

Javier Calzada Prado and Miguel Ángel Marzal

Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents

Abstract: The growing importance of data in society in general and scientific domains in particular, mirrored in the Open Data initiative and in the advent of eScience, requires public, school and academic libraries to contribute to both data and information literacy, as part of their mission to further knowledge and innovation in their respective fields of action. No specific library standards have been proposed to date, however, and most research studies conducted adopt a partial view of data literacy, stressing only the components needed in any given context. The present paper aims to contribute to the advancement of data literacy with the proposal of a set of core competencies and contents that can serve as a framework of reference for its inclusion in libraries' information literacy programs. The various definitions of data literacy are discussed, the coverage of the competencies listed in information literacy standards is described, and the competencies considered in the experiments conducted to date in education and libraries are identified. The conclusion drawn is that the model proposed can favour the development of data literacy support resources and services. Topics for further research are also specified.

Dr. Francisco Javier Calzada Prado: Assistant Professor, Library & Information Science Department, Universidad Carlos III de Madrid, Spain, Email: fcalzada@bib.uc3m.es

Dr. Miguel Ángel Marzal García-Quismondo: Professor, Library & Information Science Department, Universidad Carlos III de Madrid, Spain, Email: mmarzal@bib.uc3m.es

Introduction

The advent of the information society has brought a gradual acknowledgement of individuals' need to effectively access, handle and use information to solve problems, engage in life-long learning for the attainment of full social integration and optimal personal and professional development, and contribute actively to the societies in which they live. The suite of competencies involved, which has been dubbed "information literacy," has been associated

from the outset with democratic participation (Owens 1976) and has even come to be regarded as a basic civil right (Sturges and Gasting 2010). This state of affairs has accentuated libraries' educational task, for as key information mediators, they occupy a strategic position for the development of such competencies. This is particularly obvious in the growing involvement of school, public and academic libraries in the field. Two types of activities to further information competencies are generally conducted in these institutions: specialized reference services and information literacy training programs.

To date, these services have focused on accessing and handling sources of bibliographic information (books, articles, reports, legal texts and so on). In recent years, however, the availability and volume of statistical, scientific or technical source (or raw) data have grown exponentially. In addition to traditional sources, some of which are published and distributed under license or subscription (such as IHS Global Insight or LexisNexis Statistical) while others are available cost-free on the Web (such as UNdata, World Bank or EUROSTAT), the recent presence of open data sources is of enormous importance for researchers and the general public alike. The Open Data movement, heir in part to the Open Source and Open Access movements, encourages the free publication of data from different domains under licenses that favour their reuse.

Major initiatives in the scientific sphere are Open Knowledge and Science Commons, with prominent statements such as OECD's communiqué "Science, Technology and Innovation for the 21st Century" (OECD 2004). More recently, the Obama Administration (U.S. Office of Science and Technology Policy 2012) has introduced Big Data, and the European Commission has released its communication "Towards better access to scientific information: boosting the benefits of public investments in research" (European Commission 2012). All these endeavours are intended to favour open access to source data generated by publicly funded research.

This premise of open access to source data has acquired particular relevance in eScience. In this domain, the challenges and opportunities offered by the ever-growing data processing and analytical capacity of com-

puter technology combine with the new forms in which researchers relate to the scientific community and society with a view to facilitating methodological transparency, data preservation, sharing and reuse, and accountability. One example of the impact of this new scenario is the National Science Foundation's (NSF) 2011 decision to require a *data management plan* in all requests for funding, which must describe the resources to be devoted to managing, preserving and publishing the data generated as a result of the research conducted. Escience is usually characterized as 'data-intensive', which calls not only for the necessary technological infrastructure and resources to handle and mine data, but also the human resources able to make sense of them and oversee the process. Both matters pose substantial challenges at present for institutions and researchers alike (Soehner, Steeves, and Ward 2010).

In the public domain, Open Data has favoured freer access to the data generated by the public sector. This line of work was formalized in the expert conference held at Sebastopol in 2007 (Sunlight Foundation 2010), which paved the way for both the Obama Administration's Open Government Initiative (The White House 2009) and the European Union's PSI (Public Sector Information) directives of 2003 and 2009 (European Commission 2003, 2009). Governments the world over have, then, been publishing more and more public sector data on the Web in open format (Open Government Partnership 2012). As Berners-Lee (2010) noted, these data constitute an opportunity to enhance the effectiveness of public institutions, reinforce democratic values (insofar as they favour transparency and the public's control of government) and create business opportunities based on data reuse and mining. In practice, however, open data access appears to benefit government (as a political commitment) and the entrepreneurial class (as a business opportunity) more than the public at large, which finds data difficult to interpret (as statistics on numeracy show) and consequently depends on apps developed by others to make sense out of, use and derive value from them.

In another vein, data and their sources are increasingly present in everyday life (primarily through the media), influencing opinion and decision-making. A number of organizations, such as the International Statistical Institute (ISI) (2013), the American Statistical Association (ASA) (2011), the International Association for Social Science Information Services and Technology (IASSIST), and the United Nations (United Nations Economic Commission for Europe 2012), have stressed the need for more statistical competency instruction at all educational levels and for its inclusion as a cross-curricular subject not linked solely to mathematics, to enhance the public's statistical know-how.

The above, obviously inter-related, spheres (education systems contribute to training the public and professionals, researchers among them), pose one and the same challenge: the need to prepare individuals to make significant use of data sources. A need exists, then, 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 statistical competencies. As in the case of information literacy, libraries are well-positioned to play a strategic role in the development of such competencies. Indeed, several authors have called upon librarians, particularly university librarians, to further data literacy in information literacy training programs (Gray 2004; Schield 2004; Stephenson and Caravello 2007; Kellam 2011; Haendel, Vasilevsky, and Wirz 2012), going so far as to claim that "it is time for academic libraries to invest in data literacy programs" (Merrill 2011, 146). Nonetheless, we feel that data literacy, like information literacy, should be acquired gradually at all levels of schooling and even throughout individuals' lifetimes. We consequently believe that school and public libraries should also include data literacy, adapted as necessary, in their information literacy programs.

But, how can libraries include data literacy in their instructional programs? Exactly what competencies should be covered? And how are they related to information literacy? The scant literature published to date on data literacy and the competencies it entails contains not one, but many approaches, depending on the context and purpose sought. If one lesson is to be learned from the vast documentation on information literacy, however, it is that it calls for a comprehensive model or framework as a basis for ensuring that these competencies are consistent and meaningful in all the contexts and situations where they are to be applied, regardless of the sophistication or specificity involved.

The present paper aims to contribute to the advancement of data literacy in libraries by identifying a set of core competencies and contents that may constitute a framework of reference for the inclusion of this topic in information literacy instructional programs. It begins with a review of data literacy and its inter-relationship with other types of literacy. Data literacy coverage in the information literacy models proposed is subsequently analyzed and the most significant experiences conducted in the field to date in libraries are identified, along with the competencies addressed. This paves the way for the proposed reference framework that may serve as a guide for libraries when including data literacy in their educational programs.

Data Literacy

“Science literacy” is an educational priority world-wide. In 1999 the “Declaration on Science and the Use of Scientific Knowledge” identified its strategic role in sustainable development and democracy (UNESCO and International Council for Science [ICSU] 2002). The Rocard report (European Commission 2007) highlighted the importance of propagating scientific culture among young Europeans. President Obama, in turn, recently launched his Educate to Innovate campaign (The White House 2012), in an attempt to improve young Americans’ performance in science, technology, engineering and mathematics (STEM).

In Western higher education, particularly in North America, Australia and Europe, quantitative literacy has been encouraged since the late 1980s. It has recently been given a fresh impetus, for instance, through the British national strategy, headed by the Economic and Social Research Council (ESRC), to further training for the use of quantitative methods in social science. Quantitative literacy is defined as “using simple mathematical concepts to solve everyday problems” (Steele and Kiliç-Bahi 2008, 1), a pragmatic approach to mathematics teaching. For Steen, quantitative literacy is defined by its direct relationship to the real world. He states that “the facility we want students to acquire is not just about quantities... [R]easoning, argument and insight are as essential to QL as are numbers; so too are notions of space, chance, and data” (Steen 2004, 25).

One of the essential components of quantitative literacy is statistical literacy. Wallman defined the latter as “the ability to understand and critically evaluate statistical results that permeate our daily lives - coupled with the ability to appreciate the contribution that statistical thinking can make in public and private, professional and personal decisions” (Wallman 1993, 1). Watson, in turn, identified three abilities that determine statistical competence. In increasing order of complexity, they are: a) a basic understanding of probabilistic and statistical terminology and the ability to perform analytical and statistical calculations; b) the ability to interpret probabilistic and statistical language and concepts when they are embedded in social media contexts; c) the ability to critically evaluate statistical claims related to sampling, the distribution of raw data, appropriate use of statistics, graphs, causal claims made, and probabilistic statements (Watson 1997). In similar terms, Gal (2002, 2-4) claimed that statistical literacy comprises two inter-related competencies: “a) people’s ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse

contexts, and when relevant; b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, or their concerns regarding the acceptability of given conclusions.”

A new term, furthered primarily by the social science and open data communities, has arisen of late in connection with statistical literacy: data literacy. Some authors have equated it to statistical literacy (Stephenson and Caravello 2007), while others, such as Schield (2004), observe differentiating features. For the latter author, data literacy is the part of statistical literacy that involves training individuals to access, assess, manipulate, summarize and present data, whereas statistical literacy aims to teach how to “think critically about descriptive statistics.” The two concepts share one essential component, critical thinking, which draws from mathematical and statistical aptitudes, general knowledge and the values of the persons concerned. Moreover, this component positions both concepts near the realm of information literacy, as noted by Schield (2004), although data literacy is closer, inasmuch as it shares other elements, including its name (“data” and “information” both refer to the object of literacy) and its focus on management: data literates must be able to effectively access, handle and use data. However, as Hunt (2004, 14) points out: “in the practical implementation (or on the ground) data literacy is quite different than traditional information literacy.” The origin of this difference lies mainly in the greater complexity involved in handling data than any other type of information (Thompson and Edelstein 2005). Nonetheless, their similarities may lead to regarding data literacy as a complement to or a form of information literacy, as respectively suggested by Stephenson and Caravello (2007) and Otto (2012), which makes us think that data literacy would be the umbrella concept covering statistical literacy, rather than the vice versa.

While acknowledging the obvious interdependence between data and statistical literacy (no data manipulation, summary or presentation, to use Schield’s terms, is possible unless the operations characteristic of statistical analysis are performed, nor can statistics be critically appraised unless the fact that they entail data manipulation and interpretation is recognized), a broader perspective for dealing with data is secured if the former approach is taken as the construct. As a result, data literacy can be viewed both as a whole and as an integrated assemblage of other competencies, such as data collection, generation and management in research projects or organizations.

Using this framework, statistical literacy is envisaged as the component of data literacy involved in the criti-

cal appraisal, interpretation, processing, and statistical analysis of data. Data literacy can be defined, then, as the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data. From that perspective, information literacy and data literacy form part of a continuum, a gradual process of scientific-investigative education that begins in school, is perfected and becomes specialized in higher education and forms part of individuals' skill set throughout their lifetime. In the present paper the term data literacy is based on this approach.

Data Literacy in Information Literacy Standards

To date, no specific standards or guidelines have been established for data literacy, although some of the competencies entailed have recently been included as part of information literacy standards. The latter provide a reference framework for the competencies needed to use information efficiently in problem solving and generating new knowledge. These competencies generally include: the ability to define precisely the informational need; the ability to locate information sources suited to that need; the ability to assess critically both the sources and the ideas expressed therein; the ability to manage the information selected; the ability to analyze and synthesize information to support arguments or generate new ideas; the ability to document the sources used; and the ability to record or communicate the results in an ethical manner. In the context of this overall scheme, some standards, such as the Australian and New Zealand Information Literacy Framework (ANZIL) (Bundy 2004), fail to deal with data specifically, while others have attempted to assimilate or at least partially develop the competencies associated with data literacy.

In the domain of primary and secondary education, the American Association of School Librarians' (AASL) "Standards for the 21st-Century Learner" contain no explicit reference to data, although the AASL proposes a crosswalk to align them with the "Common Core State Standards" (AASL 2011) in English language arts, history/social studies, science and technical subjects, and mathematics.

In the university domain, both the "Seven Pillars of Information Literacy: Core Model for Higher Education" and its version for research, the "Seven Pillars of Information Literacy: a Research Lens for Higher Education" (Society of College, National and University Libraries [SCONUL]

2011a, 2011b) assume that the object of information literacy is to work with data. In fact, data and information are consistently mentioned jointly, although no specific criteria or recommendations are provided for the former. Neither do the society's working papers "The SCONUL Seven Pillars of Information Literacy through a Digital Literacy Lens" or "The SCONUL Seven Pillars of Information Literacy through an Open Content Lens" (SCONUL 2012, 2011c) accord data special treatment. And significantly, despite the focus on open content in the latter, it makes no reference to the Open Data Initiative.

The Association of College and Research Libraries (ACRL) (2000) acknowledges the value of research data and includes a number of explicit references to them in the information competencies proposed in its "Information Literacy Standards for Higher Education" (ILSHE). The ACRL has likewise published several adaptations to these standards for areas with specific needs, in which the competencies associated with data literacy acquire particular significance. These include "Information Literacy Standards for Science and Engineering/Technology" (ILSSET), "Information Literacy Standards for Anthropology and Sociology Students" (ILSAS), "Political Science Research Competency Guidelines" (PSRCG) and "Information Literacy Competency Standards for Journalism Students and Professionals" (ILCSJ) (ACRL 2006, 2008a, 2008b, 2011a).

Table 1 lists competencies explicitly associated with data literacy (they either mention data, statistics and/or quantitative methods) found in six of the standards mentioned above as particularly related to this construct: the crosswalk between the AASL and the "Common Core State Standards" (AASL/CCS) and the five ACRL standards (ILSHE, ILSSET, ILSAS, PSRCG and ILCSJ). These competencies have been grouped by type for easier identification of the importance attached to each in the standards. Note, firstly, that the competency consisting of the ability to determine and use research methods suited to the problem broached (both for collecting and using data) is present in all the standards analyzed, and stressed in particular in political science and journalism. The ability to handle (transfer, transform) and analyze data, which includes the ability to use data management tools and statistical software, is explicitly mentioned in all except the journalism standards. The competencies relating to knowledge of and access to sources and their critical assessment are likewise particularly important for most of the standards.

Table 1: Data literacy competences in information literacy standards

Data literacy competency	AASL/CCS	ILSHE	ILSSET	ILSAS	PSRCG	ILCSJ
Ability to identify the context in which data are produced and reused (data lifecycle)			1.2.f; 5.2.e	3.3.e	1.2.a; 3.2.d	
Ability to recognize source data value, types and formats	1.2.3; 4.3.2	1.2.c	1.2.e		1.2.d	
Ability to determine when data are needed		1.2.f	1.2.d		1.1.b	
Ability to access data sources appropriate to the information needed	1.1.4			2.1.a	2.1.e; 2.2.g; 2.4.c; 3.4.f	1.2.b; 1.2.d; 2.2.a; 2.3.a
Ability to critically assess data and their sources	1.1.4; 1.1.5; 1.1.7; 1.2.4; 2.2.2	3.4.e			3.2.a; 3.4.d	1.2.e
Ability to determine and use suitable research methods	1.1.1	2.3.d	2.3.d	1.2.a	1.2.f; 2.1.a; 2.3.d	1.2.c; 2.2.b
Ability to handle and analyze data	2.1.4	4.1.d	3.3.c; 4.4.b	3.1.c	2.5.a	
Knowing how to select and synthesize data and combine them with other information sources and prior knowledge	2.1.1	3.1.b		3.1.b	1.1.e; 3.3.c	
Ability to present quantitative information (specific data, tables, graphs, in reports and similar)	2.1.6; 3.1.4				4.1.c; 4.3.b	4.1.d
Using data ethically		5.2.e	4.1.d; 4.2.e			5.3.c
Ability to apply results to learning, decision-making or problem-solving	2.3.3; 2.1.3					
Ability to plan, organize and self-assess throughout the process				1.4.b	1.3.c	1.2.f

Data Literacy in Libraries' Instructional Programs and Services

While training to master some of the basic data literacy competencies has formed or is gradually becoming part of the curriculum in all levels of schooling, training delivered in libraries or with their support has focused primarily on higher education. Indeed, data have traditionally formed part of the resources in all manner of libraries, although with greater weight in academic libraries in light of their users' special research-related needs. This has led in the last few decades to the creation of data libraries or data services as specialized research library facilities. Kellam (2011) noted that in the nineteen sixties and seventies some large American universities created data support centres or archives in academic departments or computing centres, while in the 1990s libraries began to create data centres and services *per se*, such as the University of Minnesota libraries' Machine Readable Data Centre (Treadwell and Cogswell 1994). Bennett and Nicholson (2007), in an analysis of a random sample of Association of Research Libraries (ARL) members' websites, observed that this trend has continued over time, finding that libraries were competing with other centres or services within their institutions to furnish data-related resources and services.

At present, the provision of data-related services is unevenly represented in two main spheres. In formal, natural and experimental science (such as statistics, physics, mathematics, biology and astronomy), inter-departmental data labs or even supercomputing centres or grid computing infrastructures are common and have the technology and staff needed to provide highly specialized services for analyzing and managing large volumes of data. Libraries participate only sporadically in these cases, whereas in humanities and social and applied sciences (sociology, geography, economics and business administration), where such centres and services are much less common, they are provided primarily by academic libraries. Nonetheless, humanities and social and applied science are now also facing the eScience challenge (Williford and Henry 2012), which is particularly visible in areas such as geospatial information (GIS) (Williford and Henry 2012) and its applications in the various disciplines.

Academic libraries are deploying a four-fold response to eScience and the growing need to use research data: 1) hiring specialized staff (data librarians or data specialists) or furthering data management and analysis training for (generally reference) librarians; 2) intensifying the collection or compilation of and providing access to data sources; 3) participating in the development of institutional

data repositories to preserve and share original research data (Jahnke, Asher, and Keralis 2012); and 4) incorporating data literacy in their instructional programs and services (whose design should follow today's inescapable reference framework, namely the ACRL (2011b) recommendations on information literacy, as well as its "Guidelines for Instruction Programs in Academic Libraries."

This last regard constitutes the focus of the present paper, and several avenues have been identified:

- via the Web, with the publication of self-training resources (open to the public or for in-house use only);
- in the library itself, through reference service, one-on-one and on-demand or scheduled user training sessions;
- through face-to-face and online instruction, forming part of credit courses, either as specialized stand-alone instruction or, with instructors' cooperation, as instruction embedded in other subjects.

In addition to quantitative research and quantitative or statistical literacy methods, which have been in place for much longer, more recent examples of data literacy instruction provided by libraries include, for instance, the Massachusetts Institute of Technology's (MIT) Data Management and Publishing tutorial; the EDINA Research Data Management Training (MANTRA), the University of Edinburgh's Data Library and the University of Minnesota libraries' Data Management Course for Structural Engineers.

The literature has also begun to echo some of these experiences in the last few years. Stephenson and Caravello (2007), for instance, described the University of California, Los Angeles (UCLA) libraries' participation in several Sociology Department courses in which data/statistical literacy was introduced as part of information literacy training, using the ACRL (2008a) model as a reference. The primary aim was "to develop students' skills in searching for, retrieving, customizing and critically evaluating statistical resources" (Stephenson and Caravello 2007, 530). Wong (2010), in turn, reviewed the Hong Kong University of Science and Technology (HKUST) Library's experience with an elective on information literacy, which stressed the application of critical thinking to the assessment of socio-economic data sources. Miller and Fosmire (2008) described the participation of the Purdue University Libraries in the design of a geoinformatics course that provided a broader view of data than might be expected, given the specialized nature of the course, in which data acquisition management, visualization and sharing were addressed.

More recently, Stave (2012) reported the positive results attained by the instructional program for biomedical

studies delivered at Stanford University Medical School by the Lane Medical Library and Knowledge Management Centre. Among others, the centre offers a series of face-to-face courses on computation and statistics, whose key areas include data extraction, bias assessment and data analysis, R for statistics and graphics, and presenting results. Otto (2012) discussed the inclusion of data literacy in undergraduate and graduate planning methods courses as part of a cooperative program between Eastern Washington University (EWU) and the Urban and Regional Planning Department (URP), geared to developing skills for selecting, interpreting and using economic and demographic data sources. Johnston, Lafferty and Petsan (2012) described the data management training delivered by University of Minnesota (UMN) libraries for teaching staff and graduate students as support for the formulation of the data management plans required to opt for NSF funding.

As the foregoing shows, academic libraries are beginning to include a number of aspects of data literacy in their instructional programs in response to the growing importance of this issue for their users. Given such libraries' context and strategic positioning, then, their data literacy training offering can be expected to expand exponentially in the years to come. These activities may be more difficult to introduce in other types of libraries until the respective competencies are included in the information literacy models presently in place.

Data Literacy Competencies

The identification of the competencies needed to be data literate is a matter presently under study by a fairly large community of researchers and professionals in different domains. However, all these views of data literacy vary, depending on the context and purposes sought.

For the public domain, the School of Data (Open Knowledge Foundation and Peer to Peer University [P2PU] 2012) provides a number of training modules on data literacy-related areas in which the public should be proficient. These include data conceptualization, where to find data, how to use basic data analysis and graphical representation tools, critical interpretation of data, and the use of data as evidence to support an argument. The Australian Bureau of Statistics (ABS 2010), in turn, identifies the areas that it deems should be mastered by the general public:

1. Data awareness: Are the data relevant and appropriate? Where did the data come from? How were the data collected? Are the data fit for purpose?

2. The ability to understand statistical concepts: Basic forms of statistical representation; Different types of proportions; More complex statistical concepts.
3. The ability to analyze, interpret and evaluate statistical information: organize data, construct and display graphs and tables and work with different representations of the data; describe and summarize basic data; extract, understand and explain data that is presented in a variety of ways; comparison pitfalls; understand the context.
4. The ability to communicate statistical information and understandings: how are data reported? Confidentiality of ABS data.
 - Managing data in aggregation: data collections, data and users, organizational planning
 - Broader issues in science data management: archiving practices, data curation, enabling technologies, data presentation, and data sharing.

In “Statistics: Power from Data!” (2012), Statistics Canada lists a series of competencies involved in the critical use of data, such as understanding the utility of statistics in everyday life, how to design and implement a survey, the most frequent sampling methods, how data are processed, organized and presented, and how statistical information may be biased.

In pre-university education, Kent State University’s Research Centre for Educational Technology (2010) conducted a project in conjunction with SRI, funded by the National Science Foundation, from 2008 to 2010, entitled Thinking with Data (TWD). Its aim was to develop four modules for integrating data literacy in seventh grade social studies, mathematics, science and English language arts curricula. The educational resources developed sought three literacy goals: a) student understanding of the relationship between data and context; b) student use of appropriate data, tools, and representations; c) student understanding of the relationship of claims to evidence. The most prominent feature of its aims and the respective indicators was the importance attached to critical thinking in data evaluation, interpretation and use.

A number of significant university level initiatives have been developed on the competencies that define the various aspects of data literacy. Qin and D’Ignazio (2010b), for instance, stress the importance of scientific data management in future science workforce training and propose a term, science data literacy (SDL), defined to be “the ability to understand, use, and manage science data,” which entails “skills in collecting, processing, managing, evaluating, and using data for scientific inquiry.” Their Science Data Management undergraduate course at Syracuse University, designed from input provided by a group of STEM teaching staff, consists of the following modules and topics (2010a):

- Fundamentals of science data and data management: science data life cycle, databases, types of data, data sets description and data management
1. Overview of research data management: students must understand what research data are, their life cycle, the need to manage them and the impact of data management on the success of a project.
 2. Data: types, stages and formats: students must be aware of the various types of data, their formats and data collecting methods and policies.
 3. Metadata: students must know what metadata are and how to identify standards and practices applicable to a project.
 4. Data storage, backup, and security: students must understand the importance of these matters and the best practices against which to benchmark a project.
 5. Legal and ethical considerations: students must understand the ethical and legal implications of data use in terms of intellectual property, privacy and citations.
 6. Data sharing and re-use policies: students must understand the meaning of open science and open data, be aware of the conditions for data reuse, know how to convert one data format into another and be able to establish data access levels.
 7. Plan for archiving and preservation of data: students must be aware of data repository policies and purposes.

For Carlson *et al.* (2011, 633), data literacy implies “understanding what data mean, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways.” After assessing the needs specified by a group of science and engineering professors, as well as the ones detected in students of those disciplines, and analyzing information literacy standards, these authors identified the following as key areas: databases and data formats; discovery and acquisition of data; data management and organization; data conversion and interoperability; quality assurance; metadata; data curation and re-use; cultures of practice; data preservation; data analysis; data visualization; and ethics, including data citation.

Along lines similar to those drawn by Qin and D’Ignazio (2010b) and Carlson *et al.* (2011), the Lamar Soutter Library, University of Massachusetts Medical School and the George C. Gordon Library, Worcester Polytechnic Institute, developed a set of Frameworks for a Data Management Curriculum (2012) for science, health science and engineering graduate and undergraduate students. That curriculum consists of the seven modules listed below.

1. Overview of research data management: students must understand what research data are, their life cycle, the need to manage them and the impact of data management on the success of a project.
2. Data: types, stages and formats: students must be aware of the various types of data, their formats and data collecting methods and policies.
3. Metadata: students must know what metadata are and how to identify standards and practices applicable to a project.
4. Data storage, backup, and security: students must understand the importance of these matters and the best practices against which to benchmark a project.
5. Legal and ethical considerations: students must understand the ethical and legal implications of data use in terms of intellectual property, privacy and citations.
6. Data sharing and re-use policies: students must understand the meaning of open science and open data, be aware of the conditions for data reuse, know how to convert one data format into another and be able to establish data access levels.
7. Plan for archiving and preservation of data: students must be aware of data repository policies and purposes.

In the social sciences, interest focuses on data recovery, handling and critical use. Hogenboom, Phillips, and Hensley (2011, 414), working out of the University of Illinois at Urbana-Champaign library, reported that according to instructors, one priority in this area was “to help data users locate, access, extract, customize and evaluate data.” Wong (2010), in turn, stressed that students should be able to:

- describe what socioeconomic data are about;
- describe the differences between socioeconomic and scientific data;
- craft workable strategies to access various socioeconomic data;
- evaluate data quality on the basis of reliability and authority.

For Stephenson and Caravello (2007, 531), a data literacy module in a social science information literacy program should address the following competencies:

- the ability to read and critically evaluate simple 2x2 or 3-way tables;
- the ability to produce accurate bibliographic citations for data tables;
- the ability to use American FactFinder to create a table, which they could describe and cite correctly;
- the ability to read an article containing a graphical representation of data and discuss the table in relation to the article content.

Two major trends can, then, be distinguished: one that links data literacy to statistical literacy in the critical use of data and the other that relates data literacy to data management. All, however, lack a common reference framework applicable to the various contexts and purposes sought with such competencies.

Core Competencies and Contents for Data Literacy Instruction

The preceding discussion shows that many different views and competencies are currently being associated with data literacy. The question, then, is: when designing standard data literacy instruction, what competencies should be covered? A common reference framework might be useful in order to promote library adoption on wide basis. Such a framework may be built on the commonalities found among the competencies so far presented as well as including any others that might be considered appropriate. The following framework is inspired by the general

structure of information literacy standards and includes most of the aforementioned common competencies. Additionally, the framework translates competencies into instructional topics or units to facilitate interpretation and direct implementation.

1. Understanding data.

1.1. What is data?

Competency: learners need to know what is meant by data and be aware of the various possible types of data.

Contents: Data definition; Types of data (depending on origin, format, usage license and so on).

1.2. Data in society: a tool for knowledge and innovation.

Competency: learners need to be aware of the role of data in society, how they are generated and by whom, and their possible applications, as well as the implications of their use.

Contents: Data producers and consumers; Data lifecycle; Data applications: their impact on science and society; Copyright and licenses influencing data reuse.

2. Finding and/or obtaining data.

2.1. Data sources.

Competency: learners need to be aware of the possible data sources, be able to evaluate them and select the ones most relevant to an informational need or a given problem.

Contents: Data sources; Criteria for assessing data sources.

2.2. Obtaining data.

Competency: learners need to be able to detect when a given problem or need cannot be (totally or partially) solved with the existing data and, as appropriate, undertake research to obtain new data.

Contents: Main research methods for obtaining original data.

3. Reading, interpreting and evaluating data.

3.1. Reading and interpreting data.

Competency: learners need to be aware of the various forms in which data can be presented (written, numerical or graphic), and their respective conventions, and be able to interpret them.

Contents: Ways to present and represent data.

3.2. Evaluating data.

Competency: learners need to be able to evaluate data critically.

Content: Data evaluation criteria (including authorship, method of obtaining and analyzing data, comparability, inference and data summaries).

4. Managing data.

4.1. Data and metadata collection and management.

Competency: learners need to be aware of the need to save the data selected or generated and of descriptive or other data associated therewith, for due identification, management and subsequent reuse.

Content: Metadata; Reference management tools; Databases; Data management repositories: policies and practices.

5. Using data.

5.1. Data handling.

Competency: learners need to be able to prepare data for analysis, analyze them in keeping with the results sought and know how to use the necessary tools.

Contents: Data conversion; Handling data analysis tools, both locally (Excel, R, SPSS, Stata or similar) and on-line.

5.2. Producing elements for data synthesis.

Competency: learners need to be able to synthesize and represent the results of data analysis in ways suited to the nature of the data, their purpose and the audience targeted in the inquiry.

Contents: Choosing suitable data representation methods (tables, graphs or similar); Handling tools (built into analytical tools or stand-alone applications such as Gapminder, Visual.ly or IBM's Many Eyes).

5.3. Ethical use of data.

Competency: learners need to make ethical use of data, acknowledging the source when obtained or formulated by others, and making sure that used methods are deployed and results interpreted transparently and honestly.

Contents: What is the ethical use of data; How to cite data sources.

Although the scheme proposed aspires to be universal, covering the core of what would be potentially applicable to school, public or academic libraries, the key to its success will lie in the depth to or specificity with which it is developed, after adaptation to each library's particular needs. While in public and school libraries teaching will target the statistics appearing in the media, in universities

the primary target will be the data from official sources or original research. Public and school library instruction on data management, for instance, may need to cover no more than basic spreadsheet and database use, while in universities training may include such complex topics as the formulation of data management plans for research projects.

Conclusions

Data literacy today is a competency as essential as information literacy. The two are complementary and clearly form part of libraries' educational role in the furtherance of the significant use of information resources to generate knowledge and innovate. Its inclusion in libraries' instructional programs is therefore wholly justified. To date, however, data literacy has not received much attention by libraries outside higher education, where it has aroused substantial interest, driven particularly by eScience.

The framework of core competencies and contents outlined above aims to contribute to the advancement of data literacy in library instructional practice, inasmuch as it defines an understructure on which learning experiences can be built. On these grounds, libraries may, for instance, begin to design and develop a series of instructional resources (learning objects) for their websites that cover the proposed content, either created a propos or selected (and adapted, when necessary) from the relevant content openly available on the Web. This could whet the appetite of potential users to receive data literacy instructional services and serve as a basis for conducting new training activities and cooperating with teaching staff for inclusion in the wider curriculum.

These considerations raise new questions: how can the model be adapted to different types of libraries and different disciplines? and how will it impact the acquisition of data literacy by different types of users? Future studies will address these questions, among others, raised by the present research.

References

- American Association of School Librarians [AASL]. 2011. "Crosswalk of the Common Core Standards and the Standards for the 21st-Century Learner." Accessed September 23, 2012. <http://www.ala.org/aasl/guidelinesandstandards/commoncorecrosswalk>.
- American Statistical Association (ASA). 2011. "American Statistical Association Urges Support of Statistical Literacy Bill." Ac-

- cessed April 23, 2013. <http://www.amstat.org/newsroom/pressreleases/ASAUrgesSupportofSLB.pdf>.
- Association of College and Research Libraries (ACRL). 2000. "Information Literacy Standards for Higher Education." Accessed September 19, 2012. <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf>.
- Association of College and Research Libraries (ACRL). 2006. "Information Literacy Standards for Science and Engineering/Tech-nology." Accessed September 25, 2012. <http://www.ala.org/acrl/standards/infolitstech>.
- Association of College and Research Libraries (ACRL). 2008a. "Information Literacy Standards for Anthropology and Sociology Students." Accessed September 25, 2012. http://www.ala.org/acrl/standards/anthro_soc_standards.
- Association of College and Research Libraries (ACRL). 2008b. "Political Science Research Competency Guidelines." Accessed September 30, 2012. <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/PolisciGuide.pdf>.
- Association of College and Research Libraries (ACRL). 2011a. "Information Literacy Competency Standards for Journalism Students and Professionals." Accessed September 30, 2012. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/il_journalism.pdf.
- Association of College and Research Libraries (ACRL). 2011b. "Guidelines for Instruction Programs in Academic Libraries." Accessed December 24, 2012. <http://www.ala.org/acrl/standards/guidelinesinstruction>.
- Australian Bureau of Statistics. 2010. "What Is Statistical Literacy and Why Is It Important to Be Statistically Literate? (Feature Article)." *Tasmanian State and Regional Indicators* (March). Accessed September 26, 2012. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1307.6Feature+Article1Mar+2009>.
- Bennett, T. B., and S. W. Nicholson. 2007. "Research Libraries: Connecting Users to Numeric and Spatial Resources." *Social Science Computer Review* 25 (3): 302–318.
- Berners-Lee, T. 2010. "Long Live the Web: A Call for Continued Open Standards and Neutrality." *Scientific American* 2010 (12, 22 November). Accessed March 21, 2012. <http://www.scientificamerican.com/article.cfm?id=long-live-the-web&print=true>.
- Bundy, A. 2004. "Australian and New Zealand Information Literacy Framework." *Principles, Standards and Practice* 2. Adelaide, South Australia: Australian and New Zealand Institute for Information Literacy. Accessed April 7, 2013. <http://www.literacyhub.org/documents/InfoLiteracyFramework.pdf>.
- Carlson, J., M. Fosmire, C. C. Miller, and M. S. Nelson. 2011. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *Portal: Libraries and the Academy* 11 (2): 629–657.
- European Commission. 2003. "Re-use of Public Sector Information. Review of Directive 2003/98/EC." Accessed March 21, 2012. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0212:FIN:EN:PDF>.
- European Commission. 2007. "Science Education Now: A Renewed Pedagogy for the Future of Europe." Accessed September 22, 2012. http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf.
- European Commission. 2009. "Malmö Ministerial Declaration on E-Government." Accessed March 21, 2012. <http://www.egov2009.se/wp-content/uploads/Ministerial-Declaration-on-eGovernment.pdf>.
- European Commission. 2012. "Towards Better Access to Scientific Information: Boosting the Benefits of Public Investments in Research." Accessed September 12, 2012. http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf.
- Gal, I. 2002. "Adults' Statistical Literacy: Meanings, Components, Responsibilities." *International Statistical Review* 70 (1): 1–25.
- Gray, A. S. 2004. "Data and Statistical Literacy for Librarians." *IASSIST Quarterly* 28 (2/3): 24–29.
- Haendel, M. A., N. A. Vasilevsky, and J. A. Wirz. 2012. "Dealing with Data: A Case Study on Information and Data Management Literacy." *PLoS Biology* 10 (5): e1001339. Accessed April 7, 2013. <http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001339>.
- Hogenboom, K., C. M. H. Phillips, and M. Hensley. 2011. "Show Me the Data! Partnering with Instructors to Teach Data Literacy." In *Declaration of Interdependence: The Proceedings of the ACRL 2011 Conference*, March 30-April 2, Philadelphia, PA, 410-417.: Chicago, IL: ACRL. Accessed September 11, 2012. http://0-www.ala.org.sapl.sat.lib.tx.us/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/national/2011/papers/show_me_the_data.pdf.
- Hunt, K. 2004. "The Challenges of Integrating Data Literacy into the Curriculum in an Undergraduate Institution." *IASSIST Quarterly* (Summer/Fall): 12-15. Accessed April 20, 2013. http://www.iassistdata.org/downloads/iqvol282_3hunt.pdf.
- International Association for Social Science Information Services and Technology (IASSIST). 2009. "Democratizing Data: The IASSIST Strategic Plan for 2010-2014." Accessed April 23, 2013. http://www.iassistdata.org/sites/default/files/strategic_plan_2010-14.pdf.
- International Statistical Institute [(SI)]. 2013. "International Statistical Literacy Project." Accessed April 23, 2013. <http://iase-web.org/islp/>.
- Jahnke, L., A. Asher, and S. D. C. Keralis. 2012. *The Problem of Data*. Washington, DC: Council on Library and Information Resources (CLIR) Report, Pub. 154. Accessed November 22, 2012. http://digitalcommons.bucknell.edu/fac_pubs/52/.
- Johnston, L., M. Lafferty, and B. Petsan. 2012. "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach." *Journal of eScience Librarianship* 1 (2): Article 2.
- Kellam, L. M. 2011. *Numeric Data Services and Sources for the General Reference Librarian*. Oxford: Chandos Publishing.
- Kent State University, Research Center for Educational Technology. 2010. "Thinking with Data: Project Overview." Accessed December 17, 2012. <http://www.rcet.org/twd/project/projectoverview.html>.
- Merrill, A. 2011. "Library+." *Public Services Quarterly* 7 (3/4): 144-148.
- Miller, C. C., and M. Fosmire. 2008. "Creating a Culture of Data Integration and Interoperability: Librarians Collaborate on a Geoinformatics Course." *IATUL Annual Conference Proceedings* 18. Accessed April 7, 2013. <http://www.iatul.org/conferences/pastconferences/2008proceedings.asp>.
- OECD. 2004. "Science, Technology and Innovation for the 21st Century." Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January – Final

- Communique. Accessed September 12, 2012. <http://www.oecd.org/science/scienceandtechnologypolicy/name,79792,en.htm>.
- Open Government Partnership. 2012. "Participating Countries." Accessed March 21, 2012. <http://www.opengovpartnership.org/countries>.
- Open Knowledge Foundation, and Peer to Peer University (P2PU). 2012. "School of Data - Learn How to Find, Process, Analyze and Visualize Data." Accessed December 12, 2012. <http://schoolofdata.org/>.
- Otto, J. L. 2012. "Assessing and Improving Data Literacy: A Study with Urban and Regional Planning Students." *PNLA Quarterly* 76 (4). Accessed September 15, 2012. <http://unlib.unl.edu/LPP/PNLA%20Quarterly/otto76-4.pdf>.
- Owens, M. R. 1976. "The State Government and Libraries." *Library Journal* 101 (1): 19-28.
- Qin, J., and J. D'Ignazio. 2010a. "The Central Role of Metadata in a Science Data Literacy Course." *Journal of Library Metadata* 10 (2-3): 188-204.
- Qin, J., and J. D'Ignazio. 2010b. "Lessons Learned from a Two-year Experience in Science Data Literacy Education." In *The Evolving World of e-Science: Impact and Implications for Science and Technology Libraries: Proceedings of the 31st Annual IATUL Conference*, 20 - 24 June, Purdue University, West Lafayette, USA. Accessed April 7, 2013. <http://www.iatul.org/conferences/pastconferences/>.
- Schild, M. 2004. "Information Literacy, Statistical Literacy and Data Literacy." *IASSIST Quarterly* 28 (2/3): 6-11.
- Society of College, National and University Libraries (SCONUL). 2011a. "The SCONUL Seven Pillars of Information Literacy: Core Model for Higher Education." Accessed September 19, 2012. https://www.sconul.ac.uk/groups/information_literacy/publications/coremodel.pdf.
- Society of College, National and University Libraries (SCONUL). 2011b. "The SCONUL Seven Pillars of Information Literacy: A Research Lens for Higher Education." Accessed September 19, 2012. https://www.sconul.ac.uk/groups/information_literacy/publications/researchlens.pdf.
- Society of College, National and University Libraries (SCONUL). 2011c. "The SCONUL Seven Pillars of Information Literacy Through an Open Content Lens." Accessed September 19, 2012. https://www.sconul.ac.uk/groups/information_literacy/publications/OER_lens.docx.
- Society of College, National and University Libraries (SCONUL). 2012. "The SCONUL Seven Pillars of Information Literacy Through a Digital Literacy Lens." Accessed September 19, 2012. https://www.sconul.ac.uk/groups/information_literacy/publications/SCONUL%20digital_literacy_lens_v4.doc.
- Soehner, C., C. Steeves, and J. Ward. 2010. *E-Science and Data Support Services: A Study of ARL Member Institutions*. Washington, DC: Association of Research Libraries. Accessed October 10, 2012. http://www.arl.org/bm~doc/escience_report2010.pdf.
- Statistics Canada. 2012. "Statistics: Power from Data!" Accessed December 17, 2012. <http://www.statcan.gc.ca/edu/power-pouvoir/toc-tdm/5214718-eng.htm>.
- Stave, C. 2012. "Collaboration, Innovation and Diversity: Keys to Building a Cost-Effective and High-Impact Biomedical Instructional Program." Paper presented at the 2011 Special Library Association Conference, Chicago, IL, 15-18 July. Accessed October 10, 2012. <http://www.sla.org/pdfs/sla2012/edicalLifeSciDivContributedPapersChrisStave.pdf>.
- Steele, B., and S. Kiliç-Bahi. 2008. "Quantitative Literacy Across the Curriculum: A Case Study." *Numeracy* 1 (2 July). Accessed September 22, 2012. <http://scholarcommons.usf.edu/numeracy/vol1/iss2/art3/>.
- Steen, L. A. 2004. "Educational Goals for Quantitative Literacy." In *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education*, 21-26. Washington, DC: The Mathematical Association of America. Accessed September 22, 2012. http://www.in.gov/che/files/Ed_goals_for_QL.pdf.
- Stephenson, E., and P. S. Caravello. 2007. "Incorporating Data Literacy into Undergraduate Information Literacy Programs in the Social Sciences: A Pilot Project." *Reference Services Review* 35 (4): 525-540.
- Sturges, P., and A. Gasteringer. 2010. "Information Literacy as a Human Right." *Libri* 60 (3): 195-202.
- Sunlight Foundation. 2010. "Ten Principles for Opening Up Government Information." Accessed March 21, 2012. <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>.
- Thompson, K., and D. M. Edelstein. 2005. "A Reference Model for Providing Statistical Consulting Services in an Academic Library Setting." *IASSIST Quarterly* 2004 (Summer/Fall): 35. Accessed October 10, 2012. http://www.iassistdata.org/downloads/iqvol282_3thompson.pdf.
- Treadwell, W., and J. A. Cogswell. 1994. "The Machine Readable Data Center: A Model Approach to Data Services in Academic Research Libraries." *Library Hi Tech* 12 (1): 87-92.
- UNESCO, and International Council for Science (ICSU). 2002. *Science for the Twenty-First Century: A New Commitment*. Paris: Unesco. Accessed September 21, 2012. http://www.unesco.org/science/wcs/report_wcs.pdf.
- United Nations Economic Commission for Europe. 2012. *Making Data Meaningful. Part 4: A Guide to Improving Statistical Literacy*. Geneva: UNECE. Accessed September 19, 2012. http://www.unece.org/fileadmin/DAM/stats/documents/writing/Making_Data_Meaningful_Part_4_for_Web.pdf.
- United States Office of Science and Technology Policy. 2012. "Obama Administration Unveils 'Big Data' Initiative: Announces \$200 Million in New R&D Investments." Accessed September 12, 2012. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.
- University of Massachusetts Medical School, Lamar Soutter Library, and Worcester Polytechnic Institute, George C. Gordon Library. 2012. *Frameworks for a Data Management Curriculum: Course Plans for Data Management Instruction to Undergraduate and Graduate Students in Science, Health Sciences and Engineering Programs*. Amherst, MA: University of Massachusetts Medical School Lamar Soutter Library. Accessed October 9, 2012. http://library.umassmed.edu/data_management_frameworks.pdf.
- Wallman, K. K. 1993. "Enhancing Statistical Literacy: Enriching Our Society." *Journal of the American Statistical Association* 88 (421, March 1): 1-8.
- Watson, J. M. 1997. "Assessing Statistical Thinking Using the Media." In *The Assessment Challenge in Statistics Education*, edited by I. Gal and J. B. Garfield, 107-121. Amsterdam: IOS Press and The International Statistical Institute. Accessed

- September 21, 2012. <http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter09.pdf>.
- The White House. 2009. "Open Government Directive." Accessed March 21, 2012. <http://www.whitehouse.gov/open/documents/open-government-directive>.
- The White House. 2012. "Educate to Innovate." Accessed September 22, 2012. <http://www.whitehouse.gov/issues/education/k-12/educate-innovate>.
- Williford, C., and C. Henry. 2012. *One Culture: Computationally Intensive Research in the Humanities and Social Sciences*. Washington, DC: Council on Library and Information Resources.
- Accessed November 22, 2012. <http://www.clir.org/pubs/reports/pub151/pub151.pdf>.
- Wong, G. K. W. 2010. "Facilitating Students' Intellectual Growth in Information Literacy Teaching." *Reference and User Services Quarterly* 50 (2): 114–118.
- Editorial history
received 10 January 2013
final version received 7 February 2013
accepted 11 February 2013