regurgitation, are essential in today's market. Deep thinking and problem-solving are required for successful integration of students and workers into a knowledge-based economy. The article also highlights eight "meta" skills that are becoming increasingly important:

- “Comprehension and understanding of information systems, a sort of “information understanding”, where the stakeholder is himself able to perceive the various types of information systems to weigh them up, to identify their value, and use the right attitudes to them
- “Information knowledge” i.e working declarative knowledge related to information and the dissemination of existing tools. The challenge of this MS is to possess the vocabulary of expression and representations linked to them. Our work shows that the media, technical devices, tools for processing information etc. are not conceived of in the same way by individuals, undoubtedly causing all sorts of misunderstanding and misconceptions
- Procedural knowledge related to technical issues (or “information applications”) where the goal is basically to be able to use effectively and efficiently the main technical tools in order to meet a need and perform a task
- The ability to assess the informational potential of the environment or the technique used (or “information potential”). It is clear that the individuals interviewed take for granted the potential of a system more than they really test it, and often discover belatedly the offers and features available to them. Strengthening the use and integration of new technological media requires this ability to project oneself and to appraise one’s own strengths and weaknesses
- “Actional” strategies oriented to the organization and perpetuation of one’s memory of one’s work. Transliteracy aims at adopting procedures for processing personal content for later use in new professional and/or learning situations
- The “ability” to stand back from one’s own daily, and sometimes even mechanical and systematic, reception of information. Several studies show that media users eventually get locked into multiple repetitions without discerning what could be done differently (an effect called the “tunnel” effect). Stepping back means that information may be received otherwise and new techniques can be used that are flexible and not repetitive.
- The technologies and technical devices are calling more and more sensory, physical and optic clues. Alan Liu (2012) points out, for example, the impact of the visual culture in defining and understanding the informational transliteracy. Schools, universities and companies will likely have to reinforce training and help in the identification and control of the sensory cues and physical media spaces at our disposal.
- Anthropocentric and consists in the assessment of how to identify and characterize one’s own cognitive styles. Indeed, to what extent are we dependent or not on the field of technology and media? Do we respond individually by impulsivity or reflectivity? Do we centre our gaze or rather scan during reading on the screen, etc.” (5-6)

**Citation:** Mackey, T.P., & Jacobson, T.E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, 72(1), 62-78. doi: 10.5860/crl-76r1

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

This article by Thomas P. Mackey and Trudi E. Jacobson focuses on social media and participation in online communities using collaborative Web 2.0 technologies. The authors argue that the emergence of these technologies and dynamic online communities has changed the traditional definition of information literacy. Due to information taking many forms online, in order to fully understand and collaborate using Web 2.0, information literacy must be taught in conjunction with other literacies, and should be reframed as the foundation for metaliteracy.

According to the authors, metaliteracy promotes critical thinking and collaboration and provides a comprehensive framework to effectively participate in social media and online communities, as well provide a basis for lifelong learning. It is critical for individuals to have this comprehensive understanding of information in order to critically evaluate, share, and produce content.

Data literacy is not mentioned specifically, but could be assumed to be included among the wider literacies

that make up metaliteracy. Furthermore, competencies for putting metaliteracy into practice are also directly applicable to data. The competencies put forward by the authors are as follows:

- Understand format and delivery mode
  - Information seekers must be able to determine extent of information (or data) required, as well as the format and delivery mode of said information (or data)
- Evaluate user feedback as active researcher
  - Critical thinking is key in order to continuously filter through what information (and data) is usable and unusable
- Create a context for user generated information
  - Information (including data) is now created and released by multiple users and/or organizations.
  - Requires an increased (and often ongoing) evaluation on information sources, including metadata
- Evaluate dynamic content critically
  - Users must have the ability to synthesize disparate forms of information (including data)
- Produce original content in multiple formats
  - Producers of digital content must be able to make critical choices about the format/tools to articulate and explain ideas
  - Must also be aware of the transferability of content/data
- Understand personal privacy, information ethics, and intellectual property issues
  - Proper attribution of sources is key (e.g. data citation)
  - Requires an ongoing exploration of legal, economic, political, and social issues that mediate our access to technology define the types of data we use
- Share Information in participatory environments
  - Metaliteracy encourages collaboration in the development and distribution of original content in synchronous and asynchronous online environments

**Citation:** MacMillan, D. (2014). Using open access resources in data literacy instructions: Renewing the IL curriculum by aligning it with changing needs. *Library Instruction West 2014*. Paper 24.  
[http://pdxscholar.library.pdx.edu/liw\\_portland/Presentations/Material/24](http://pdxscholar.library.pdx.edu/liw_portland/Presentations/Material/24)

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

This PowerPoint presentation by Don MacMillan, Liaison Librarian at the University of Calgary, provides an overview of the the library's efforts to renew their existing information literacy (IL) program, and incorporate IL education into courses in biology and biochemistry.

The author points out that data-intensive disciplines such as the sciences and social sciences naturally benefit from data literacy competencies. Along these lines, the author worked with instructors to create enquiry-based lab exercises using real-world solutions and domain repositories. This ensured authentic pathway activity that replicated research workflow, and emphasized interoperability.

When replicating these types of exercises, the desired outcomes for learning should be:

- Innovative learning experience wherein data integrates into and informs content
- Students able to manage and analyze their data more efficiently
- Students able to find real world solutions to research questions;
- Students able to prepare presentations demonstrate deeper understanding of subjects

The author also identifies a number of best practices that emerged from the project. Exercises must be integrated well with other course content and worth a percentage of a student's grade. A module or scaffolded approach to teaching competencies is also useful, with sequential steps going from simple to complex. Students also seem to learn best by hands-on activity. Interactivity and flexibility are crucial, and



data sources should be tailored to specific questions, Especially important is consistent instruction across all lab sections, as teaching assistants (TAs) do not always have the requisite expertise.

The PowerPoint also provides a list of tools appropriate for each discipline with the hope of integrating them into other courses in the future. Appendix C on page 26/7 further provides marking rubrics for presentations.

**Citation:** Mandinach, E.B, Parton, B.M., Gummer, E.S., and Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, 96(5), 25-28. Retrieved from <http://web.b.ebscohost.com.ezproxy.library.dal.ca/ehost/pdfviewer/pdfviewer?sid=c0f93557-337c-4e0f-96eb-24b20658a586%40sessionmgr110&vid=1&hid=106>

**Theme(s):** Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Teach The Teachers

**Contribution:**

This article by Ellen B. Mandinach et al. looks at the issue of teachers using data to guide teaching policy and decisions made in the classroom, with specific focus put on the ethical use of said data. The authors argue that teachers must be familiar themselves with how to use data ethically in order to effectively make decisions, and work with data in the classroom. There are related points of interest regarding data literacy that are worth mentioning.

As outlined by the authors, data literacy can be considered the the ability to transform information into actionable knowledge and practices by collecting, analyzing, and interpreting all types of data. It is integral that data utilization in this manner fall within ethical bounds, including aspects of data security and confidentiality. It also relates to how data is presented, visualized, and described, i.e. not misrepresenting data. In terms of data literacy instruction, ethics is a critical component, and should be included as a module and/or key theme.

**Citation:** Martin, E., and Leger-Hornby, T. (2012). Framework for a data management curriculum: course plans for data management instruction to undergraduate and graduate students in science, health science, and engineering programs. Retrieved from [http://library.umassmed.edu/data\\_management\\_frameworks.pdf](http://library.umassmed.edu/data_management_frameworks.pdf)

**Theme(s):** Delivery and Assessment

**Contribution:**

Martin and Leger-Hornby's study focuses on a course that incorporates undergraduate and graduate students from science, health science and engineering programs from University of Massachusetts Medical School. The course consists of seven modules and four case studies

1. overview of research data management
2. types, formats, and stages of data
3. contextual details needed to make data meaningful to others
4. data storage, backup, and security
5. legal and ethical considerations for research data
6. data sharing and reuse policies
7. plan for archiving and preservation of data

Learning objectives, lecture content, activities, assessment, and readings are also available, but with such a focused course on only a few of the competencies, I did not see the reason for incorporating this, but they are available from the link provided above. Although, it does provide a template that can be adapted to curricula

**Citation:** McAuley, D., Rahemtulla, H., Goulding, J., and Souch, C. (2010). *How Open Data, Data Literacy and Linked Data Will Revolutionize Higher Education*. Pearson Centre for Policy and Learning, 88-93. Retrieved from [http://pearsonblueskies.com/wp-content/uploads/2011/05/21-pp\\_088-093.pdf](http://pearsonblueskies.com/wp-content/uploads/2011/05/21-pp_088-093.pdf)

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions

**Contribution:**

This paper by Derek McAuley et al. theorizes on the impact of open data, Big Data, and linked data on higher education. The authors state that data literacy within higher education is essential for the growth of innovative data usage and repurposing.

Data Literacy as defined by the authors is “the ability to identify, retrieve, evaluate, and use information to both ask and answer meaningful questions.” (p. 89). The opportunity for discovery and optimizing of raw data through mechanisms such as open and Big Data will only continue to grow, the so-called ‘data fusion mechanisms’. Therefore, educating students on how to distinguish between good and bad data critically is of paramount importance. Higher learning and post-secondary institutions must implement policy encouraging release of data in linkable forms, offer educational programs with data literacy applications and resources, and engage students.

**Citation:** Mitrovic, Z. (2015). *Strengthening the Evidence-Base for Open Government in Developing Countries: Building Open Data Capacity Through e-Skills Acquisition*. Mitrovic Development and Research Institute. Retrieved from <http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper3.pdf>

**Theme(s):** 21st Century Skills and Literacies, Barriers to Effective Data Literacy Instruction,

**Contribution:**

Mitrovic’s article focuses on the necessity to educate the general public with e-skills, e-awareness, ICT skills, technological literacy, information literacy, media literacy, and data literacy, relating to Open Government Data (OGD). These skills overlap in core principles, lending to the learning process, and will decrease the ‘data divide’; an unequal disbursement of skills will result in issues concerning socio-economic reach of initiative, meaning a special push must be made to communities in poverty.

It is noted that in 2014, the United Nations report stressed the need for a data literate world population. The OGD lends to this need, because it is a multinational initiative with data freely available to access, use, reuse and publish without restriction, and promotes transparency and accountability for governments.

Overall, Mitrovic’s study indicated most students and academic professionals believed that they had insufficient skills needed to access, and use OGD or e-skills. This includes determining context, organizational social structure, task, and technology, using who, what, where, why, whom, and how. The most relevant skill is *data literacy* because it provides the foundation of an innovative knowledge economy. The School of Data offers online courses to people who already possess e-Literacy Requisite skills, that are widely accessible for various levels of learning.

**Citation:** Mooney, H., and Carlson, J. (2014). *Developing Data Literacies for Graduate Students in the Social Sciences*. Libraries Faculty and Staff Presentations. Paper 51. Retrieved from [http://docs.lib.purdue.edu/lib\\_fsres/51](http://docs.lib.purdue.edu/lib_fsres/51)

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

Mooney and Carlson's PowerPoint presentation centres around interviews done with faculty and students, and argues that teaching data literacy in alignment with disciplinary cultures and local practices is key to successful graduate level education.

They outline 12 competencies that are essential:

1. Data processing and analysis
2. data management and organization
3. data preservation
4. database and data formats
5. ethics and attribution
6. data quality and documentation
7. data curation and re-use
8. data conversion and interoperability
9. data visualization and representation
10. discovery and acquisition
11. metadata and data description
12. cultures of practice

**Citation:** Qin, J., and D'Ignazio, J. (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Librarians, 31st Annual Conference*, Paper 5. Retrieved from <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>

**Theme(s):** Barriers to Effective Data Literacy Instruction, Delivery and Assessment, Data Literacy Competencies and Skills

**Contribution:**

This report by Jian Qin and John D'Ignazio provides an overview of the Science Data Literacy (SDL) Project that occurred over a two-year period from May 2007 to May 2009 at Syracuse University (SU) iSchool. The main goal of the project was to assess the needs for scientific data literacy education by collaborating with the science, technology, engineering, and mathematics (STEM) faculties, and using information gathered to design a data management course.

Data literacy as defined by the authors is the ability to understand, manage, and use (science) data. SDL education has the two main goals of training students to become e-science data literate. and become e-science data management professionals (i.e. skilled in collecting, processing, managing, evaluating, and utilizing data). The SDL Project **identified two gaps** that the SU iSchool could address:

1. What is required of the next generation workforce to manage a new type of information resource (i.e. data), and the corresponding cyberinfrastructure?
2. How aware is the STEM faculty of their own shifting expectations of the skills required by individuals to assist them in research (i.e. research assistants), as well as information support (e.g. from the university library)?

The authors created a full-year data management course, open to both undergraduate and graduate students. The course utilized a module-based approach, and was designed to represent several different research fields. The course included student presentations on specific case studies, quizzes, regular and guest lectures and corresponding readings, interview practice, group work (undergraduates paired with graduate students), and a multiple stage project based work being done by faculty members. Content included:

- Data types
- Data formats
- Data management principles
- Relational databases, E-R diagram creation

- XML schema
- Science data lifecycle
- Metadata and data citation

The authors conducted an evaluative survey at the beginning and end of each semester. Pertinent student feedback included:

- Positive:
  - Databases and websites are very useful knowledge; case studies were easier to learn through group study; project work offered real world experience
- Negative:
  - Highly technical terms and jargon; readings were overwhelming; exercises were ambiguous; no background knowledge in databases made it very difficult to learn

The authors noted some barriers and challenges. Graduate students who were most suited for the course were bogged down with other courses, and found it difficult to fully commit. The interdisciplinary nature of the course also caused challenges, as case studies had to apply to all students. SDL education should be provided at different levels via different venues and be adaptive to students' skill levels.

**Citation:** Romani, J.C.C. (2009). Strategies to promote the development of e-competencies in the next generation of professionals: European and international trends. Publisher's version. Retrieved from <http://ora.ox.ac.uk/objects/uuid:da0007a3-b504-4c20-858b-21dd359e3cae>

**Theme(s):** 21st Century Skills and Literacies, Teach The Teachers, Delivery and Assessment

**Contribution:**

Romani's article focuses on primary and lower secondary level education, but is applicable to higher education in the grand scheme. The article was written in response to OECD referring to the mismatch of skills taught in school and workplace demand, in e-competencies and ICT, affecting employability of graduates in this knowledge-based economy. There is an increased demand for highly-qualified graduates able to perform cognitive, analytical and interactive complex tasks upon entrance to the workforce. It should be noted that knowledge of ICT is not enough, students must be able to apply this knowledge in diverse ways to varying situations, therefore requiring diverse teaching methods.

Romain studies the benefits of adopting ICT as instruments to improve the learning process, to bridge this gap, and current trends that are likely to impact e-competency development in policies, strategies, and programmes in European society. This study uses 16 international case studies from Europe, North America, Israel and Columbia 2001-2009, to identify trends related to impact of ICT in school, with the hope of improving student preparedness entering the workforce. A review was conducted of 16 studies concerning ICT use and results, which produced an inconclusive conclusion of 10 points (p.10-11). Generally, ICT training is informal, independent learning is key, teachers primarily using for administrative purposes, in class use is not driver of success, lack of coordination between adoption of technology and teaching-learning strategies.

The "British Qualifications and Curriculum Authority (2008) report that some of the abilities required by teachers are: Identify problems and defining tasks; Searching and selecting information; Organizing and structuring information; Analysing and interpreting information; Combining and refining information Modeling; Controlling events and devices; Exchanging information; Presenting information; Controlling events and devices; Exchanging information; Presenting information; Reviewing, testing and evaluating and assessing the impact of ICT" (40). Reskilling the teachers to increase the impact of ICT learning and outcomes is essential to the success of any initiative. This can be done through self-learning and informal peer-learning, which are of the most significant approaches to obtaining ICT skills, and the overarching ability to create, connect, enrich and transfer knowledge among people. Table 6 provides a list of competence standards for teachers.

Romani argues that new technology being adopted by educational institutions is not enough, the teachers must adopt new teaching and learning paradigms. Some approaches that are proven to improve the learning process include project-based learning, real-world problems, self-learning activities, collaborative and interdisciplinary learning, constant acquisition of new competencies and knowledge transfer (33). Digital devices have been known to increasingly stimulate other 21st century literacies, for example soft skills to support creativity, innovation, experimentation, problem-solving, collaborative work, and critical thinking (34). There has been found a strong association between students' performance and the use of ICTs at home (OECD, 2005), which highlights the role of parents and the family (and independent motivation/more comfortable at home); it is not the hours of technology used in the classroom, but the motivation to use it in diverse ways outside of the classroom that promotes learning with technology. There is evidence showing that independent learning outside of the classroom is conducive, but the perception that "digital natives" are competent is not necessarily true. Self evaluation ranks them higher than reality in assessment of e-skills.

The article posits that the definition and underlying concepts must be identified to create a useable standard for ICT literacy; this can be done through reviewing international standards to help in this development, such as the European e-Competence Framework and the European Computer Driving Licence Foundation. These are both focused on creating a framework, standard and certification based on context; it is also noted that standards and practices should continuously be evaluated and improved.

Five underlying concepts of e-competency are listed:

- *e-awareness* (understanding of relevance of ICT in society),
- *technological literacy* (ability to interact with hardware and software, as well as apply, communicate and manage applications),
- *media literacy* (how traditional and new media are merging; understanding the nature of media; skills to identify, judge, and discriminate content and services),
- *digital literacy* (proficiency to build new knowledge based on the strategic employment of ICTs; critical, creative, and innovative thinking combined and empowered with information management skills; access, retrieve, store, organize, manage, synthesise, integrate, present, share, exchange, and communicate), and
- *informational literacy* (ability to understand, assess and interpret information from a variety of sources)

"E-competencies are a set of capabilities, skills and abilities to exploit tacit and explicit knowledge, enhanced by the utilization of digital technology and the strategic use of information" (43). This article also provides definitions of ICT specialist, advanced users, basic users, ICT practitioner skills, ICT user skills, and E-business skills, for reference when implementing targeted changes to education.

Stakeholders must be involved with the training of these skills, because context and practice are important to implementing the correct level of education to prepare graduates for the workforce, and these people and organizations can provide perspective to initiatives and tools. Also increases engagement in the system for both future employers and employees. Involving these groups in the discussion allows for collaboration that works toward a set of common goals that could be influential in informal, industry-based, and government supported initiatives. Research and evaluation of initiatives should be addressed regularly, and align with complexities of the knowledge-based-economy.

**Citation:** Sapp Nelson, M., Zilinski, L., and Van Epps, A. (2014). Developing Professional Skills in STEM Students: Data Information Literacy. *Libraries Faculty and Staff Scholarship and Research*. Paper 85. DOI:10.5062/F42V2D2Z [http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1087&context=lib\\_fsdocs](http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1087&context=lib_fsdocs)

**Theme(s):** 21st Century Skills and Competencies, Data Literacy Competencies and Skills, Data Literacy Best Taught At The Commencement of Post-Secondary Studies

**Contribution:**

Sapp Nelson, et al., provide a variety of competencies based on context

- Technical competencies relating to data consumption for entry level engineer positions include:
  - identify problems through data collection and analysis
  - apply logical processes to analyse information and draw conclusions
  - identify inconsistent or missing information
  - critically review, analyse, synthesize, compare, and interpret information
- Competencies related to discovery and acquisition of data, data conversion and interoperability, and data quality and documentation
- Related competencies that were particularly appropriate for an undergraduate audience include:
  - ask a question and find a dataset that will have the data required
  - understand that there are fields within datasets
  - understand what the fields within a dataset mean
  - understand relationships between fields within a dataset
  - develop a question based on the data in a dataset
  - read and interpret charts, graphs, and other data visualizations

The authors also provide the reader with a Data Credibility Checklist, which includes important factors in recognizing quality in a data object

- Documentation- is there a content map or guide of some sort? What is covered? what is not covered? is there metadata included?
- Authority- who created the data? who is managing it? who paid for the data? what bias might be implicit? is the data object currently maintained? are there any references on how this data object have been used in the past? are there clear release versions and updates information?
- Format expectations- are there clear format expectations? what units are used? what fields are present? what naming conventions are used? are the dates of creation or last update easily located
- Quality control- is quality control explicitly outlined? who is in charge of checking for quality? what process do they use? how is missing data handled?
- Human readable/machine readable- can a file be opened and a user understand the content? is the file available for download in an open format? is there a clear process to download?

<b>Content map</b>	<b>Authoritative</b>	<b>Format expectations</b>	<b>Quality control</b>	<b>Human readable/Machine readable</b>
What is covered?	Who created the data?	What units are used?	Who is in charge of checking for quality?	Can you open a file and understand what is in it?
What is not covered?	Who is managing the data?	What fields are present?	What process do they use?	Is the file available for download in an open source format?
Is it relevant to	Who paid for the	What naming	How is missing	Is there a clear

my research question?	data?	conventions are used?	data handled?	process for download?
How is it relevant to my research question?	What bias might be implicit?	Are the dates of creation or update easy to find?		
Is there metadata included?	Is it currently maintained?			
	Has someone else used this data object for reuse in the past? How?			
	Are there clear release versions, updates with release dates?			

**Citation:** Scheitle, C.P. (2006). Web-based data analysis: creating a sociological laboratory. *Teaching Sociology*, 34(1), 80-86.

**Theme(s):** Barriers to Effective Data Literacy Instruction, Delivery and Assessment, Data Literacy Competencies and Skills

**Contribution:**

This paper by Christopher P. Scheitle looks at how to encourage students in university sociology courses to “think sociologically”, by applying theory to real-world applications (Scheitle, 2006, p. 80). Although there are traditional methods to encourage this style of learning in-class, the author argues that they are typically limited to student experience and do not allow for the exploration of larger issues and research. To overcome this, the author puts forth the idea of using sociological laboratories, using analysis of quantitative data to supplement traditional course activities and assignments.

Data analysis in the social sciences at the undergraduate level is often left out of curriculum. This is due to a lack of knowledge on the part of students of quantitative data, difficulty in obtaining and using advanced tools, and due to students feeling that this content is boring. However, the author argues that the use of web-based data analytics can negate some of these barriers (Consortium for Political and Social Research (ICPSR), SegMaps, US Census Bureau Data, etc.). Using these tool as opposed to data analytics software packages or database management systems allows for greater accessibility, and many are free (negating financial cost as a barrier). Web-based data analytics programs often also simpler interfaces, which lessens the learning curve for students new to working with data. By working with data, students can effectively translate the ‘chalk and talk’ of classrooms into more real terms, e.g. briefing notes on issues.

To this end, Scheitle also includes examples of web data-based assignments.

**Citation:** Stout, A., and Graham, A. (2007). The Data Dilemma. *ASEE Annual Conference*. Retrieved from <http://dspace.mit.edu/bitstream/handle/1721.1/39640/amystout.pdf>

**Theme(s):** Barriers to Effective Data Literacy Instruction, Data Literacy Competencies and Skills

**Contribution:**

This conference report by Amy Stout and Anne Graham looks at how the increasing proliferation of data has caused a paradigm shift in how researchers manage and utilize data. With so much data and information being disseminated and collected, it is more important than ever to improve skills related to information retrieval. This includes knowledge of metadata, sophisticated databases and search engines, as well as legal and social issues regarding data sharing and ethical use of data.

The authors interviewed faculty at MIT in 2005 in order to garner their perspective on data and how academic libraries could best assist them in terms of data management and curation. Relevant themes that emerged were:

- Technical barriers to data storage and organization
- Haphazard metadata
- Privacy issues and loss of control
- Time and effort needed to manage data effectively;
- Hope that librarians will provide commercial databases (knowledge of relational databases)
- Need for earlier access to datasets
- Perceived need for centralized rather than distributed storage
- Data curation, and retention

The authors argue that librarians are skilled data managers who can provide leadership in the above area, but must be aware of the social, technical and legal issues that often hinder data creation and management (and data itself). Some questions and issues for them to consider when thinking about how to provide data literacy related service or training are:

- What are the benefits of storing, preserving, and sharing your data?
- What happens to research data when researchers leave the institution?
- Issues surrounding Open Access, Creative Commons, and copyright
- Technical issues surrounding data storage
- Preservation issues
- How to access data created by other researchers
- Metadata creation, schema, and the Semantic Web

**Citation:** Swan, A., & Brown, S. (2008). *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practices and Future Needs*. Key Perspectives LTD: Truro, UK.

Retrieved from

<http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf>

**Theme(s):** 21st Century Skills and Literacies, Barriers to Effective Data Literacy Instruction, Data Literacy Competencies and Skills, Delivery and Assessment, Data Literacy Best Taught At The Commencement of Post-Secondary Studies

**Contribution:**

This report by Alma Swan and Sheridan Brown was prepared for Jisc United Kingdom (UK), a registered charity aimed at championing the use of digital technologies in research and education. The report was based on two recommendations from *Dealing With Data: Roles, Rights, Responsibilities and Relationships* (Lyon, 2007) reflecting the need for a study to examine the role and career development of data scientists, as well as the value and potential of offering training to undergraduate and postgraduate students. The report contains useful perspectives on the future of data science and the related data literacy competencies and teaching.

The report is based mostly on a qualitative analysis of fifty-seven semi-structured interviews and four focus



groups from across the UK, as well as online survey.

Respondents indicated a lack of formal training data training in education, and there was agreement that data literacy skills should be taught early to undergraduate students. However, there has been push-back from ut institutions arguing that their curriculum is too full to add more training. That said, data literacy competencies could be integrated into complementary skills (e.g. relational database training). Formal skills assessment is not as favored as practical assessment. Continuous training and life-long-learning is also important due to the rapidly changing nature of data. Therefore, discipline-targeted, intensive, short-term courses are the prefered method of formal learning. Cross-disciplinary workshops could facilitative learning, and build on strengths and experiences of different disciplines. Major barriers to creating new courses are time, financial resources, and full curriculum schedules.

In terms of competencies, respondents agreed teaching should tie in with fundamentals such as statistics, laboratory practices, methods, and recording findings. This includes:

- Relational databases
- XML
- Principles of data curation
- Documenting work
- Task tracking

Librarians can be integrated into the teaching of these fundamental skills either formally or informally. However, in order to do so, there is a need to educate librarians more comprehensively in data science, as academics are expecting them to be familiar with skills that are not necessary taught in library school.

Despite a preference for data training at the undergraduate level, there are potential benefits to instruction at the post-graduate level:

- Could help researchers understand the importance of the data lifecycle and their role in generating, handling, curating, archiving, and preserving data successfully
- May provide the impetus for some researchers to go on to become data scientists
- Will equip new data scientists with the basic knowledge they need to get started

The UK Data Archive has further developed a set of training materials consisting of 8 modules that could be useful for teaching data literacy skills:

- Developing research consent agreements
- Anonymous data gathering techniques
- Data description and metadata
- Data formats and software
- Copyright and IPR
- Data storage, backups, and security
- Digitization
- Providing access to data

**Citation:** Twidale, M.B., Blake, C. and Grant, J. (2013). Towards a data literate citizenry. In Schamber, L. (Ed.). *iConference 2013 proceedings*, 247-257. Retrieved from <https://www.ideals.illinois.edu/bitstream/handle/2142/38385/189.pdf?sequence=4>

**Theme(s):** [Delivery and Assessment](#), [Barriers to Effective Data Literacy Instruction](#), [Data Literacy Competencies and Skills](#), [21st Century Skills and Literacies](#), [Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions](#)

**Contribution:**

This article by Michael B. Twidale, Catherine Blake, and Jon Gant focuses on building data literacy within civil society in order to improve citizens' engagement in the democratic process, and to understand and

participate in data-driven decision-making processes. The article contains relevant points relating to data literacy instruction and related competencies, barriers to data literacy, and tools for data literate users.

The authors argue that debates on societal and policy issues are themselves informed by analyses of data, as are other debates on priorities, finance, resource allocation, health care, climate change, etc. Therefore, it is important that citizens are able to understand data and statistics, how this data can be used to inform decisions, and how to differentiate between 'good' and 'bad' representations of data. At its base level, comfort with using computational technologies, i.e. basic computer literacy, is an important precursor to being data literate, as many sources of data or tools for interpreting data rely on computers and the requisite software. Moreover, due to the overlap of issues regarding data, data literacy is also intrinsically linked to other literacies such as statistical and information literacy.

Links to information literacy are especially important, and could serve as a baseline from which to develop delivery and content for teaching data literacy. For example, the Big6 approach developed by Eisenberg & Berkowitz (2011) on teaching information literacy to all ages and grade levels could be readily applied to teach data literacy:

#### **1. Task Definition**

- 1.1 Define the information problem
- 1.2 Identify information (data) needed

#### **2. Information Seeking Strategies**

- 2.1 Determine all possible sources
- 2.2 Select the best sources

#### **3. Location and Access**

- 3.1 Locate sources (intellectually and physically)
- 3.2 Find information within sources

#### **4. Use of Information**

- 4.1 Engage (e.g., read, hear, view, touch)
- 4.2 Extract relevant information

#### **5. Synthesis**

- 5.1 Organize data from multiple sources
- 5.2 Present (data visualization) the data

#### **6. Evaluation**

- 6.1 Judge the product (effectiveness)
- 6.2 Judge the process (efficiency)

The authors put particular emphasis on the importance of metadata instruction (i.e. what is it, how is it used, what can it tell a user) in regards to teaching data literacy. Areas that could serve as core concepts/modules further include:

- Bibliometric search
- Data mining
- Text mining
- Data visualization
- Claims analysis

In regards to tools that help users understand terminologies and methods, the use of 'popular' or accessible examples/case studies when teaching data literacy can help citizens/students connect to the data, and bring higher concepts down to a more understandable level. Barriers to data literacy include math and computer phobia, and misconceptions about data and statistics. Misconception analysis can be useful in overcoming these and other barriers.

**Citation:** Womack, R. (2014). Data Visualization and information literacy. *IASSIST Quarterly*, 38(1), 12-17. Retrieved from <http://dx.doi.org/doi:10.7282/T3X92CZF>

**Theme(s):** Data Literacy Competencies and Skills, Delivery and Assessment, Data Literacy Best Taught At

## The Commencement of Post-Secondary Studies

### Contribution:

This article by Ryan Womack focuses on the importance of data visualization as a basic component of information and data literacy. The article goes on to make recommendations for delivery of data visualization content at the postsecondary level, and possible curriculum design.

Data visualization represents a major tool for the communication of complex results from large and often heterogeneous data sources, is a key competency of data information literacy (and could even be considered another type of literacy alongside statistical and information literacy). Owen et. al. (2013) classify data visualizations into three areas: 1) scientific or data visualization, in which the data dimensions correspond to physical reality (e.g., remote sensing); 2) information visualization, for multidimensional data from a defined field of interest; and 3) visual analytics, which is massive and heterogeneous.

The author asserts that most existing data management courses are taught at the graduate level, focusing on training students methods to handle and present their data and findings. Although useful, teaching data management (and data visualization techniques) this late in a student's educational career limits their data visualization skills to very specific, discipline-focused areas. Instead, data visualization should be taught early on (e.g. first year) at the postsecondary level. This capitalizes on students' still being new to research methodology, and instills good practices as they build their general study habits.

ACRL Information Literacy Competency Standards for Higher Education 3 and 4 are most relevant in terms of teaching data visualization at a post-secondary level:

- 3. The information literate student evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system
- 4. The information literate student, individually or as a member of a group, uses information effectively to accomplish a specific purpose.

Focus should be placed on creating an Intellectual/theoretical framework for data visualization (built off these two standards), as the tools and technology used for data visualization will continue to change rapidly. Data visualization curriculum should further focus on three areas:

- **Evaluation**
  - Refers to the ability to establish quality, accuracy, and reliability of a data visualization
  - Students should be able to 'interrogate the image' and find out the source of the data, reliability of the source, appropriateness of the visualization for that kind of data (e.g. Tufte's Lie Factor test).
  - Understanding the methodology behind the data visualization is also key
- **Critique**
  - Critique involves comparison among different data visualizations in order to develop understanding of which visualizations exemplify best practices. General principles such as striving for clarity, avoiding clutter, and emphasizing the most relevant data apply to most visualizations.
- **Use**
  - Emphasis is on putting data visualization into practice.
  - Guiding students through introductory examples, working in sandbox environments, and using various demos and examples will lead students through the process of actually developing their own visualizations based on the choices before them.
  - The actual form that the 'Use' component takes will be determined by current technology and research needs in a particular setting.

As a concluding example from the literature, Kelleher and Wagener (2011) offer 10 simple guidelines that apply to almost any kind of data visualization.

**Citation:** Yeh, K., Xie, Y., and Ke, F. (2011). Teaching computational thinking to non-computing majors using spreadsheet functions. 41st ASEE/IEEE Frontiers in Education Conference: Session F3J. Rapid City, SD.

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills

**Contribution:**

This paper by Kuo-Chuan (Martin) Yeh, Ying Xie, and Fengfeng Ke, discusses the ability of non-computer science students at recall, application and problem-solving, and how to best maximize computational thinking (CT) learning potential due to its importance in society and the workforce where tasks are growing in scope and complexity. CT is a key 21st century skill that is useful for facilitating data literacy learning and skills.

CT is defined as “an ability or skill to analyze a problem, create abstraction, and solve it effectively” (p. 1). Characteristics of CT are:

- Automation of abstractions: focuses on the ability to manage complex situations by generating abstractions and maintaining the relationships among them
- Precise representations: to generate abstractions, we need to have formal representations that reflect our cognitive processes and structures (discerning aspects of the situation)
- Systematic analysis: enable us to generate hypothesis and search for a plausible solutions systematically
- Repetitive refinements: during problem solving, we consistently evaluate the current situation against our previous experience or our prediction until the best solution is reached (p.1)

The authors carried out a survey of 126 undergraduate non-computing major students who took a required computer literacy course (introduction to spreadsheets and databases) as part of their major. Students were recruited while taking their course, with participation being voluntary (albeit participants received 2% extra credit). The survey was designed to motivate participants to use their problem-solving and analysis skills through Excel (with no outside help i.e. online tutorials). Questions included:

- **Recall:** Participants were asked to explain the purpose/meaning of a function and its arguments(s)
- **Application:** Participants were asked to use data and a particular function to generate a correct answer. They were cued by the data available for which function they should use.
- **Problem-solving:** Participants were asked to choose functions freely to solve two problems, with no cue as to function

Results showed that when the complexity of a task increased, performance would decrease. This proved the hypothesis of the authors that recall of purpose and syntax of a function was simple, but the use of a function proved much more difficult. Common errors included wrong type of argument, missing arguments, wrong logic statements, and wrong function syntax. Also students' data representation was weak, especially when differentiating between types of data, and function syntax.

Strategies that could promote learning include problem-based learning or case-based reasoning. These could help students internalize the CT skills and individualize their problem solving styles.

**White Papers**

**Citation:** ALA. (2000). *ACRL Information Literacy Competency Standards for Higher Education*. Retrieved from <http://www.ala.org/acrl/standards/informationliteracycompetency>

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

ALA defines information literacy as “set of abilities requiring individuals to ‘recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information” (2). This is increasingly important due to technological advancements and available information resources. This standard forms the basis for lifelong learning, and is common to all disciplines, environments, and levels (as is data literacy). Implementing this standard in institutions should be a collaborative effort between professors, faculties, and administration.

An information literate people should be able to:

- determine the extent of information needed
- access the needed information effectively and efficiently
- evaluate information and its sources critically
- incorporate selected information into one’s knowledge base
- use information effectively to accomplish a specific purpose
- understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally (2-3)

A comprehensive table is provided with seven standards (listed above), as well as corresponding performance indicators and learning outcomes.

**Citation:** ALA. (2011). *ACRL Visual Literacy Competency Standards for Higher Education*. Retrieved from <http://www.ala.org/acrl/standards/visualliteracy>

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

ALA in association with ACRL developed a set of standards for helping implement this into university and colleges to help students develop competencies for academic and professional futures. These standards can be adapted fully or partially and according to disciplinary need. They are presented in a linear way, but can be employed as needed. Partnerships and shared implementation strategies across departments is helpful in successful education: faculty, librarians, curators, archivists, visual resources professionals and learning technologists.

In an interdisciplinary, higher education environment, a visually literate individual is able to (2)

- determine the nature and extent of the visual materials needed
- find and access needed images and visual media effectively and efficiently
- interpret and analyze the meanings of images and visual media
- evaluate images and their sources
- use images and visual media effectively
- design and create meaningful images and visual media
- understand many of the ethical, legal, social, and economic issues surrounding the creation and use of images and visual media, access and use visual materials ethically

A comprehensive table is provided with seven standards (listed above), as well as corresponding performance indicators and learning outcomes.

**Citation:** Association of College and Research Libraries. (2013). Working Group on Intersections of Scholarly Communication and Information Literacy. *Intersections of Scholarly Communication and Information Literacy: Creating Strategic Collaborations for a Changing Academic Environment*. Chicago, IL: Association of College and Research Libraries. Retrieved from <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Intersections.pdf>

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills

**Contribution:**

This white paper by the Associations of College and Research Libraries (ACRL) Working Group on Intersections of Scholarly Communication and Information Literacy looks at scholarly communication and information literacy, and provides an overview of intersecting literacies (i.e. digital, data, and transliteracy). Although the article focuses on changes required to scholarly communication in libraries due to the evolution of information literacy, there are some relevant points relating specifically to data literacy.

Data literacy centers around “understanding how to find and evaluate data, the version of data being found and used, who is responsible for it, how to cite it, and the ethics of data procurement” (p. 10). Although competency standards and teaching programs for media and visual literacy are focused on undergraduates, key questions about teaching data literacy tend to focus on graduate students and faculty. Specifically the curation, presentation, and interpretation of research data has increased in importance due to the ease and proliferation of data, thus requiring new approaches. Students, both as users and as future creators of data, should be trained to understand how their choices affect access, reuse, and preservation. Other competencies of importance include:

- Use of data management and visualization tools
- Reuse of content in diverse ways not imagined by the creator;
- Data ownership and rights
- Data as preservation and curation

Librarians can be key teachers in this regard, and should be involved in developing ways to handle data issues, such as determining who should have access, how that access can be managed (given the wide variety of formats and technologies involved), and what steps will be necessary to ensure that these data collections remain available over time. The report concludes that data literacy is an area where the impact of external forces, ranging from increasing demand on students to find and use data to funder mandates to have data management plans, point to a critical area of intersection between scholarly communication and information literacy.

**Citation:** Association of College and Research Libraries. (2014). Top trends in academic libraries. *College & Research Libraries News*, 75(6), 294-302. Retrieved from <http://crln.acrl.org/content/75/6/294.full.pdf+html>

**Theme(s):** Data Literacy Competencies and Skills, Teach The Teachers

**Contribution:**

This white paper looks at the top trends in academic libraries for 2014, as identified by the Association of College and Research Libraries (ACRL) Research Planning and Review Committee. The document is part of a series released on an annual basis. The paper covers a variety of trends, with an overall theme of deeper collaboration between academic libraries and other university actors. Data literacy is identified as one key trend, and there are some relevant take-away points.

Libraries, IT services, administration, and grant support will increasingly have to collaborate in order to provide the expertise necessary for data management support as part of the academic research process. In order to facilitate this, a **shared vocabulary for data literacy** is required. This would also make it easier to develop strategies/plans for teaching data literacy competencies that are truly transferable.

In terms of current efforts for teaching data literacy, more and more universities are creating graduate programs, or certificate programs centered on preparing professionals for analysis and manipulation of big data, e.g. the Institute for Advanced Analytics at North Carolina State University, and graduate certificate in data mining offered by Stanford University are placing demands on libraries for cross-disciplinary expertise in data collection access, metadata, curation, and preservation. Other key areas

where libraries can lead the way in terms of preparing students to work with data are:

- Teaching students the skills, as well as the resources that enable them to verify, re-use, and cite data correctly.
- Teaching students to use tools to manipulate, clean, and transform data, e.g. Open-Refine

may also need to become more familiar with aspects of data literacy, in order to develop and provide robust support to students and faculty.

**Citation:** Benito, E.M. (2009). Learning.com joins the Partnership for 21st century skills. *Business Wire*. Retrieved from [https://global.factiva.com/ha/default.aspx#!?&\\_suid=143231383276103258974098134786](https://global.factiva.com/ha/default.aspx#!?&_suid=143231383276103258974098134786)

**Theme(s):** 21st Century Skills and Competencies, Delivery and Assessment

**Contribution:**

The Partnership for 21st century skills brings stakeholders together in business, education, and policy. It recommends that skills building should be infused with tools and resources to help facilitate and drive change. Six key elements of 21st century education are identified:

1. Core Subjects
2. 21st Century Content
3. Learning and Thinking Skills
4. Information and Communication Technology (ICT)
5. Life Skills
6. 21st Century Assessment

**Citation:** Doucette, L., and Fyfe, B. (2013). Drowning in research data: Addressing data management literacy of graduate students. *ACRL*. Retrieved from [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe\\_Drowning.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf)

**Theme(s):** Data Literacy Competencies and Skills, Data Literacy as the Ability to Understand and Use Data Effectively to Inform Decisions

**Contribution:**

This paper by Lise Doucette and Bruce Fyfe discusses the authors' research study on the graduate students' level of research data management (RDM) literacy. The research includes an assessment attitude and behaviours, as well as informal and formal data literacy education experience.

The authors pre-empt the discussion of their findings with a brief look at the importance of data management.

The proper management of data including organization, protection, preservation, and sharing are "essential for productivity, securing grant funding, enabling collaboration and ensuring future use of data" (p. 165). Moreover, strong RDM competencies and the ability to create an RDM plan allows for the re-use of publicly-funded research data. Graduate students should be competent in terms of the research data lifecycle:

- Production
- Dissemination
- Long-term management
- Discovery/repurposing

The research project consisted of a survey of graduate students in the social science and science disciplines across Canada. Subject areas included geography, psychology, sociology, chemistry, physics, and earth sciences. Key areas of interest were attitudes and behaviours toward research data management, and literacy and education related to RDM (self-assessment, formal, and information

education). Relevant results were:

- Most respondents agreed that it was important to effectively manage data within their research group
- 90% responders 'confident' in the ability of their data skills;
- Several respondents often had to re-collect data because of lost or unopenable files;
- There was agreement (78%) on the importance of reuse of data, but disagreement (38%) over how to actually carry it out
- 37.8% of respondents did not have written or oral policies related to research data management
- 20.8% of respondents took a research methods course in which RDM was discussed
- 22.3% took another course where RDM was discussed
- 15.4% participated in a workshop where RDM was discussed
- Only 4.7% of respondents discussed the management of research data with a librarian
- 56.1% of respondents agreed or strongly agreed with the statement "I educate myself on best practices for preserving my data"

Based on the results of the survey, the authors made some relevant conclusions. Social sciences, and doctoral students have the highest t-test scores for understanding of data literacy and management. Social science students also had a statistically higher formal education score than science students. Most students ranked their confidence high in data management, but in the proceeding questions prove that their practices are in actuality quite poor, leading to inefficiency, low productivity, and costliness (especially in terms of duplication of data collection and analysis). Further research could focus on this disconnect between confidence in skills and actual lack of data management experience. Further formal education (e.g. courses, workshops, instruction from data librarians, etc.) could serve to close this gap.

**Citation:** Hogenboom, K., Holler Phillips, C.M., and Hensley, M. (2011). Show me the data! Partnering with instructors to teach data literacy. *ACRL 2011*, Philadelphia, Pennsylvania. 410-417.

**Theme(s):** [21st Century Skills and Literacies](#), [Barriers to Effective Data Literacy Instruction](#), [Delivery and Assessment](#), [Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions](#), [Teach The Teachers](#)

**Contribution:**

Hogenboom, et al., in association with ARCL, define data literacy as "the ability to read and interpret data, to think critically about statistics, and to use statistics as evidence" (410). They refer to many perspectives on this topic: Caravello and Stephenson suggest integrating data literacy into an established information literacy course, which would allow students to build on existing and complementary skills. Karen Hunt recognizes that the technical tools and skills required to work with data creates a complication to the integration process in universities, but recommends using Electronic Data Center or the UK Data Archives, which provide datasets to courses ready for online analysis, enabling learning without technical issues. This process would allow for students to become familiar with skills at an accessible level, before learning more technical programming and formatting skills.

On the flip side of beginning slowly, Czarnocki and Khouri recommend introducing equipment and software to students through data-intensive assignments. Others recommend that inclass librarian instruction is helpful in introducing the students to support through data services - collaboration with faculty would enable the library to provide focused learning modules for students needs, and possibly providing subject-based data pages that have already been converted into readily-usable formats; increasing referrals from faculty to the library for instruction in overall critical thinking and skills development would help bridge the gap between full curriculum and intense skill development with support.

University of Illinois at Urbana-Champaign conducted a survey to find how instructors are using data in teaching, demand for software and hardware to support teaching, and prioritizing training for staff. Many thought that the library was the first point of reference for students with data related questions. Frustration



existed primarily in the tools and access to specific data. Many believed that broad training would be helpful, but also specific tool related training would be valuable as well, but in class, traditional training was least favored format, because of limited time and copious amount of information understanding decreases

**Citation:** Johnson, L., Adams Becker, S., Estrada, V., and Freeman, A. (2015). NMC Horizon Report: 2015 Higher Education Edition. Austin, Texas: The New Media Consortium. Retrieved from <http://www.nmc.org/publication/nmc-horizon-report-2015-higher-education-edition/>

**Theme(s):** 21st Century Skills and Literacies, Teach The Teachers, Barriers to Effective Data Literacy Instruction

**Contribution:**

Johnson, et al., posit that universities must foster conditions for innovation to happen: creativity, risk-taking, collaboration can lead to entrepreneurial spirit. systems to be more adaptive with evidence, and data-based decision-making. This white paper outlines key trends, significant challenges and important technological developments likely to impact higher education in the next five years.

- Trends:
  - Policy: measuring learning through data-driven practice and assessment is currently on the rise in universities, and will only become more important in 3-5 years
  - Leadership: collaboration between different higher education institutions-innovation can scale better when ideas are shared between institutions
  - European Commission's goals "stimulating a more open research environment, fostering stronger partnerships with businesses, and rethinking how qualifications are recognized"
  - The non-profit organization Unizin "focuses on interoperability and open standards...facilitating learning analytics with the aim of improving student outcomes."
  - University of Maryland has a predictive analytics reporting framework. Discussion to incorporate data analytics into professional development. Analytics are becoming key to data-driven learning and assessment.
  - Channel 9 provides online learning for a variety of in-demand skills including computer coding and programming through streaming videos, interactive events and more. "Getting and Cleaning Data" video lectures and online quizzes on obtaining data through APIs and databases. More social experience with peer-to-peer evaluation
- Challenges
  - Must educate the educators
  - Creating policies that better advance digital literacy. Key component is finding training techniques that prioritize creativity, and being able to leverage technologies for innovation is vital to fostering real transformation in higher learning. Intention, reflection, and generativity
  - "digital literacy is an iterative process that involves students learning about, interacting with, and then demonstrating or sharing their new knowledge" (24), as well as leveraging digital tools to navigate, evaluate, create, and critically apply information
  - "Complex thinking is the applications of systems thinking, which is the capacity to decipher how individual components work together as a part of a whole, dynamic unit that creates patterns over time. Computational thinking is another higher-order thinking that complements complex thinking, and it entails logical analysis and organization of data; modeling, abstractions, and simulations; and identifying, testing, and implementing possible solutions" (28)
  - Providing a more competency-based curriculum, increasing workforce preparedness
- Technologies to help in learning:
  - consumer technologies, digital strategies, enabling technologies, internet technologies, learning technologies, social media technologies, and visualization technologies

**Citation:** Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. Retrieved

from [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

This industry white paper presents a current issue of Big Data facing society today. It is an issue because there will be a severe shortage, approximately 140,000-190,000, of trained professionals to deal with this. Big Data is only growing due to the increasing number of devices being used, which is valuable only if it is being used. If this shortage is addressed, and professionals can embrace big data it can become key to productivity growth and consumer surplus, although privacy will become a growing issue with society in relation to personal data use, as well as the sharing of data for public welfare.

Seven Key Insights are presented:

1. Data have swept into every industry and business function and are now an important factor of production;
2. Big data creates value in several ways: creating transparency; enabling experimentation to discover needs, expose variability, and improve performance; segmenting populations to customize actions; replacing/supporting human decision making with automated algorithms; and improving new business models, products, and services;
3. Use of big data will become a key basis of competition and growth for individual firms;
4. The use of big data will underpin new waves of productivity growth and consumer surplus;
5. While the use of big data will matter across sectors, some sectors are poised for greater gains;
6. There will be a shortage of talent necessary for organizations to take advantage of big data;
7. Several issues will have to be addressed to capture the full potential of big data: data policies, technologies and techniques, organizational change and talent, access to data, industry structure

**Citation:** McKendrick, J. (2015). Data driven and digitally savvy: The rise of the new marketing organization. *Forbes Insights*. Turn: New York, N.Y. Retrieved from [https://www.turn.com/livingbreathing/assets/089259\\_Data-Driven\\_and\\_Digitally\\_Savvy\\_The\\_Rise\\_of\\_the\\_New\\_Marketing\\_Organization.pdf](https://www.turn.com/livingbreathing/assets/089259_Data-Driven_and_Digitally_Savvy_The_Rise_of_the_New_Marketing_Organization.pdf)

**Theme(s):** Data Literacy Competencies and Skills, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions

**Contribution:**

This Forbes Insight Report looks at the rise and importance of data-driven marketing and decision-making in the business world. The report includes sections covering leadership skills and competencies that are required to engage and succeed in this new data-driven business world. Part of the results of the report were based on a survey to top executives (with salaries of \$500,000+) concerning their thoughts on data-driven marketing and the role of technology and interviews with six executives and thought leaders to provide context.

Data-driven decision-making is central to all successful marketing today. It provides actionable insights from the copious amount of data available, and is critical to answer the 'why' and 'so what' of business. Moreover, marketing "is now an ongoing process of engagement and learning" (p. 5), and the global economy demands data-driven campaigns to be successful in this hypercompetitive environment.

The key findings of the report were:

- Demonstrable Results
  - Data-driven marketing has delivered demonstrable results in terms of customer loyalty, customer engagement, and market growth.
- Commitment to Data

- Committed efforts to use data are difficult due to siloing within organizations.
- Integration
  - Data-driven action requires expertise and innovative thinking from many individuals and teams, integration and interdisciplinary cooperation are key.
- Training
  - Training is essential. Majority of businesses do not offer training and education to develop data-driven marketing skills. Successful enterprises encourage employee development usually through online programming or informal mentoring or coaching

The top industries making use of data-driven decision-making and marketing according to the report are: technology (39%), telecommunications (39%), retail (37%), advertising and/or marketing (22%), the media (19%), travel (13%), banking (10%), consumer goods (10%), automotive (7%), energy (5%).

Using data and analytics allows businesses to carry out targeted customer group campaigns, instead of putting a large amount of resources into one generalized campaign (thereby increasing cost-effectiveness). The most important aspect of a data-driven marketing campaign is integrating data from a variety of sources into a comprehensive picture in order to tell a 'story' and create understanding. Types of data being collected for data driven marketing and decision-making by firms include: customer data, website usage data, CRM data, campaign metrics, social metrics, online transactions, behavioral data, and demographic data. This is expected to grow, as 7/10 executive leaders see reliance of data only increasing over the next three years. As one executive puts it, "[Marketing] Leaders always measure results with analytics, strive to make data-driven decisions and actively transform personnel roles to be more digitally savvy...data moves freely and is consistent across all channels, and is considered trustworthy and timely...[and] is integrated into operations across the enterprise and its partners" (p. 8).

A pronounced organizational trend occurring is the need to bring in new skills sets to be able to manage data collection, analysis, and decision-making process. According to the report, data analytics and critical thinking are highly valued skills, almost more so than basic marketing skills. Integrated thinking, the ability to recognize the different types of data available with a good fundamental understanding of people, is crucial (i.e. human interaction and active engagement).

Important competencies are:

- Input side
  - Knowledge of data collection and analysis tools
  - Understanding of data modelling
  - Critical analysis skills
- Output side
  - Communication and synthesis skills (including the ability to communicate results to a non-technical audience)
  - Ability to comprehend results (i.e. data comprehension) and make decisions based on analysis

Knowledge of specific technologies is also key. Technologies currently employed for data-driven marketing include website analytics (56%), data management platforms (47%), analytics software (38%), marketing automation software (29%). However, the report also states that 17% of executives are unsure or do not know what kind of technology their firms use to carry out data management. This points to a knowledge gap, but also an opportunity for growth. It is essential to hire people who are comfortable analyzing and looking at various type of data, and have a sense of curiosity (i.e. people who can separate the signal, from the noise).

**Citation:** Oceans of Data Institute. (2014). Profile of a big-data-enabled specialist: Executive summary. Waltham, MA: Education Development Center, Inc.

[http://oceansofdata.edc.org/sites/oceansofdata.org/files/ODI%20Panel%20Exec%20Summary\\_FINAL.pdf](http://oceansofdata.edc.org/sites/oceansofdata.org/files/ODI%20Panel%20Exec%20Summary_FINAL.pdf)

**Theme(s):** 21st Century Skills and Literacies, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions

**Contribution:**

The Oceans of Data Institute did a cross industrial analysis skills required by data specialists. They identified essential tasks: defining problems and articulating questions; developing deep knowledge of data sources; developing methods and tools; and staying current on emerging technologies, data types, and methods. They recognize that industries across the board put emphasis on practices such as ethical standards, and protecting the data and results; critically evaluating the results of analysis to determine the level of confidence in the results and estimate the precision and accuracy of answers; and telling a “data story” to convey insights, identify limitations, and provide recommendations based on the results of data analyses.

Overall, the ODI’s key findings included unexpectedly “soft skills” such as analytical thinking, critical thinking, and problem solving dominates the 20+ big data skills and knowledge requirements identified by the panel and endorsed by experts who completed the validation survey; apply statistical methods also ranked among the most important skills, as well as knowledge of algorithms; and among other behaviours, experts and reviewers said that a successful big-data-enabled specialist is willing to question, and is a lifelong learner, a seeker of patterns, open minded, and curious.

**Citation:** Oceans of Data Institute. (2014). Profile of a big-data-enabled specialist. Waltham, MA: Education Development Center, Inc. **OR** Ippolito, J., and Malyn-Smith, J. (2014). Profile of a big-data-enabled specialist. *Ocean’s of Data Institute*. Retrieved from <http://oceansofdata.org/our-work/profile-big-data-enabled-specialist>

**Theme(s):** 21st Century Skills and Literacies

**Contribution:**

The previous executive summary is in relation to this document, which lists skills in, knowledge of, behaviours, equipment/tools/supplies, trends/concerns, and duties of a data specialist. This document can be very useful when creating a course with the goal of preparing students to enter the workforce with usable and employable skills.

**Citation:** OECD Skills Outlook 2013. Retrieved from [http://skills.oecd.org/OECD\\_Skills\\_Outlook\\_2013.pdf](http://skills.oecd.org/OECD_Skills_Outlook_2013.pdf)

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills

**Contribution:**

The Organisation for Economic Co-operation and Development (OECD)’s Skills Outlook provides an overview and assessment of the skillsets of OECD member states. The skill levels set by the OECD tend to influence policy of member states, and provide a benchmark for competency. The report identifies literacy, numeracy, and problem-solving as the foundation for effective and successful participation in the social and economic life of advanced economies. The assessment of literacy and data/technology usage is especially telling.

Finland and Japan have the highest performance in literacy, enabling them to “perform multiple-step operations to integrate, interpret, or synthesize information for complex inferences and appropriately apply background knowledge as well as interpret or evaluate subtle truth claims or arguments”. Citizens of these states are further able to analyze and engage in complex tasks involving data, statistics, and chance, spatial relationships. They can also perform tasks involving multiple steps and select appropriate problem-solving strategies and processes, as well as understand arguments and communicate well-reasoned

explanations for answers or choices.

Out of five levels, 14% of adults are level 1 with no data usage; 33% level 2 with interpretation of relatively simple data and statistics in text, tables, and graphs; 34.4% at level 3 with basic analysis; 11.4% level 4 more complex analysis involving statistics and chance, spatial relationships, change, proportions, and formulae; and 1.1% with level 5.

Problem solving in technology-rich environments is defined as the ability to use digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks. The assessment focuses on the abilities to solve problems for personal, work, and civic purposes by setting up appropriate goals and plans, and accessing and making use of information through computers and computer networks.

- Technology- hardware devices, software applications, commands and functions, and representations (ie. text, graphics, video, etc.).
- Tasks- intrinsic complexity, and explicitness of the problem statement
- Cognitive strategies: Set goals and monitor progress, plan, acquire and evaluate information, and use information
- Context: work-related, personal, and society and community

The description of proficiency levels is as follows:

- **Below level 1:** tasks are based on well-defined problems involving the use of only one function within a generic interface to meet one explicit criterion without any categorical or inferential reasoning, or transforming or information. Few steps are required and no sub-goal has to be generated.
- **Level 1:** tasks typically require the use of widely available and familiar technology applications; little or no navigation to access information to solve problems and require few steps and minimal operators, and simple forms of reasoning.
- **Level 2:** tasks require the use of both generic and more specific technology applications; some navigation of pages and applications is required to solve the problem; use of tools can facilitate resolution. Tasks may involve multiple steps and operators, with higher monitoring demands, unexpected outcomes or impasses may occur; some integration and inferential reasoning may be needed.
- **Level 3:** typically require the use of both generic and specific technology applications, and some navigation across pages and applications is required for solving problem. The use of tools is required to make progress toward solution, and may include multiple steps and operators. Goal to be defined and criteria may or may not be explicit, with high monitoring demands, and likely unexpected outcomes/impasses. Tasks may require evaluating the relevance and reliability of information, as well as integration and inferential reasoning may be needed to a large extent (p. 88)

The Programme for International Student Assessment (PISA) is provided by the OECD and tests literacy, numeracy, and problem-solving proficiencies in 15 year olds, and could be a useful as a baseline for building other assessment tools. The OECD recommends formal and informal education, as well as practice to ensure that literacy and technology skills don't depreciate.

**Citation:** Open Data Institute. (2012). *Open Data Institute Business Plan 2012-2017*. Retrieved from <https://theodi.org/about-us>

**Theme(s):** 21st Century Skills and Literacies, Delivery and Assessment

**Contribution:**

This White Paper consists of the business plan for the United Kingdom (UK)'s Open Data Institute (ODI), an independent non-profit limited guarantee company set up by the British Government. The purpose of the ODI is to serve as a centre to innovate, exploit, and research opportunities created by the UK's Open Data

Policy. It does so by providing a policy, research, and training hub for academia, business, public sector, and international actors regarding open data. The ODI also has a number of international open data hubs that assist in projects and research, including Open Data Toronto.

The ODI provides a number of courses and training streams. The two main open data training streams are:

- 3-month intensive short course in Open Data Technology leading to a post-graduate diploma in Open Data Technology
  - Designed to equip students with tools, techniques, and business methods of Open Data publication and application construction
- Open Data Fellows program for professionals
  - Same core material of the as the 3-month Open Data Technology stream
  - Fellows will be involved in the acquisition of experience and knowledge around open data policy, standards, and mentoring skills suitable for developing open data capabilities within organizations
  - Fellows will be able to create sustainable knowledge and understanding of open government data

Follow-up course content can be found on the ODI website, <https://theodi.org/courses>

**Citation:** Swan, K., Vahey, P, Kratcoski, A., van 't Hooft, M., Rafanan, K., and Stanford, T. (2009). Challenges to cross-disciplinary curricula: Data literacy and divergent disciplinary perspectives. Presented at the Annual Conference of the American Educational Research Association, San Diego, CA. Retrieved from <http://www.sri.com/work/publications/challenges-cross-disciplinary-curricula-data-literacy-and-divergent-disciplinary-p>

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills, Data Literacy As The Ability To Understand And Use Data Effectively To Inform Decisions, Delivery and Assessment, Barriers to Effective Data Literacy Instruction

**Contribution:**

Swan, et al., define data literacy as the ability to formulate and answer data-based questions; use appropriate data, tools, and representations; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions” (1). They argue that data literacy is a civic skill, and should be taught through relating the learning to real-world events/data to bridge the gap between learning facts and acquiring inquiry skills, critical reasoning, argumentations, and communication.

Cross-curricular education and real world data provide varying perspectives and context, promoting genuine inquiry, reflective discourse, and fosters students’ understanding that data can be queried to help make informed decisions about relevant problems. Although this study focuses on middle school students, there is opportunity to integrate this type of teaching method into higher education.

Thinking With Data (TWD) consists of four, two week modules, respectively in social studies, mathematics, science, and English language arts. They “address issues of data representation, common measure, and proportional reasoning, using real data accessed from real world media sources in discipline-specific, problem-solving contexts and align with relevant subject area standards” (2). This requires students to formulate and answer data-based questions; use appropriate data, tools, and representations; and develop and evaluate data based inferences based on world water issues. This relates to the approach *Preparation for Future Learning* that highlights structure, internalizing key dimensions and applying it in a variety of contexts.

The modules consist of social studies - analyzing data for water availability in Turkey, Syria and Iraq, and how best to share it; mathematics - proportional reasoning and data analysis; science - defending/disputing

various hypotheses concerning quality and availability, and use data to compare to US watersheds; English language arts - using research to develop persuasive arguments identifying major water issues and proposing solutions. The lecture-and-apply process is flipped, allowing students to recognize their inadequate knowledge, which makes learning more practical.

The assessment measures if students could (a) interpret complex tables of data; (b) understand arguments that used the tables of data; and (c) create their own proportional measures, which provide a more complete understanding of the data and arguments that one would have without the proportional measures. Students were asked to complete a pre and post course test measuring skill levels, which indicated that the students were able to improve their understanding through engagement in more sophisticated activities

Overall, the cross-disciplinary approach is most enriched teaching method, because it helps students in real world decision making, where differences are obscured, and things are not clearly laid out, critical thinking and the ability to recognize which approach is appropriate are important skills to hone in middle school to build on in higher education.

**Citation:** Zalles, D.R. (2005). Designs for assessing foundational data literacy. *Center for Technology in Learning, SRI International*. Retrieved from <http://serc.carleton.edu/files/NAGTWorkshops/assess/ZallesEssay3.pdf>

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills, Delivery and Assessment

**Contribution:**

This paper by Dr. Daniel R. Zalles describes four different assessment methods developed by SRI International that measure data literacy at the elementary and middle-school level. The paper is written through the lens of assessing foundational data literacy skills (examples outlined include sample size, sample selection, database structure, data distribution, central tendency, natural variability, and measurement error) required for courses in geosciences. The author outlines how data literacy (in the United States) is recognized as a key element of national standards relating to science, math, and social studies curricula (i.e. transliteracy). Students must be able to analyze evidence and data, carry out data manipulation, and report on conclusions based on evidence presented.

The paper goes on to describe the assessments designed by SRI International, designed to engage students in investigating real data sets, score them for deeper understanding, and provide an overall assessment of data literacy competency. These assessments are outlined below.

**EPA Phoenix** is an assessment carried out at the eighth grade level. It is a modular program that aims to assess the outcomes of information communication technology (ICT) learning in schools. A problem is posed to students, and they must use stylized data to provide the best solution. Students must assess graphs showing air quality ratings in Phoenix that have been generated from the Environmental Protection Agency's Air Quality System (AQS) database. They must then compare air quality trends to ozone ratings in other states, and come up with an evidence-based recommendation.

Assessed data literacy outcomes included:

- Comparing trend lines on graphs
- Transferring relevant data about air quality from one type of representation (line graphs) to another (data table) in order to facilitate analysis
- Critiquing the relevance of specific data for answering a research question; and
- Synthesizing data from different representations to formulate an overall conclusion

Solar Power (no longer active) was a modular program designed to assess the outcomes of ICT learning in schools. Students were required to use GIS representations to compare and contrast air temperature data, as well as compare and contrast model-generated data about incoming solar radiation. Students were also

required to also weather data, including the percentage of cloudy days within specific areas and seasons. The culminating activity of the Solar Power project was for the student to create an evidence-based recommendation on whether or not a given state should invest in solar power for energy generation.

GLOBE Integrated Investigation Assessments (no longer active) consisted of a number of web-based assessment tools and frameworks that provided teachers and students updates on GLOBE initiatives related to data literacy. Participating students were required to take atmosphere, hydrology, soil samples, etc. and post their data on the internet. Students would then be required to create maps and graphs with the data. Assessed data literacy outcomes included::

- Finding observable trends in the data
- Examining data for possible measurement or entry errors
- Identifying relationships between two variables
- Representing data in a graph or tables
- Using data to generate new representations in order to analyze trends
- Summarizing graphed data (i.e. mode, mean, median, range)
- Comparing data sets
- Generating evidence-based conclusions based on the data

The Thinking With Data (TWD) (see Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., & Stanford, T., 2012) consists of a number of foundational tools for data literacy designed to assess outcomes of integrated math/life science units at the fourth and sixth grade levels. Fourth grade students are required to collect, organize, and analyze data about pulse rate from samples of people drawn from different populations. Sixth grade students are required to conduct experiments in which they grow sets of "fast plants" under different conditions, test hypotheses, and analyze results. At the time of writing, both grade levels utilized the computer program Tabletop to view data and carry out analysis. Assessed data literacy outcomes included:

- Understanding what types of research questions can be answered by collecting data
- Determining the appropriateness of different data representations for different analyses; and
- Analyzing data distributions (i.e. strength of relationships between variables, detecting measurement error, detecting central tendency, and critiquing the viability of conflict claims about data)

The author makes the point that teaching the fundamentals of data literacy can straddle multiple content levels (at the K-12 level). Therefore, cooperation and collaboration among teachers is crucial in order to create a systematic and comprehensive approach to teaching data literacy.

## Websites

**Citation:** P21: Partnership for 21st Century Learning. (2012). Teaching critical thinking skills through project based learning [blog]. Retrieved from <http://www.p21.org/news-events/p21blog/1097-teaching-critical-thinking-skills-through-project-based-learning>

**Theme(s):** [21st Century Skills and Literacies](#)

### Contribution:

This blog post by Dr. John Mergendoller looks at the importance of critical thinking as a foundational 21st century skill, and how this skill can be built within post-secondary students. The author argues that Project Based Learning (PBL) can serve as a useful pedagogy to teach students to think critically. As critical thinking is a key underlying component of becoming data literate, there are some relevant take-away points. Moreover, practical exercises in data literacy workshops or courses could potentially benefit from using the PBL approach.

PBL requires that to be planned around topics that lend themselves to thoughtful consideration. Projects



must be structured to demand deliberative, reflective thought, with examples provided of what is correct. Students should be provided with the tasks, supports and scaffolds needed to develop critical thinking tools and strategies. Project topics or questions should be “non-Googleable”, in order to provoke deeper thinking.

Competencies relating to PBL include:

- Define terms of a project
- Consider whether information and concepts vary according to context
- Ability to weigh multiple explanations, evaluate evidence, and compare alternative actions based on their probability of success
- Ability to create a project plan

Actionable feedback from teachers and peers is essential for the development of skills and evaluation of arguments and reasoning can help students further develop in their own processes.

## Policies

**Citation:** Chinien, C., and Boutin, F. (2011). *Defining Essential Digital Skills in the Canadian Workforce*. Human Resources and Skills Development Canada. Retrieved from [http://www.nald.ca/library/research/digi\\_es\\_can\\_workplace/digi\\_es\\_can\\_workplace.pdf](http://www.nald.ca/library/research/digi_es_can_workplace/digi_es_can_workplace.pdf)

**Theme(s):** 21st Century Skills and Literacies, Data Literacy Competencies and Skills, Delivery and Assessment

### Contribution:

Purpose of the HRSDC policy was to ensure that the federal government’s Essential Skills Framework was accurate and current, in response to the G20 Summit’s recognition that a skilled workforce is essential to ensuring a strong, sustainable and balanced growth (11) (Government of Canada 2010). This framework was developed for Canadian workers through amalgamating various skills concepts/skill clusters:

1. foundational skills;
2. transversal skills;
3. \*technical digital skills
4. \*digital information processing skills, including cognitive and metacognitive skills. (6)
  - All include subskills and represent various economic sectors
  - \* is identified as most important skills

The method used in this study included a national and international literature review, as well as interviews and consultation with key informants. These professionals identified important skills:

1. use of digital systems and tools
2. use of software applications
3. applications of security measures in digital environments
4. processing of digital information
  - Note: there are varying levels of skill which should be indicated in any framework
  - 80% of key informants believed this framework to be useful and comprehensive, and that the clusters were relevant, accurate, and clearly defined
  - Information processing skills: ability to scan information visually, analyze information, interpret information, make effective and efficient use of information, use information to solve problems, and behave and act ethically in handling information, were ranked very important

“The most common ones [terms]are: IT literacy, ICT literacy, digital literacy, digital competence, ICT fluency, computer literacy, ICT skills, e-Skills, technological literacy, media literacy, information literacy, e-literacy,

generic skills, 21st century skills, multiliteracies, and new literacies. The overlaps between the various literacy concepts can be explained by: —the evolution of literacies from a skills focus through an applications focus towards a concern with critique, reflection, and judgement, and the identification of generic cognitive abilities or processes, or meta-skills. (Martin & Grudziecki, DigEuLit: Concepts and Tools for Digital Literacy Development, 2006, p. 253).” (15) Chinien & Boutin present several ways to assess competencies: self-assessment, demonstration of digital tools, pre-screening and testing, on-line assessment tools, performance assessment as an indicator, and test of workplace essential skills (TOWES). They recommend consulting a variety of resources depending on purpose:

- Educational Testing Service (ETS) developed an assessment for ICT skills, this includes three sections: cognitive proficiency, technical proficiency, and ICT proficiency - uses scenario-based tasks to measure
- UK Skills for Life Survey is a large scale survey and is a test for computer skills, not broader digital skills, and has 2 parts: test of awareness and test of practical skills
- Qualitative analysis of skills can be useful, but is much more relevant joined with quantitative analysis of data - this can be done through scenario-based testing, which measures cognitive and technical skills
- The Internet and Computing Core Certification 2005 is an internationally recognized standard, and covers three areas: Computer hardware, computer software, and using an operating system.
- Internet Digital Skills Assessment consists of 4 clusters: operational, formal, information, and strategic; these are measured with 3 indicators: successful task completion, main outcome achieved, and time spent on task completion
- European (International) Computer Driving License (ECDL) involves standardization, assessment, and certification for various levels. It is a non-proprietary program in partnership with not-for-profit, and so more widely available and affordable for the everyday person or organization. There are 12 modules:
  - 1. Concepts of information and communication technology; 2. Using the computer and managing files; 3. Word processing; 4. Spreadsheets; 5. Using databases; 6. Presentation; 7. Web browsing and Communication; 8. 2D Computer-aided design; 9. Image editing; 10. Web editing; 11. Health information systems usage; and 12. IT security
- Figures 24 and 25 offer assessment tools for indemand/important digital skills according to key informants Levels 1-5; Figure 27 offers skills-based assessment broken down into explained levels 1-4

Several countries were evaluated in consideration with what a successful digital skills framework should incorporate; although most of this article refers to “digital skills” and “ICT” there is much overlap with data literacy competencies and skills.

- Australia and New Zealand Skills Framework developed from ACRL:
  - Use: ICT infrastructure, devices to find content and services
  - Understanding and Interpretation: ability to understand and evaluate media content of various forms in order to judge the quality and trustworthiness of online information
  - Creation and Participation: ability to participate in social media and to generate digital content
  - Customer Protection/Security: understand cyber threats and be able to protect oneself against cyber crimes
- UK ICT and digital skills framework:
  - Improving productivity using IT: ability to plan, evaluate and improve the use of ICT to improve efficiency and productivity
  - Using IT systems: ability to use ICT systems safely, securely, sensibly, and purposefully to meet needs
  - Using IT to find and exchange information: ability to access, search, retrieve and exchange information using ICT, using digital networks and communication systems
  - Use IT software applications: ability to select and use software applications to process data and to produce and present information
- US - Technical, cognitive and ICT proficiencies are all essential to developing a comprehensive framework and an engaged society

- define: use digital technology to identify information needs; access: use digital technology to collect and/or retrieve information; manage: organize and classify digital information; integrate: interpret, summarize, compare and contrast digital information; evaluate: judge the quality, relevance, usefulness or efficiency of digital information; create: adapt and apply existing information to generate new knowledge; and communicate: use digital technology to exchange information with others
- EU Digital Competence
  - statement: to state clearly the problem to be solved or task to be achieved and the actions likely to be required
  - identification: to identify the digital resources required to solve a problem or achieve successful completion of a task
  - locate: to locate and obtain the required digital resources
  - evaluate: to understand the meaning conveyed by a digital resources
  - organization: to organize and set out digital resources in a way that will enable the solution of the problem or successful achievement of the task
  - integration: to bring digital resources together in combinations relevant to the problem or task
  - analysis: to examine digital resources using concepts and models which will enable solution of the problem or successful achievement of the task
  - synthesis: to recombine digital resources in new ways which will enable solution of the problem
  - creations: to create new knowledge objects, units of information, media products or other digital outputs which will contribute to task achievement
  - communication: to interact with relevant others whilst dealing with the problem or task
  - dissemination: to present the solutions or output to relevant others
  - reflection: to consider the success of the problem-solving or task-achievement process, and to reflect upon one's own development as a digitally literate person
- Netherlands Digital Skills Framework:
  - operational skills: skills to operate digital media
  - formal skills: skills to handle the structures or digital media
  - informational skills: skills to locate information in digital media
  - strategic skills: skills to employ the information contained in digital media towards personal (and professional) development
- The UNESCO framework can be used to facilitate digital skills development, assessment, and certification:
  - Information Literacy: definition and articulation of information; location and access of information; assessment of information; organization of information; use of information, communication and ethical use of information; and other information skills
  - ICT Skills-Media Literacy: Digital technology use; use of communication tools; use of networks; sift media messages; analyze media messages; and other ICT media skills
  - Literacy: Reading, writing, numeracy, and other basic skills
  - Reasoning: thinking skills

Proposed Canadian digital skills framework:

- determine information needs: recognize, define and articulate digital information needs
- access information: locate, select and retrieve digital information
- create information: generate new digital contents and knowledge by organizing, integrating, adapting and applying digital information
- assess information: judge the quality, relevance, usefulness, validity and applicability of digital information
- integrate information: interpret, analyze, summarize, compare and contrast, combine, repurpose and represent digital information
- apply information: use information of various digital formats effectively and efficiently to perform job tasks
- organize information: decode, restructure, and protect digital information

- input information: identify, recognize, record and store digital information to facilitate retrieval and use
- communication information: share digital information with others at work

Overall, Chinien & Boutin argue that “[d]igital skills are essential survival skills for the 21st century” (8) people must be able to process more complex cognitive problems quickly, effectively, and efficiently, because of the emergence of the global knowledge-based economy; in this type of environment human resources and technical infrastructure are the assets, because even today it is reported that almost all (92%) new employees being hired must have at least some basic level of digital technology skills (e-skills), and will be a key driver for job creation, and are essential to fully participate in the economy, which is only increasing in momentum.

**Citation:** Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 10 to 12: Computer Studies*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/computer.html>

**Themes:** 21st Century Skills and Literacies, Teach The Teachers

**Contribution:**

This Ontario Ministry of Education document outlines the Computer Studies curriculum currently being followed in Ontario high schools, and the associated outcomes for student learning. Although data literacy is not mentioned explicitly, there are related themes and competencies that warrant inclusion.

At its core, the Computer Studies curriculum involves defining and analyzing problems, and designing and testing solutions. Critical thinking, and analysis skills are integral. The document further explains that computer studies is relevant for all studies, as it incorporates a broad range of transferable skills, e.g. critical problem solving, logical thinking, knowledge strategies required for research, creative design, synthesis of ideas and data, evaluation, and communication. These skills all tie in with the notion of developing life-long learning habits that will help students adapt to new technology in the 21st century workplace and wider world. In particular, the ability to locate, question, and evaluate information allows a student to become an independent, life-long learner. It is also meant to reinforce mathematical and information literacy.

In terms of delivery, teachers are encouraged to use scaffolding or a modular style, wherein skills and competencies that reinforce each other are built by students successively. Teachers are also expected to model the skills that they expect students to learn. The document also states that teachers should take into account the diversity of students’ abilities, educational backgrounds, interests, level of technological capability, and learning styles, and tailor material appropriately. Categories of learning for Computer Studies are defined as follows:

- Knowledge and understanding
- Thinking:
  - Planning skills, i.e. focusing research, gathering information, selecting strategies, organizing projects
  - Processing skills, i.e. analyzing, interpreting, assessing, reasoning, generating ideas, evaluating, synthesizing, seeing different perspectives
- Critical/Creative Thinking, i.e. problem solving, decision-making, research
- Ethical use of computers
  - Students should be made aware of issues relating to privacy, safety, and responsible use of computers and the internet (also applies to data)

Students are also expected to develop an awareness of environmental stewardship and sustainability relating to computer studies, as well as critical thinking, citizenship, and personal responsibility when using computers and the Internet. Relevant course learning outcomes related to data literacy are as follows:

- Course: Computers and Society
  - Explain on privacy of techniques for collecting and processing data

- Describe legal and ethical issues relating to computers (also applies to data literacy)
- Course: Programming Concepts and Skills
  - Data types and expressions, including using one-dimensional arrays of compound data types

**Citation:** Ontario Minister of Education. (2008). *The Ontario Curriculum Grades 9 and 10: Technological Education*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/teched.html>

**Themes:** [21st Century Skills and Literacies](#)

**Contribution:**

This Ontario Ministry of Education document outlines the Technological Education curriculum currently being followed in Ontario high schools for grades 9 and 10, and the associated outcomes for student learning. Although data literacy is not mentioned explicitly, there are some related themes and competencies that warrant inclusion.

The document outlines how technological innovation influences all areas of life, both in daily interactions as well as the wider global social, business, and government contexts. Stemming from this, students must be able to meet the challenges and opportunities of the 21st century. To this end, students must be technically literate, i.e. able to understand, work with, and benefit from a range of technologies. The document makes mention of the Ontario Skills Passport (OSP); a bilingual web-based resource designed to enhance school-workplace connections, and which provides clear descriptions of Essential Skills (e.g. reading, writing, text, computer use, etc.).

In terms of delivery, technological education courses should lend themselves to a wide range of approaches, requiring students to discuss issues, identify problems, plan solutions, work collaboratively, conduct research, and think critically.

Course content for Technological Education ranges broadly from communications technology to hairstyling and aesthetics. The most relevant course learning outcomes fall within the field of computer technology, but still fall more towards the programming side:

- Computer Technology Fundamentals
  - Data Representation and Digital Logic
    - Describe how computers and represent and process data using the binary number system (e.g, binary counting, binary codes, ASCII code)
    - Derive the truth tables of the fundamental logic gates (e.g. AND, OR, NOT, NOR, NAND, XOR)
    - Write Boolean equations for the fundamental logic gates

**Citation:** Ontario Minister of Education. (2008). *The Ontario Curriculum Grades 11 and 12: Technological Education*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/teched.html>

**Themes:** [21st Century Skills and Literacies](#)

**Contribution:**

This Ontario Ministry of Education document outlines the Technological Education curriculum currently being followed in Ontario high schools for grades 11 and 12, and the associated outcomes for student learning. Although data literacy is not mentioned explicitly, there are some related themes and competencies that warrant inclusion. This document is a continuation of the fundamental learning outcomes covered in the prior grades (9 and 10).

Technological courses at this level should involve an open, collaborative, activity/applied approach that

takes into account students' interests (e.g. allowing them to pick from a range of practical projects in which to apply theory and taught content). Literacy, mathematical literacy, and inquiry/research skills are critical to students' success, and key aspects of technological courses. Students also learn to analyze the context and background of challenges in order to explore a variety of possible solutions to said challenges. The document states that the ability to locate, question, and evaluate information is crucial for a student to become an independent, life-learn learner.

Course content and delivery is mostly in-line with the prior grades, but teachers should work collaboratively in order to plan and deliver different courses/modules when possible, and avoid unnecessary overlap.

**Citation:** Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 9 and 10: Mathematics*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/math.html>

**Themes:** [21st Century Skills and Literacies](#)

**Contribution:**

This Ontario Ministry of Education document outlines the Mathematics curriculum currently being followed in Ontario high schools for grades 9 and 10, and the associated outcomes for student learning. Data literacy is not specifically mentioned, but hard skills/competencies relating to it are covered.

Processes that are key for students to be successful in mathematics (and related to data literacy) include problem solving, reasoning and proving, reflecting, selecting tools and computational strategies, connecting, representing, and communication. Representing mainly involves representing mathematical ideas and models in different ways, e.g. numeric, geometric, graphical, etc., as well as using dynamic software to create representations (i.e. Excel). Students should be able to go from one representation to another, and identify and understand differences, similarities, and other relationships. Communication includes being able to create visualizations of data.

Relevant course learning outcomes

- Grade 9 Math Linear Relations unit
  - Using data management to investigate relations
    - Interpret the meaning of points or plots on a graph representing a linear equation
    - Design and carry out an investigation or experiment involving relationships between two variables, including the collection and organization of data, using appropriate methods, equipment, and/or technology (e.g. surveys, scientific probes, online sources, etc.), and techniques (i.e. generating tables and graphs to visualize data)
    - Describe trends and relationships observed in data

**Citation:** Ontario Ministry of Education. (2008). *The Ontario Curriculum Grades 11 and 12: Mathematics*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/math.html>

**Themes:** [21st Century Skills and Literacies](#)

**Contribution:**

This Ontario Ministry of Education document outlines the Mathematics curriculum currently being followed in Ontario high schools for grades 11 and 12, and the associated outcomes for student learning. Although data literacy is not mentioned explicitly, there are some related themes and competencies that warrant inclusion, especially within the curriculum's data management component. This document is a continuation of the fundamental learning outcomes covered in the prior grades (9 and 10).

The curriculum makes direct mention of preparing students for working, living, and contributing to the 21st century society. It goes on to say that in order to meet the demands of the world in which they live students will need to adapt to changing conditions and learn independently. This includes using technology effectively to process large amounts of quantitative data.

The most relevant areas to data literacy relate to the Data Management stream of math courses offered in Grade 11 and 12. Students apply methods for organizing and analyzing large amounts of information, solve problems involving probability and statistics, and carry out a culminating investigation that integrates statistical concepts and skills. Students have a choice in which math stream they enrol in, and thus not all students will be learning the foundational data management competencies offered at the secondary level.

Relevant course learning outcomes are as follows:

- Mathematics of Data Management (Grade 12)
  - Organization of Data For Analysis unit
    - Recognize and describe the role of data in statistical studies
    - Identify and explain reasons why variability is inherent in data
    - Distinguish different types of statistical data
    - Determine and describe principles of primary data collection
    - Explain the distinction between the terms population and sample, and what characterizes a good sample
    - Collect data from primary and secondary sources
  - Statistical Analysis unit
    - Analyze, interpret, and draw conclusions from one variable and two variable data
    - Evaluate validity of data
  - Culminating Data Management Investigation
    - Pose a significant problem that requires the organization and analysis of a set of primary or secondary quantitative data
    - Design a plan to study the problem
    - Gather and organize data related to the study of the problem
    - Interpret, analyze, and summarize data
    - Draw conclusions from analysis of data
    - Evaluate strength of evidence
    - Specify limitations, next steps
    - Compile a comprehensive report
      - Present a summary to peers using technology (e.g. PowerPoint)
      - Answer questions and respond to critiques

**Citation:** Ontario Ministry of Education, (2008). *The Ontario Curriculum Grades 11 and 12: Interdisciplinary Studies*. Retrieved from <https://www.edu.gov.on.ca/eng/curriculum/secondary/interdisciplinary.html>

**Themes:** [21st Century Skills and Literacies](#)

**Contribution:**

This Ontario Ministry of Education document outlines the Interdisciplinary Studies curriculum currently being followed in Ontario high schools for grades 11 and 12, and the associated outcomes for student learning. Data literacy is not specifically mentioned, but there are certain related themes and skills that warrant mentioning.

The content of Interdisciplinary Studies at the secondary level recognizes the changing 21st century context, and the need to respond to challenges with insight and innovation. These challenges often arise in areas that combine or cut across different subject disciplines, e.g. space science, information management, alternative energy technologies, etc. To deal with these issues, students require competencies and skills from a range of discrete disciplines. To this end, the Interdisciplinary Studies curriculum aims to build the

following skills in students:

- Research processes
- Information management
- Collaboration
- Critical and creative thinking; and
- Technological applications

The document also makes note of the importance of teaching students to be information literate. Students must be able to combine diverse models of research and inquiry, integrate a range of information management skills and technologies, and apply the processes of information organization, storage, and retrieval to new situations and across disciplines. These principles could be readily applied to data literacy as well. Actual courses can be built at the discretion of teachers, as long as these learning outcomes are addressed in the course content and evaluation techniques.

## Courses and Workshops

McCaffrey, M. (2015). INF 2115H Data Librarianship [Syllabus]. iSchool, University of Toronto. Retrieved from <http://mccaffrey.ischool.utoronto.ca/2115/syllabus.pdf>

### Key Points:

- Assignments:
  - Statistics- definitions, literacy, concepts, interpretation, identification, location, and retrieval
  - Data- identification and location. Extraction, recoding, subsetting, manipulation, etc. (from pre-determined datasets)
  - Statistics and Data- advanced (reference questions based)
  - Group Assignment
- Class Schedule
  - 1- Syllabus, expectations, outcomes, technical requirements, introductions to statistics
  - 2- Introduction to statistics part II, statistical literacy, the administrative environment
  - 3- Canadian statistical programmes and policies
  - 4- Finding Canadian statistics
  - 5- American statistics, international statistics
  - 6- Microdata, part I
  - 7- Microdata, part II
  - 8- Canadian Microdata
  - 9- Foreign and international microdata, other repositories, business data
  - 10- Data visualization, geographical information systems
  - 11- Reference work
  - 12- Data services administration, issues and trends
- Lists readings for each session

MIT. (2014). [Pamphlet] *Tackling the Challenges of Big Data*. Retrieved from <https://mitprofessionalx.mit.edu/courses/course-v1:MITProfessionalX+6.BDx+5T2015/about>

### Key Points:

This is an online course held over 6 weeks with 5 modules, 18 topics, 20 hours of video, case studies, community wikis, and discussion forums (provoking though in medicine, social media, finance, and transportation. It covers areas such as: data collection (smartphones, sensors, the Web); data storage and processing (scalable relational databases, Hadoop, Spark, etc); extracting structured data from unstructured data, systems issues (exploiting multicore, security); analytics (machine learning, data compression,



efficient algorithms); visualizations, and a range of applications.

The course aims to reduce the time from research to industry dissemination and expose participants to some of the most recent ideas and techniques in Big Data. Students will learn the state-of-the-art in Big Data, and will be able to:

- Distinguish what is Big Data (volume, velocity, variety), and learn where it comes from, and what are the key challenges
- Determine how and where Big Data challenges arise in a number of domains, including social media, transportation, finance, and medicine
- Investigate multicore challenges and how to engineer around them
- Explore the relational model, SQL, and capabilities of new relational systems in terms of scalability and performance
- Understand the capabilities of NoSQL systems, their capabilities and pitfalls, and how the NewSQL movement addresses these issues
- Learn how to maximize the MapReduce programming model: What are its benefits, how it compares to relational systems, and new developments that improve its performance and robustness
- Learn why building secure Big Data systems is so hard and survey recent techniques that help; including learning direct processing on encrypted data, information flow control, auditing, and replay
- Discover user interfaces for Big Data and what makes building them difficult

The course consists of 5 modules:

- Module One: Introduction and Use Cases The introductory module aims to give a broad survey of Big Data challenges and opportunities and highlights applications as case studies. > Introduction: Big Data Challenges > Case Study: Transportation > Case Study: Visualizing Twitter
- Module Two: Big Data Collection The data capture module surveys approaches to data collection, cleaning, and integration. > Data Cleaning and Integration> Hosted Data Platforms and the Cloud
- Module Three: Big Data Storage The module on Big Data storage describes modern approaches to databases and computing platforms. > Modern Databases > Distributed Computing Platforms> NoSQL, NewSQL
- Module Four: Big Data Systems The systems module discusses solutions to creating and deploying working Big Data systems and applications. > Multicore Scalability > Security> User Interfaces for Data
- Module Five: Big Data Analytics The analytics module covers state-of-the-art algorithms for very large data sets and streaming computation> Machine Learning Tools> Fast Algorithms I > Fast Algorithms II > Data Compression> Case Study: Information Summarization> Applications: Medicine> Applications: Finance

## Associations and Organizations

**Citation:** BizEd Editorial Staff. (2014). Big data gets bigger on campus. *AACSB International*, (March/April). Retrieved from <http://www.bizedmagazine.com/archives/2014/2/research/big--data-gets-bigger-on-campus/>

### Contribution:

Authors argue that analytics-related positions are predicted to increase 25% by 2018, and universities are developing focused training:

- Babson College in Wellesley, Massachusetts-undergraduate and graduate concentrations available

in data analytics with four paths functional depth, marketing analytics, financial analytics, and industry sectors. Students must complete courses in foundation, application, and depth courses

- MIT Cambridge provides a four week online course for technical professionals and executives “Tackling the Challenges of Big Data”
- John Molson Executive Centre at Concordia University in Montreal offers a three week \$4000 online and face-to-face certificate program in data analytics for middle managers
- Goodman School of Business at Brock University in Ontario offers a specialization in business analytics to its two year MBA program
- IBM Analytics Talent Assessment measures preparedness for real world. It takes 30-40 minutes and includes sections that measure cognitive ability, verbal and logical reasoning and provides students with report
- Pilot testing big data analytics programs: Fordham University, George Washington University, Illinois Institute of Technology, Northwestern University, The Ohio State University, Southern Methodist University, University of Massachusetts Boston, and the University of Virginia.
- Scott Moore, Babson’s undergraduate dean: “this kind of knowledge is becoming increasingly important..soon, every business organization will be an analytics-focused organization” (para. 9).

Digital Edition 2015 <http://www.bizedmagazine.com/current-issue/~media/5B840464B0AB425D918342919A9D8699.ashx>

The edition provides information on American post-secondary institutions that offer courses and programs that are data centred, as well as information on a new program for real-time student evaluations, from a Canada based learning solutions provider eXplorance. They have introduced bluepulse, a real-time online social collaborative hub and course evaluation tool that allows professors to solicit student feedback on and suggestions for the course as it progresses. (60).

The website is business centered, and provides the user with seminar sessions based on skill building activities, and School Profile Search through DataDirect (most comprehensive business education database in the world <http://www.aacsb.edu/knowledge/data/datadirect/>).

**Citation** Data Science Central. (2015). Homepage. Retrieved from <http://www.datasciencecentral.com/>

**Contribution:**

This website is an online resource for big data practitioners. It offers the users a wide variety of tools and skills relating to data, including workshops beginning at Intro to Data Science to more advanced topics such as programming tools, data wrangling, data story, inferential statistics, and machine learning

[https://www.mysliderule.com/workshops/data-science-intensive?utm\\_source=DSC&utm\\_medium=newsletter&utm\\_campaign=DS2Workshop](https://www.mysliderule.com/workshops/data-science-intensive?utm_source=DSC&utm_medium=newsletter&utm_campaign=DS2Workshop)

**Citation:** Cavique, L. (2015, June 18). 3 issues with big data. [Blog]. *Data Science Central*. Retrieved from <http://www.datasciencecentral.com/profiles/blogs/big-data-3-data-issues>

**Contribution:**

Cavique explains that Big Data refers to amounts too large to be processed by traditional tools, won't fit into a single server, and is unstructured meaning cannot fit into a database. 3 Vs: Volume-too big; update Velocity-too continuously flowing; and Variety of formats-too unstructured. Structured consists of SQL

databases, data warehouses, XML databases, and recently NoSQL databases, and unstructured of plain text that is not tagged or specifically formatted.

Three subproblems are identified due to their complicated nature: video production increasing exponentially, but difficult to search for, because of the many variables associated with it; NoSQL allows us to deal with large volumes of data and provides flexibility, but there are also a large number of consultants; and event logs refers to the execution of a computer command program recorded, highly used surveillance, and can be key in Business Process Management technology.

This website also provides many relevant articles for people interested in learning:

- How to avoid a data disaster <http://www.datasciencecentral.com/profiles/blogs/how-to-avoid-a-data-disaster-infographic>
- 10 common NLP terms explained for the text mining novice <http://www.analyticbridge.com/profiles/blogs/10-common-nlp-terms-explained-for-the-text-mining-novice>
- Harnessing Big data for Security <http://www.datasciencecentral.com/profiles/blogs/harnessing-big-data-for-security-intelligence-in-the-era-of-cyber>

**Citation** Government of Canada. (2015). The Federal Geographic Data Committee Retrieved from <https://www.fgdc.gov/>

**Contribution:**

This website provides users with online lessons:

- Geospatial Data Discovery and Access: Geospatial One-Stop Portal; Geospatial Metadata; and Geospatial Web Services
- Geospatial Data Integration: NSDI (National Spatial Data Infrastructure) Standards; NSDI Data Themes; and The National Map
- Geospatial Partnership, Policy and Planning: NSDI Policies and Practices; NSDI Partnership Opportunities; and Geospatial Business Planning
  - Each includes PowerPoint slides with notes
  - Module based learning

**Citation:** NASA. (2015). NASA Earth Data. Retrieved from <https://earthdata.nasa.gov/earth-observation-data/tools>

**Contribution:**

Provides users with tools to use while manipulating earth data. They focus on search and order tools; data handling (reading/ingesting, format conversion, and data manipulation); subsetting and filtering tools (temporal, spatial, parameter, and channel), geolocation, reprojections, and mapping tools; and data visualization and analysis tools. Additionally, online webinars and tutorials are available through YouTube, all centred around earth data in some facet

- **SIPS** (Science Investigator-led Processing System) provides support for targeted learning <https://earthdata.nasa.gov/about/science-investigator-led-processing-systems>