

**This item is not the definitive copy.
Please use the following citation when referencing this
material:**

Koltay, T. (2016). Data Governance, Data Literacy and the Management of Data Quality. *IFLA Journal*, 42(4): 303-312.

Data Governance, Data Literacy and the Management of Data Quality

Abstract

Data governance and data literacy are two important building blocks in the knowledge base of information professionals, involved in supporting data-intensive research, and both address data quality and research data management.

Applying data governance to research data management processes and data literacy education helps in delineating decision domains and defining accountability for decision making. Adopting data governance is advantageous, because it is a service based on standardised, repeatable processes and is designed to enable the transparency of data-related processes and cost reduction. It is also useful, because it refers to rules, policies, standards; decision rights; accountabilities and methods of enforcement. Therefore, although it received more attention in corporate settings and some of the skills related to it are already possessed by librarians, knowledge on data governance is foundational for research data services, especially as it appears on all levels of research data services, and is applicable to big data.

Keywords

Data-intensive research, data librarian, data governance, data literacy, research data services

Introduction

Data intensive science, coupled with mandates for data management plans and open data from research funders, has led to a growing emphasis on research data management both in academia and in academic libraries. The role of the latter is changing, so academic librarians are often integrated in the research process, first of all in the framework of research data services (RDSs) (Tenopir et al., 2015). Therefore, it comes as no surprise that supporting data-intensive research is a top trend in academic library work (ACRL, 2014; NMC, 2014). It is in focus especially because it gives chance to change the present situation, where faculty and researchers regard the library not as a place of real-time research support, but only as a dispensary of books and articles (Jahnke, Asher and Keralis, 2012).

On this background, a review of the literature was done in order to identify and examine significant constituents of the knowledge base that is crucial for information professionals, involved in supporting data-intensive research. The first constituent is data governance (DG), which is extensively dealt with mainly in the corporate (business) sector, and is explored in this paper with the belief that bringing it into the picture will enable better research data services. The second one is data literacy, about which there is a massive body of literature, among others in the form of review articles (MacMillan, 2014; Koltay, 2015a, Koltay 2015b). Data literacy is closely related to research data services that include research data management (RDM). As the concept of RDSs itself and data literacy education are still evolving, their relationship to data governance requires examination that may lead to some kind of synthesis. The management of data quality is also inspected in order to determine to what extent it plays the role of an interface between these two constituents.

Accordingly, this writing is built on three core terms. *Data governance* can be defined as the exercise of decision-making and authority that comprises a system of decision rights and accountabilities that is based on agreed-upon models, which describe who can take what actions, when and under what circumstances, using what methods (DGI, 2015a). While the various definitions of *data literacy* will be discussed below, we define it here as the ability to process, sort, and filter vast quantities of information, which requires knowing how to search, how to filter and process, to produce and synthesize it (Johnson, 2012). This definition is in accordance with the idea, expressed by Schneider (2013), that the boundaries between *information* in information literacy and data literacy are blurring, because these boundaries never have been rigid.

Research data services consist of a wide spectre of informational and technical services that a library offers to researchers in managing the full data life cycle (Tenopir, Birch and Allard, 2012).

Research data services and the paradigms of academic library management

A better understanding of the academic libraries' role in the data-intensive environment can be obtained if we place them into the context academic librarianship's past and present development paradigms, outlined by Martell (2009). The first paradigm, called the 'Ownership' or 'Collections' paradigm evolved after the World War II and reached its zenith in the 1960s. It was built on the assumption that campus library systems would be able to collect all documents that can satisfy adequately the institutions' scholarly and teaching needs. Such support adequately allowed for a broad range of interpretations, but it proved to be unsustainable and was supplanted by the 'Access' paradigm

that directed more attention to and made use of resource sharing from the late 1970s until the end of the 20th century. Widespread access to digital material, in particular the availability of electronic full text of serials made ownership in its traditional sense not practical, so the 'iAccess' paradigm took place. More recently, the emergence and growing prevalence of social media creates an opportunity to add social dimension to iAccess, forging in this way the 'sAccess' paradigm.

While social media undoubtedly plays a role in Research 2.0, it is often difficult to disentangle the relationship between features of induced by its presence and by the growing importance of data. Social media influences academic libraries in many ways. It produces enormous quantities of (big) data that can be analysed, published and reused mainly by researchers in the social sciences (Boyd and Crawford, 2012). It also changes the ways, how research is done, even though the lack of trust in social media channels for scholarly communication lessens its impact (Nicholas et al, 2014). Therefore, it is a demanding task to define to what extent data-intensive research pertains to iAccess and to sAccess. In any case, both paradigms have influence in it to some degree.

Data governance in detail

As said above, data governance is a subject of interest for the business sector. Therefore, it is rarely addressed by the LIS literature. A notable exception is the work of Krier and Strasser (2014) that focuses on data management in libraries. A review of definitions of data government by Smith (2007) clearly shows the close ties of DG to the business sector. Besides providing a set of definitions that relate it to companies, enterprises and business, Smith underlines that "the process of data governance is to exercise control over the data within a corporate alignment".

It seems to be clear that the academic sector, librarianship, as well as library and information science also should pay attention to DG, albeit it raised attention mainly in the business sector. Even though rather implicitly, this need is asserted by DosSantos (2015), who points out that the role of the data governor must shift to be something more akin to a data librarian in order to make data governance the driving force behind business innovation, instead of being an impediment to data. This goal can be attained by delivering information technology as a service and by enabling the processes of locating and organizing the best available data.

The expression *data governance* could refer to organizational bodies; rules, policies, standards; decision rights; accountabilities and methods of enforcement. DG enables better decision-making and protects the needs of

stakeholders. It reduces operational friction and encourages adopting common approaches to data issues. Data governance also helps building standard, repeatable processes, reducing costs and increasing effectiveness through coordination of efforts and by enabling transparency of processes. It is governed by the principles of integrity, transparency, auditability (DGI, 2015a). DG also delineates decision domains, i.e. what decisions must be made to ensure effective management and use of the organization's assets. It also defines the locus of accountability for decision making by defining, who is entitled to make decisions in a given organization, and who is held accountable for the decision-making related to data assets (Weill and Ross, 2004; Khatri and Brown, 2010). Seiner (2014) adds to this that valid data governance may require identifying "people who informally already have a level of accountability for the data they define, produce and use to complete their jobs or functions". One of the reasons for this is that correct and efficient governance depends as much on technology as on organizational culture, despite the fact that good governance technology makes data transparent, gives it accountability and helps identify areas, where performance can be improved (ORACLE, 2015). Accountabilities, the main components of which are stewardship and standardization, are defined in a manner that introduces checks-and-balances between different teams, between those who create and collect information, others who manage it, those who use it, and those who introduce standards and compliance requirements (DGI, 2015b). As stewardship appears in this list and is also present in several resources, related to research data management (Bailey, 2015), and because it is sometimes used interchangeably with DG, some clarification is needed. Data stewardship is concerned with taking care of data assets that do not belong to the stewards themselves, thus data stewards represent the concerns of others, and ensure that data-related work is performed according to policies and practices as determined through governance. In contrast, data governance is an overall process that brings together cross-functional teams (including data stewards and/or data governors) to make interdependent rules or to resolve issues and to provide services to data stakeholders (Rosenbaum, 2010). To be successful, data governance needs to have clear definitions of its objectives, processes, and metrics. It has to create its own processes and standards. Besides roles and responsibilities for all data governance roles, communities of practice for governance, stewardship, information management have to be established. Change management processes also have to be instituted, and – last, but not least – there have to be rewards for good data governance behaviour.

Data governance should not be optional, because it contributes to organizational success through repeatable and compliant practices. In the sense of managing, monitoring and measuring different aspects of an organization, governance can be related to managing information technology, people and other tangible resources. Data is everywhere, thus DG runs horizontal. Definitions of the data and how to use it are part of the data management process, while integrating data into the organization and establishing individuals to oversee the administration of data processes pertain to data governance. DG also must include metadata, unstructured data, registries, taxonomies and ontologies (Smith, 2007).

The traditional principles of DG also apply to big data. From among big data types, data from the web and from social media, as well as machine to machine data deserve distinguished attention. Big data governance is especially important in regard to the acceptable use of data (Soares, 2012). In environments, where big data plays substantial role, one of the most common data integration mistakes is underestimating data governance (ORACLE, 2015). Although big data integration differs from traditional data integration by many factors (Dong and Srivastava, 2013), it demonstrates the complexity and importance of data governance. Data integration itself can be defined as the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. It helps to understand, cleanse, monitor, transform and deliver data, thus it supplies trusted data from a variety of sources (IBM, 2016). Data integration solves the problems related to combining data of varied provenance by presenting a unified view of these data (Lenzerini, 2002).

As Sarsfield (2009) put it, DG is like an elephant in a dark room. It can be perceived depending on where you touch it. If you touch its tail, it feels like a snake. If you touch one of its legs, it feels like a tree. Therefore, cross-functional perspectives on data governance vary, and we will take this variability into consideration to couple it with data quality and data literacy.

In research settings, the stakeholders of DG are researchers, research institutions, funders, publishers and the public at large. A good understanding of data governance also addresses researchers' fear of lost rights and benefits. Governance structures are needed for managing human subjects-related data, as well, because taking care of sensitive information requires not only establishing standards and norms of practice, but fostering culture change towards better data stewardship (Hartter et al., 2013). Besides of these functions, data governance in this environment enables proper access and sharing (Riley, 2015), even if data ownership is often ambiguous, because if someone has a stake in research data, it does not mean that they are owners of

that data (Briney, 2015). Many of DG skills, such as dealing with licensing terms and agreements, as well as knowledge about copyright are already possessed by librarians (Krier and Strasser, 2014).

Altogether, data governance is the starting point for managing data. A formal data governance program has to provide answers to questions, such as the availability and access possibilities, provenance, meaning, trustworthiness. As a shared responsibility among all constituents of an institution, it is also required to provide coordinated, cross-functional approaches and to facilitate best practices. It both prevents the misuse of institutional data assets and encourages more effective use of these same data assets by the institution itself (ECAR, 2015). Being knowledgeable of data governance's nature is foundational for RDSs and well-developed data governance is one of the necessary conditions for open data (Weber, Palmer and Chao, 2012), even though it is also one of the most challenging issues of data sharing (Krier and Strasser, 2014).

Data governance and managing data quality

Data governance also “guarantees that data can be trusted and that people can be made accountable for any adverse event that happens because poor quality” (Sarsfield, 2009: 38). In a similar vein, Khatri and Brown (2010) underline that governance includes establishing who in the organization holds decision rights for determining standards for data quality. Data management involves determining the actual criteria employed for data quality, while DG is about designating who should make these decisions. According to Seiner (2014), DG formalizes not only behaviour related to the definition, production, and usage of data, but its quality. Similarly, a white paper by Information Builders emphasizes that data governance is a critical component of any data quality management strategy (Breaking Big, 2014). Another white paper, titled *Successful information governance through high-quality data*, underlines that the success of an information governance program depends on robust data quality that can be achieved if we reduce the proliferation of incorrect or inconsistent data by continuous analysis and monitoring (IBM, 2012).

Data quality is one of the cornerstones of the data-intensive paradigm of scientific research. This is true, even if it is difficult to appraise data, because appraisal requires deep disciplinary knowledge and manually appraising data sets is very time consuming and expensive, while automated approaches are in their infancy (Ramírez, 2011). In the academic sphere, the problem of data quality has been relatively well elaborated, thus an exhaustive further treatment of it is not needed. Nonetheless, let us repeat its most notable factors, which are availability and discoverability, trust and authenticity, acceptability, accuracy

(comprising correctness and consistency), applicability, integrity, completeness, understandability and usability (IBM, 2012). It is also clear that research data services, offered by academic libraries could play a critical role as data quality hubs on campus, by providing data quality auditing and verification services for the research communities (Giarlo, 2013). While caring for the availability of data would be a self-explanatory requirement, set towards data librarians, being knowledgeable of the ways to assess the digital objects' authenticity, integrity and accuracy over time also would be useful (Madrid, 2013). More recently, Zilinski and Nelson (2014) identified some other factors of data quality as coverage and relevance to the given research question and format, comprising of fields and units used, naming conventions, dates of creation and update. They also direct our attention to a set of quality control attributes that are akin to data governance that answer the question if quality control is explicitly outlined by examining who is in charge of checking for quality and what processes do they use.

Successful data governance depends not only on provisions related to roles in general, but responsibilities connected with appropriate data standards and managed metadata environments (Smith, 2007). Therefore, managing metadata is one of the key quality-related processes of data governance because it enables – among others – documenting the provenance of data that enables securing its quality (ORACLE, 2015).

Data governance, data quality and data literacy

To illustrate the importance of appropriate DG, we can take the case study, presented by Soares (2012) about the unfortunate events surrounding the Mars Climate Orbiter. In 1999, a navigation error directed the Orbiter into a trajectory 170 kilometres lower than the intended altitude above Mars, because NASA's engineers used English units (pounds) instead of NASA specified metric units (newtons). This relatively minor mistake resulted in a huge miscalculation in orbital altitude and in the loss of \$328 million. With appropriate attention to data governance principles, to the actual details and having mobilized data literacy skills, this accident could have been avoided.

Even though data literacy is going through a gestation period (Carlson and Johnston, 2015), being data literate begins to be widely accepted as a crucial ability for information professionals, involved in supporting data-intensive research (Qin and D'Ignazio, 2010; Schneider, 2013; Koltay, 2015b). On the other hand, the terminology in the field of data literacy is still not standardized. There is *science data literacy* (Qin and D'Ignazio, 2010) and *research data literacy* (Schneider, 2013). Carlson et al. (2011) argue for *data information literacy* because – according to their approach – it differs from a more restricted

meaning of data literacy, i.e. the ability of reading graphs and charts appropriately, drawing correct conclusions from data, and recognizing when data is being used in misleading or inappropriate ways. In the following, naming differences will be disregarded, and we will vote for the term *data literacy* first of all because this term is simple and straightforward (Koltay, 2015a), while it does not seem to have the limitation, mentioned by Carlson et al. (2011). Besides of this, while the terms differ, definitions and competence lists show convergence. If we look to the development of data literacy's definitions, we can see that in 2008 Fosmire and Miller spoke simply about information literacy in the data world (Fosmire and Miller, 2008). Two years later, data literacy was defined plainly as the ability to understand, use and manage data (Qin and D'Ignazio (2010). According to a 2013 definition, data literacy enables individuals to access, interpret, critically assess, manage, handle and ethically use data (Calzada Prado and Marzal (2013).

As mentioned above, Johnson (2012) described data literacy in much more detail, defining it as the ability to process, sort, and filter vast quantities of information, which requires knowing how to search, how to filter and process, to produce and synthesize. It is clear that these attributes are basically identical with the characteristics of information literacy as they appear in the well-known and widely accepted definition of information literacy, which comprises the abilities to recognize information need, identify, locate, evaluate, and use information to solve a particular problem (ALA, 1989). Nonetheless, it has to be added that – while information literacy seems to essentially enable us to efficiently process all types of information content (Badke, 2010) – the community of practice for data librarians differs from the one of information literacy (Carlson and Johnston, 2015).

As to the similarities to information literacy, it has to be added that several authors emphasize it. The Australian and New Zealand Information Literacy Framework, edited by Alan Bundy states that information literate persons obtain, store and disseminate not only text, but data, as well (ANZIIL, 2004). Andretta et al. (2008) identified presenting, evaluating and interpreting qualitative and quantitative data as a learning outcome of information literacy. According to Hunt (2004), data literacy education should borrow heavily from information literacy education, even if the domain of data literacy is more fragmented than the field of information literacy. Schneider (2013) also defined data literacy as a component of information literacy.

Both the SCONUL Seven Pillars information literacy model (2011) and the information literacy lens on the Vitae Researcher Development Framework (Vitae, 2012) stress that to identify which information could provide the best material to answer an information need, finding, producing and dealing with

research data is important, as information literacy today not only encompasses published information and underlying data. This is in accordance with a broader interpretation of information literacy, which recognizes that the concept of information includes research data (RIN, 2011). Carlson et al. (2011) underline that expanding the scope of information literacy to include data management and curation is a logical development. Si et al., (2013) state that data-related services should be supported by professionals with excellent information literacy skills.

Even though without referring to data literacy, Wang (2013) mentions that reference librarians frequently conduct information literacy sessions that educate the users about the existing data resources for their specific study areas.

Calzada Prado and Marzal (2013) state that information literacy and data literacy form part of a scientific-investigative educational continuum, a gradual process of education that begins in school, is perfected and becomes specialized in higher education and becomes part of lifelong learning. When suggesting a new framework for data literacy education, Maybee and Zilinski (2015) also point towards the close relationship between information literacy and data literacy.

Beyond definitions, applying and analysis of several information literacy standards, Calzada Prado and Marzal (2013) identified a number of abilities, some of which clearly show their origin in this best known definition of information literacy (ALA, 1989) and the *Information Literacy Competency Standards for Higher Education* (ACRL, 2000).

- Determining when data is needed;
- Accessing data sources appropriate to the information needed;
- Recognizing source data value, types and formats;
- Critically assessing data and its sources;
- Knowing how to select and synthesize data and combine it with other information sources and prior knowledge;
- Using data ethically;
- Applying results to learning, decision making or problem-solving.

They also emphasize the ability of identifying the context in which data is produced and reused. By mentioning these two main components of the data lifecycle they are in line with contemporary views of information literacy that incorporate the understanding of how information is produced (ALA, 2013).

Mandinach and Gummer (2013) identify data literacy as the ability to understand and use data effectively to inform decisions. With this, they give weight to data literacy's role in supporting decisions making. Therefore, they

bring data literacy up to data governance, recognising that it may be tied to the world of business.

Data literacy, as it is understood by the Association of College and Research Libraries, focuses on understanding how to find and evaluate data, giving emphasis to the version of the given dataset, the person responsible for it, and does not neglect the questions of citing and ethical use of data (ACRL, 2013).

Taking all these definitions together, data literacy can be defined as a specific skill set and knowledge base, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage, and ethically use data (Koltay, 2015).

Searle et al. (2015) identify data literacy as one of RDSs activities that support researchers in building the skills and knowledge required to manage data well. Therefore, we can say that data literacy is related to practically all processes that are covered by RDSs, and build the main framework of libraries' involvement in supporting the data-intensive paradigm of research (Tenopir et al., 2014). RDSs are undoubtedly comprehensive, thus covering its aspects, makes data literacy overarching and comprehensive.

When taking the closeness of data literacy to information literacy into consideration, it is intriguing to contemplate if there is such a thing as generic information literacy.

According to Carlson et al. (2011), data information literacy programs have to be aligned with current disciplinary practices and cultures. A bibliometric study by Pinto, Pulgarin and Escalona (2014) shows that information literacy both in the health sciences and the social sciences have a respective specific personality. In general, newer approaches to information literacy underline that information is used in the different disciplinary contexts (Maybee and Zilinski, 2015). In this context, the case of chemical information literacy is especially interesting. Bawden and Robinson (2015) examined its history and found that – while chemical information literacy contains some generic elements – it is more strongly domain specific than any other subject. As Farrell and Badke (2015) underline it, in order to meet the demand of the information age for skilled handlers of information, information literacy education must become situated within the socio-cultural practices of the disciplines by an expanded focus on epistemology and metanarrative. Truly situated information literacy will therefore require that librarians or disciplinary faculty invite students into disciplines. Therefore, information literacy has to be understood as information practices belonging to a discipline.

Data literacy skills are also regarded to be discipline-specific (Carlson and Johnston, 2015). As to the required skills and abilities, data literate persons have to know how to select and synthesize data and combine it with other

information sources and prior knowledge. They also have to recognize source data value, be familiar with data types and formats (Calzada Prado and Marzal, 2013). Other skills include knowing how to identify, collect, organize, analyse, summarize, and prioritize data. Developing hypotheses, identifying problems, interpreting the data, and determining, planning, implementing, as well as monitoring courses of action also pertain to required skills and add the need for tailoring data literacy to the specific uses (Mandinach and Gummer, 2013).

Ridsdale et al. (2015) set up a matrix of data literacy competencies with the intention to foster an ongoing conversation about standards of data literacy and learning outcomes in data literacy education. The perhaps most important activity in this matrix is quality evaluation that includes assessing sources of data for trustworthiness and for errors or problems. Evaluation appears already when we collect data and data interpretation clearly shows the mechanisms that also characterize information literacy. Even data visualization comprises evaluating and critically assessing graphical representations of data.

A pilot data literacy program on data literacy offered at Purdue University was built around the following skills:

- Planning;
- Lifecycle models;
- Discovery and acquisition;
- Description and metadata;
- Security and storage;
- Copyright and licensing;
- Sharing;
- Management and documentation;
- Visualizations;
- Repositories;
- Preservation;
- Publication and curation (Carlson and Stowell Bracke, 2015).

The fact that data quality plays a distinguished role in data literacy is also demonstrated by Carlson et al (2011), who compiled the perspectives of both faculty and students. Generally, faculty in this study expected their graduate students to be able to carry out data management and handling activities. Both major responsibilities and deficiencies in data management of graduate students included quality assurance. Quality assurance is seen as a blend of technical skills that materialises in familiarity with equipment; disciplinary knowledge, and a metacognitive process that requires synthesis. Even though partly superseded by the *Framework for Information Literacy for Higher Education* (ACRL, 2015), data literacy can be seen through the prism of the

Information Literacy Competency Standards for Higher Education (ACRL, 2000). Standard Three of these Standards (Evaluate information critically) contains the requirement of understanding and critically assessing sources by determining if the given data is reputable and/or if the data repository or its members provide a level of quality control for its content.

As mentioned above, managing metadata is one of the key quality-related processes of data governance. In the same time, the appraisal of metadata is part of quality assurance that should be included in data literacy programs. Quality assurance in this context comprises utilising metadata to facilitate understanding of potential problems with data (Ridsdale et al., 2015).

Data literacy education has a dual purpose. The first one is rather self-explanatory, i.e. to achieve that students, faculty and researchers become data literate science workers. As Carlson and Johnston (2015) underline it, we must raise awareness of data literacy among faculty, students and administrators by sending clear messages to our stakeholders' needs. Some of these messages could have their roots in business environments. Conveying corporate messages may even strengthen the credibility of such messages. The second goal is to educate information professionals (Qin and D'Ignazio, 2010; Schneider, 2013).

Imparting data literacy to faculty is hampered by the circumstance that educating them is a delicate issue. As Duncan, Clement and Rozum (2013) pointed it out, faculty members rarely like to hear what they are doing something in a wrong way. Exner (2014) also confirms that it is not easy to reach faculty, especially if we do not understand their lives properly. Faculty members are busy, and being experts in their fields, they usually require different approaches to instruction than students (Carlson and Johnston, 2015).

Conclusion

Although being familiar with data governance did not receive considerable attention in academia, it brings substantial knowledge to the work of the data librarian. Despite differences, between them, both data governance and data literacy are indispensable for managing data quality, thus – by their overarching nature – making use of them is a prerequisite of effective and efficient data management that substantiates research data services.

Making use of the lessons learnt from data governance could substantially enhance the effectiveness of research data management processes in academic libraries. The reasons for this are manifold. First, in delineating decision domains and defining accountability for decision making, applying practices adopted from data governance can improve data management in the library. Second, data governance is a service that is based on standardised,

repeatable processes and is designed to enable the transparency of data-related processes and cost reduction, thus it can be used also in the academic library. Third, it refers to rules, policies, standards; decision rights; accountabilities and methods of enforcement. Therefore, it would serve as a pragmatic addition to already existing data quality principles, practices and tools of the library. Fourth, the practice of data governance can also be helpful in managing change and negotiating big data issues.

These lessons can speak for themselves and may be built into data literacy programs. It is important for the library profession to take this challenge seriously and acquire the skills needed to provide effective data literacy education, irrespective of the fact that its competencies extend beyond the knowledge and skills of a typical librarian, or a faculty member. Paying attention to the management of data quality (also taking data governance into consideration) is an important step in direction of making all our target audiences to accept the library's mission to provide research data services and offer these services to their full satisfaction.

References

ACRL (2000) *Information Literacy Competency Standards for Higher Education*. Chicago, IL: Association of College and Research Libraries. Available at: <http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf> (accessed 2 May2016).

ACRL (2013) *Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment*. Chicago, IL: Association of College and Research Libraries. Available at: <http://acrl.ala.org/intersections/> (accessed 2 May2016).

ACRL (2014) ACRL Research Planning and Review Committee. Top ten trends in academic libraries. A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News* 75(6): 294–302.

ACRL (2015) *Framework for Information Literacy for Higher Education*. Chicago, IL: Association of College and Research Libraries.

ALA (1989) *Final Report, American Library Association Presidential Commission on Information Literacy*. Chicago, IL: American Library Association. Available at: <http://www.ala.org/acrl/publications/whitepapers/presidential> (accessed 2 May2016).

Andretta S, Pope A and Walton G (2008) Information Literacy Education in the UK. *Communications in Information Literacy* 2(1): 36–51.

- ANZIIL (2004) *Australian and New Zealand Information Literacy Framework*. 2nd ed., Bundy A (ed.). Adelaide: Australian and New Zealand Institute for Information Literacy.
- Badke, W. (2010) Information overload? Maybe not. *Online*, 34(5): 52–54.
- Bailey Ch (2015) *Research Data Curation Bibliography, Version 5*. Houston, TX: Digital Scholarship. Available at: <http://digital-scholarship.org/rdcb/rdcb.htm> (accessed 2 May2016).
- Bawden D and Robinson L (2015) 'An intensity around information': the changing face of chemical information literacy. *Journal of Information Science* 0165551515616919.
- Boyd D and Crawford K (2012) Critical Questions for Big Data. *Information, Communication and Society* 15(5): 662–669.
- Breaking Big (2014) *Breaking Big: When Big Data Goes Bad. The Importance of Data Quality Management in Big Data Environments*. New, York, NY: Information Builders.
- Briney, K (2015) *Data Management for Researchers. Organize, maintain and share your data for research success*. Exeter: Pelagic Publishing
- Calzada Prado JC and Marzal MÁ (2013) Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri* 63(2): 123–134.
- Carlson J, Fosmire M, Miller C and Nelson MS (2011) Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2): 629–657.
- Carlson J and Johnston LR (2015) (eds.) *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*. West Lafayette, IN: Purdue University Press.
- Carlson J and Stowell Bracke MS (2015) Planting seeds for data literacy: Lessons learned from a student-centered education program. *International Journal of Digital Curation* 10(1): 95–110.
- DGI (2015a) *Definitions of Data Governance*. Available at: http://www.datagovernance.com/adg_data_governance_definition/ (accessed 2 May2016).
- DGI (2015b) *Data governance: The basic information*. Data Governance Institute. Available at: http://www.datagovernance.com/adg_data_governance_basics/ (accessed 2 May2016).
- Dong XL and Srivastava D (2013) Big data integration. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference*. New York, NY: IEE, pp. 1245–1248.

- DosSantos J (2015) What librarians can teach us about managing Big Data. *InFocus*. Available at: https://infocus.emc.com/joe_dossantos/what-librarians-can-teach-us-about-managing-big-data (accessed 2 May2016).
- Duncan J, Clement K and Rozum B (2013) Teaching Our Faculty. Developing Copyright and Scholarly Communication Outreach Programs. In Davis-Kahl S and Hensley M K (eds.) *Common Ground at the Nexus of Information Literacy and Scholarly Communication*. Chicago IL: Association of College & Research Libraries, pp. 269–286.
- ECAR (2015) *The Compelling Case for Data Governance*. EDUCAUSE ECAR Working Group. Available at: <http://www.educause.edu/library/resources/compelling-case-data-governance> (accessed 2 May2016).
- Exner N. (2014) Research Information Literacy: Addressing Original Researchers Needs. *Journal of Academic Librarianship* 40(5): 460-466.
- Farrell R and Badke W (2015) Situating information literacy in the disciplines. *Reference Services Review* 43(2), 319-340.
- Fosmire M and Miller C (2008) Creating a culture of data integration and interoperability: librarians and Earth Science Faculty collaborate on a geoinformatics course. *Proceedings of the IATUL Conferences*, Paper 16, Available at: <http://docs.lib.purdue.edu/iatul/2008/papers/16> (accessed 2 May2016).
- Giarlo M (2013) Academic Libraries as Quality Hubs. *Journal of Librarianship and Scholarly Communication* 1(3): 1–10.
- Hartter J, Ryan SJ, MacKenzie CA, et al. (2013) Spatially Explicit Data: Stewardship and Ethical Challenges in Science. *PLoS Biology* 11(9): e1001634. doi:10.1371/journal.pbio.1001634
- Hughes H and Bruce CS (2012) Snapshots of informed learning: LIS and beyond. *Education for Information* 29(3-4): 253–269.
- Hunt K (2004) The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly* 28(2): 12–15, Available at: http://www.iassistdata.org/downloads/iqv01282_3hunt.pdf (accessed 2 May2016).
- IBM (2012) *Successful information governance through high-quality data*. Somers, NY: IBM Corporation
- IBM (2016) *What is data integration?* Available at: <http://www.ibm.com/analytics/us/en/technology/data-integration/> (accessed 2 May2016).
- Jahnke L, Asher A and Keralis S D (2012) *The problem of data*. Washington, DC: Council on Library and Information Resources.

- Johnson CA (2012) *The Information Diet: A Case for Conscious Consumption*. Sebastopol, CA: O'Reilly Media.
- Khatri V and Brown CV (2010) Designing data governance. *Communications of the ACM* 53(1): 148–152.
- Kim J (2015) Competency-based Curriculum: An Effective Approach to Digital Curation Education. *Journal of Education for Library and Information Science* 56(4): 283–297.
- Koltay T (2015b) Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science* Epub ahead of print November 26, 2015, doi: 10.1177/0961000615616450
- Kowalczyk ST (2014) Where Does All the Data Go: Quantifying the Final Disposition of Research Data. *Proceedings of the American Society for Information Science and Technology* 51(1): 1–10.
- Koltay T (2015a) Data literacy: in search of a name and identity. *Journal of Documentation* 71(2): 401–415.
- Krier L and Strasser CA (2014) *Data Management for Libraries*. Chicago, IL: American Library Association.
- Lenzerini M (2002) Data integration: A theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. New York, NY: Association of Computing Machinery, pp. 233–246.
- MacMillan D (2014) Data Sharing and Discovery: What Librarians Need to Know. *Journal of Academic Librarianship* 40(5): 541–549.
- Madrid MM (2013) A study of digital curator competences: A survey of experts. *The International Information and Library Review* 45(3-4): 149–156.
- Mandinach E and Gummer E (2013) A systemic view of implementing data literacy in educator preparation. *Educational Researcher* 42(1): 30–37.
- Martell Ch (2009) sAccess: The social dimension of a new paradigm for academic librarianship. *The Journal of Academic Librarianship* 35(3): 205–206.
- Maybee C and Zilinski L (2015) Data informed learning: a next phase data literacy framework for higher education. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. Silver Spring, MD: American Society for Information Science, 108–111.
- Nicholas D, Watkinson A, Volentine R et al. (2014) Trust and Authority in Scholarly Communications in the Light of the Digital Transition *Learned Publishing* 27(2): 121–134.
- NMC (2014) *NMC Horizon Report: 2014 Library Edition*. Austin, TX: The New Media Consortium, Available at: <http://redarchive.nmc.org/publications/2014-horizon-report-library> (accessed 2 April 2015).

- ORACLE (2015) *The Five Most Common Big Data Integration Mistakes to Avoid*. Redwood Shores, CA: Oracle Corporation.
- Pinto M, Pulgarin A and Escalona M (2014) Viewing information literacy concepts: A comparison of two branches of knowledge. *Scientometrics* 98(3): 231–232.
- Qin J and D'Ignazio J (2010) Lessons Learned from a Two-year Experience in Science Data Literacy Education. In *Proceedings of the 31st Annual IATUL Conference* Available at: <http://docs.lib.purdue.edu/iatul2010/conf/day2/5> (accessed 2 May2016).
- Ramírez ML (2011) Whose role is it anyway? A library practitioner's appraisal of the digital data deluge. *Bulletin of the American Society for Information Science and Technology* 37(5): 21–23.
- Ridsdale Ch, Rothwell J, Smit M, et al. (2015) *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. Halifax, NS: Dalhousie University. Available at: http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf (accessed 2 May2016).
- Riley, A C (2015) Data management and curation: Professional development for librarians needed. *College & Research Libraries News* 76(9): 504–506.
- RIN (2011), *The Role of Research Supervisors in Information Literacy*, Research Information Network, Available at: http://www.rin.ac.uk/system/files/attachments/Research_supervisors_report_for_screen.pdf (accessed 2 May2016).
- Rosenbaum S (2010) Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Services Research* 45(5): 1442–1455.
- Sarsfield S (2009) *The Data Governance Imperative: A Business Strategy for Corporate Data*. Ely: IT Governance Publishing.
- Schneider R (2013) Research Data Literacy. in Kurbanoglu S et al. (eds.), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*. Cham: Springer International, pp. 134-140.
- SCONUL (2011) *The SCONUL Seven Pillars of Information Literacy. Core Model for Higher Education*. London: Society of College, National and University Libraries Working Group on Information Literacy, Available at: <http://www.sconul.ac.uk/sites/default/files/documents/coremodel.pdf> Available at: <http://docs.lib.purdue.edu/iatul/2008/papers/16> (accessed 2 May2016).
- Searle S, Wolski M, Simons N, et al. (2015) Librarians as partners in research data service development at Griffith University. *Program: electronic library and information systems* 49(4): 440-460.
- Seiner RS (2014) *Non-invasive data governance: the path of least resistance and greatest success*. Basking Ridge, N.J.: Technics Publications.

- Smith AM (2007) Data Governance Best Practices – the Beginning. *EIMInsight* (1)1, Available at: <http://www.eiminstitute.org/library/eimi-archives/volume-1-issue-1-march-2007-edition/data-governance-best-practices-2013-the-beginning> (accessed 2 May2016).
- Soares S (2012) *Big data governance: an emerging imperative*. Boise, ID: MC Press.
- Tenopir C., Birch B., Allard S. (2012) *Academic Libraries and Research Data Services. Current Practices and Plans for the Future*. Chicago, IL: Association of College and Research Libraries.
- Tenopir C, Sanduski R.J, Allard S, et al. (2014) Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research* 36(2): 84–90.
- Tenopir C, Hughes D, Allard S, et al. (2015) Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship* 4(2): e1085. <http://dx.doi.org/10.7191/jeslib.2015.1085>
- Weber NM, Palmer CL and Chao TC (2012). Current Trends and Future Directions in Data Curation Research and Education. *Journal of Web Librarianship* 6(4): 305-320.
- Vitae (2011) *Researcher Development Framework*. Cambridge: Careers Research and Advisory Centre.
- Weill P and Ross JW (2004) *IT governance: How top performers manage IT decision rights for superior results*. Boston, MA: Harvard Business School Press.
- Zilinski LD and Nelson MS (2014) Thinking critically about data consumption: Creating the data credibility checklist. *Proceedings of the American Society for Information Science and Technology* 51(1): 1–4.