Volume 42 Number 4 December 2016

Contents

Special Issue: Research Data Services

Guest Editors: Michael Witt and Wolfram Horstmann

Guest Editorial

Articles

Aims and Scope

IFLA Journal is an international journal publishing peer reviewed articles on library and information services and the social, political and economic issues that impact access to information through libraries. The Journal publishes research, case studies and essays that reflect the broad spectrum of the profession internationally. To submit an article to IFLA Journal please visit: http://ifl.sagepub.com

IFLA Journal
Official Journal of the International Federation of Library Associations and Institutions
ISSN 0340-0352 [print] 1745-2651 [online]

Published 4 times a year in March, June, October and December

Publisher

SAGE, Los Angeles, London, New Delhi, Singapore, Washington DC and Melbourne.

Annual subscription (4 issues, 2017) Free to IFLA members. Non-members: full rate (includes electronic version) £304/$561. Prices include postage. Full rate subscriptions include the right for members of the subscribing institution to access the electronic content of the journal at no extra charge from SAGE. The content can be accessed online through a number of electronic journal intermediaries, who may charge for access. Free e-mail alerts of contents listings are also available. For full details visit the SAGE website: sagepublishing.com

Student discounts, single issue rates and advertising details are available from SAGE, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP, UK. Tel: +44 (0) 20 7324 8500; e-mail: subscriptions@sagepub.co.uk; website: sagepublishing.com. In North America from SAGE Publications, PO Box 5096, Thousand Oaks, CA 91359, USA.

Please visit ifl.sagepub.com and click on More about this journal, then Abstracting/indexing, to view a full list of databases in which this journal is indexed.

Printed by Henry Ling Ltd, Dorset, Dorchester, UK.

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 251–252

DOI: 10.1177/0340035216678726
ifl.sagepub.com

International approaches to research data services in libraries

Michael Witt
Purdue University, West Lafayette, Indiana, USA

Wolfram Horstmann
Göttingen State and University Library, University of Göttingen, Germany

Libraries and archives around the world are acquiring new skills and applying the principles of library and archival sciences to solve challenges and provide new services related to research data management. Librarians are helping researchers address needs throughout the research data lifecycle, for example, by conducting assessments and outreach, consulting on data management plans and metadata, incorporating data into information literacy instruction and collection management, providing reference services to help patrons find and cite data, and providing data publication and preservation solutions. They are creating web guides and tutorials, training colleagues within and outside of the library, contributing to discussions related to research data including policy development and planning, and in some cases, participating directly with researchers on data-intensive projects.

The degree to which libraries are offering or planning to offer data services has been explored in detail for member organizations of the Association of College and Research Libraries in North America by Tenopir et al. (2012, 2015) and more recently for LIBER (Ligue des Bibliothèques Européennes de Recherche) in Europe (Tenopir et al., 2016); however, these studies only begin to paint a part of the picture in terms of the variety of services that libraries around the world are designing and deploying. And how are they creating and providing these services?

This is the central motivation for the next two special issues of *IFLA Journal*: to gather the latest theory, research, and state-of-the-art practices from libraries that are informing and innovating effective data services from an international perspective. The idea for this theme was inspired by the high level of attendance and interest in a conference program on the topic of research data and libraries that was created in collaboration with IFLA and the Research Data Alliance for the 81st IFLA World Library and Information Congress in Cape Town in 2015. The number and quality of manuscripts submitted in response to the call for papers enjoined the expansion from one special issue to two and the allowance of some additional time for reviewing and editing.

Major themes that emerged from the submissions include the assessment of researcher needs and practices, training for librarians, examples of different data services and approaches to designing and offering them, and data information literacy. The papers are presented in this order as a sensible progression that many libraries are undertaking themselves: to identify patron needs related to research data, then learn the skills required to help meet their needs, then design and offer services, and lastly assist patrons in using the services and related resources.

In this issue, researchers were interviewed at Griffith University in Australia with the application of the A-COM-B conceptual framework to better understand researcher behaviors regarding research data and related practices. Their results suggest that attitude is the key element to be addressed in designing strategies to support researchers in data management. Librarians at two Swedish universities extended the Data Curation Profile instrument and used it in interviewing researchers from a variety of different disciplines to explore their needs for effective data management, in particular for subject control and other requirements for descriptive metadata. In the Netherlands, the 'Essentials 4 Data Support' taught

Corresponding author:
Michael Witt, Purdue University, West Lafayette, Indiana, IN 47907, USA.
Email: mwitt@purdue.edu

data support skills to over 170 Dutch librarians and information technology professionals. Learning outcomes are achieved by either a six-week blended course of in-person and coached online course, or an online course that is accompanied by coaching, or a self-directed online-only course. Two papers present examples of data services: ETH-Bibliothek in Switzerland takes a data lifecycle approach that incorporates researchers working in private, shared, and public domains with corresponding levels of access control and other functionalities; the Qualitative Data Repository hosted by Syracuse University in the United States is tailored to meet the needs of qualitative and multi-method social inquiry research methods with emphases on protecting data that involve human subjects and providing the capability to relate data to published text through scholarly annotation. A literature review and discussion conducted in Hungary explores and brings attention to the relationship between data governance and data literacy. Finally, a group of American librarians examine why data information literacy should be integrated in the areas of business and public health and discuss how it can be accomplished.

A second special issue of *IFLA Journal* on research data services in libraries, forthcoming in March 2017, will continue and build upon these four themes while incorporating approaches and perspectives from a broader range of countries and libraries. It is our hope that these two issues will add to our shared understanding of the latest developments and practices of libraries that are designing and providing research data services.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Tenopir C, Birch B, and Allard S (2012) Academic libraries and research data services: Current practices and plans for the future. Association of College and Research Libraries.

Tenopir C, et al. (2015) Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship*, 4(2).

Tenopir C, et al. (2016) Research Data Services in European Academic Research Libraries (Submitted). Pre-available at: http://libereurope.eu/blog/2016/10/13/research-data-services-europes-academic-research-libraries (accessed 17 October 2016).

Article

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 253–265

DOI: 10.1177/0340035216673856
ifl.sagepub.com

Modifying researchers' data management practices: A behavioural framework for library practitioners

Susan Hickson
Griffith University, Gold Coast Campus, Australia

Kylie Ann Poulton
Griffith University, Nathan Campus, Australia

Maria Connor
Griffith University, Gold Coast Campus, Australia

Joanna Richardson
Griffith University, Nathan Campus, Australia

Malcolm Wolski
Griffith University, Nathan Campus, Australia

Abstract

Data is the new buzzword in academic libraries, as policy increasingly mandates that data must be open and accessible, funders require formal data management plans, and institutions are implementing guidelines around best practice. Given concerns about the current data management practices of researchers, this paper reports on the initial findings from a project being undertaken at Griffith University to apply a conceptual (A-COM-B) framework to understanding researchers' behaviour. The objective of the project is to encourage the use of institutionally endorsed solutions for research data management. Based on interviews conducted by a team of librarians in a small, social science research centre, preliminary results indicate that attitude is the key element which will need to be addressed in designing intervention strategies to modify behaviour. The paper concludes with a discussion of the next stages in the project, which involve further data collection and analysis, the implementation of targeted strategies, and a follow-up activity to assess the extent of modifications to current undesirable practices.

Keywords

Attitude, behaviour, behavioural framework, capability, libraries, motivation, opportunity, research data management

Submitted: 13 May 2016; Accepted: 7 September 2016.

Introduction

In recent years, advances in data management, data curation, dissemination, sharing and research infrastructure have transformed the provision of library services to researchers. A considerable amount of literature has been published on the emergence in academic libraries of research data services for their faculty and students (Peters and Dryden, 2011; Si et al., 2015; Wang and Fong, 2015; Weller and Monroe- Gulick, 2014), although many more institutions do not yet provide targeted outreach programmes (Tenopir et al., 2013).

Corresponding author:
Susan Hickson, Griffith University, Gold Coast Campus, University Drive, Southport, Queensland 4222, Australia.
Email: s.hickson@griffith.edu.au

This paper reports on the progress of a broader research project that investigates whether the actions of library and information specialists to improve data management practices can be enhanced by understanding the attitudes and behaviour of researchers. Wolski and Richardson (2015: 1) have proposed a behavioural framework (A-COM-B) for service delivery teams to 'a) better understand the cohort with which they are engaging, b) identify where and when to focus their attention, and c) develop more effective plans to bring about changed researcher practices'. The objective of the research is to use the A-COM-B model as a lens for improving understanding of data management practices in order to develop, plan and provide interventions to change researchers' data management practices and to investigate the results of such interventions. This paper reports on progress to date.

Environmental research context

International

The international research environment promotes the sharing of information awarded by publicly funded research and encourages the reuse of data sets for the betterment of the community by ensuring that practice reflects policy and guidelines. For example, in the Research Councils UK's (RCUK) (2014a) Research Outcomes Overview, it is stipulated that researchers must 'demonstrate the value and impact of research supported through public funding' and 'Information on research outcomes attributed to all RCUK funded awards are now collected online'. In addition, the RCUK policy (Research Council UK, 2015) mandates that research data be open and discoverable and that policies and procedures should be in place to ensure compliance with these practices. RCUK actively works with *researchfish* (Research Councils UK, 2014b), which is a portal where researchers can log the outputs, outcomes and impacts of their work.

Likewise, the United States' National Science Foundation (NSF) is 'dedicated to the support of fundamental research and education in all scientific and engineering disciplines'(National Science Foundation, 2016a). The NSF's webpage on the dissemination and sharing of data states that all applicants must submit a data management plan and describe how it meets the policy on the dissemination and sharing of research results (National Science Foundation, 2016b).

Similarly the European Commission (EC) has instituted an Open Research Data Pilot as part of Horizon 2020 (European Commission, 2016). All research proposals must include a section on research data management which clearly outlines how the data will be disseminated, shared and retained. This informs the Digital Single Market, which aims to open up digital opportunities for citizens' access to information and promote modern open government.

Australia

Similar to the United Kingdom and the United States, the Australian Research Council (ARC) (2014) considers data management planning an important part of the responsible conduct of research. The ARC strongly encourages the depositing of data arising from a research project in an appropriate, publicly accessible subject and/or institutional repository. Additionally, the National Health and Medical Research Council's (2014) *Policy on the Dissemination of Research Findings* states that the metadata from all NHMRC projects should also be made available in an open source journal within 12 months of publication.

The Australian Code for the Responsible Conduct of Research (National Health and Medical Research Council, 2007) promotes integrity in the proper management of research data, publishing, dissemination and attribution of authorship by use of a best practice framework. Researchers and institutions each have a responsibility to comply with legislation, guidelines, policy and codes of practice as part of responsible research.

Griffith University

Griffith University is a comprehensive, research-intensive university, ranking 37th in the *2015/16 QS University Rankings Top 50 Under 50* (Quacquarelli Symonds, 2015). Located in the rapidly growing corridor between Brisbane and the Gold Coast in Southeast Queensland, the University offers more than 200 degrees across five campuses to more than 43,000 students from 130 countries studying at undergraduate through to doctoral level in one of four broad academic groups: arts, education and law; business; science; and health. Griffith's strategic research investment strategy has positioned it to be a world leader in the fields of Asian politics, trade and development; climate change adaptation; criminology; drug discovery and infectious disease; health; sustainable tourism; water science; music and the creative arts.

The Division of Information Services (INS), with which the authors are affiliated, has a long and proud tradition of providing quality service to Griffith students and staff. It also has an international reputation for being innovative and cutting-edge in the

deployment of emerging technologies (Brown et al., 2015; Searle et al., 2015; Stanford Libraries, 2013).

In 2014 Griffith University introduced its *Best Practice Guidelines for Researchers: Managing Research Data and Primary Materials*, in response to the Australian Code for the Responsible Conduct of Research (National Health and Medical Research Council, 2007). The subsequent *Griffith University Code for the Responsible Conduct of Research* was endorsed by the University's Academic Committee in July 2012 and subsequently updated in November 2015. According to the principles of the Code, the University is required to provide research infrastructure, training and professional development opportunities as well as elicit an understanding from researchers that they are to properly manage their data, while adhering to all applicable policies, legislation and standards with honesty and integrity.

Griffith University has invested significantly in the provision of solutions and services for capturing, storing, analysing, managing, sharing and publishing data. Examples of institutional storage solutions include (a) Research Storage Service (https://research-storage.griffith.edu.au/), a self-serve private cloud solution based on the open source application, ownCloud, and (b) a research data repository. Since implementing the ownCloud solution in 2015, as of May 2016 there were 580 Griffith users. Although a pleasing start, this constitutes only approximately 15% of potential users at the institution. Given the status of Griffith as a research intensive university, a minimum of 30% take-up by May 2016 seemed a reasonable expectation. However, it became apparent that there is a lack of incentives for researchers to use enterprise services. The research project described in this paper is but one response to bring these services to the attention of the academic community.

Griffith University is currently developing an institutional approach to research data support by utilising the Australian National Data Service (ANDS) Data Management Framework (Australian National Data Service, 2016b) and the Capability Maturity Model Integration (CMMI) (Australian National Data Service, 2016a). Key stakeholders of the University, i.e. Griffith Graduate Research School, Office for Research, and Information Services, are seeking a common understanding for the strategic direction of data management, education, training, services and solutions in order to develop a data support model and an institutional data service.

The research project described in this study, along with others currently being undertaken, will inform the process of developing a suitable framework.

Literature review

In the context of this paper, *data management* follows the definition by O'Reilly et al. (2012: 2): 'all aspects of creating, housing, delivering, maintaining, and retiring data'. Data management is a part of the everyday research process and work habits rather than a specific task undertaken separately within the research lifecycle. 'Data management is not necessarily a formalised process, but rather actions taken in response to a researcher's current information needs and work goals' (Fear, 2011: 71).

The literature exploring data management practices of academics has significantly increased in the last decade, with many universities surveying the data management practices of their academics and researchers (Fear, 2011; Henty et al., 2008; Jahnke et al., 2012; Kennan and Markauskaite, 2015; Peters and Dryden, 2011; Schumacher and VandeCreek, 2015; Weller and Monroe-Gulick, 2014). As the amount of research data being generated grows exponentially, universities are taking notice of how their researchers manage research data. Historically the creator was solely responsible for their own data; however, as new policies, both at a national and publisher level, now require research data to be made publicly available, institutions are required to provide adequate infrastructure which allows researchers to safely access, store and archive their data.

That said, institutional data storage services are infrequently used as they do not adequately meet researcher needs. O'Reilly et al. (2012: 13) report that: 'Data storage is often inadequate, with researchers resorting to suboptimal data storage methods, resulting in data that are often unreliable and short lived'. Westra (2014) has reported on the need to address this issue as part of the University of Oregon's decision to develop an integrated research data management strategy. Jahnke (2012) urges institutions to make networked storage more readily available to 'multi-institutional research projects' (p. 17), integrate 'the data preservation system with the active research cycle' (p. 19), and enhance storage systems with 'intuitive live linking visualization tools' which could entice researchers to use the systems and assist with 'curatorial decision making'.

Researchers are 'largely unaware of the basic principles of . . . data management' (Schumacher and VandeCreek, 2015: 107) and receive little to no training other than what they learn through doing research (Fear, 2011: 64; Peters and Dryden, 2011: 395). As a result, researchers are largely left to create their own unique, ad hoc approach to organising their research data. Furthermore, data management

reflects individuals' organisational style, which is not always easy for others to interpret (Fear, 2011: 64). This can cause problems for those trying to re-use data later on.

Not surprisingly, researchers consistently and primarily manage their research data on their PCs, a USB device and/or CDROM, closely followed by cloud-based services such as Google Drive and Dropbox (Schumacher and VandeCreek, 2015; Weller and Monroe-Gulick, 2014; Wolff et al., 2016). All of these solutions pose potential privacy and security challenges, particularly the freely available alternatives to institutional data storage, e.g. Dropbox, Figshare, and SurveyMonkey. Jahnke (2012: 12) elaborates:

> . . . the terms of service for these products are often poorly understood by researchers and the research participants. Furthermore, the terms of service may not be sufficient to meet the data protection and confidentiality standards that researchers and their institutional review boards (IRBs) require. Dropbox's well publicized June 2011 security glitch, which left all Dropbox accounts open to access without a password for several hours, is indicative of this problem. Applying additional security measures, such as encrypting files locally prior to sharing them via a cloud service, is beyond the technical skills of many researchers . . . .

The data management practices of STEM researchers are heavily represented in the literature (Fear, 2011; Peters and Dryden, 2011). A number of studies have surveyed academics from across all disciplines within their institutions (Schumacher and VandeCreek, 2015; Weller and Monroe-Gulick, 2014), and social science and humanities researchers are often discussed within these broader studies. Jahnke et al (2012) solely investigate the data management practices of social science researchers. Their study explores the 'nonlinear nature of the research process' (Jahnke et al., 2012: 10) and how this impacts data management practices, along with concerns around data sharing, infrastructure and technical issues.

Martinez-Uribe and Macdonald (2009: 314) found that 'the curation of research data requires trusted relationships achieved by working and conversing with researchers . . . '. Jahnke et al. (2012: 19) recommend that 'extensive outreach to scholars is necessary to build the relationships that will facilitate data preservation'.

This paper adds to the literature by applying a newly developed conceptual framework to the attitudes and behaviours of researchers in regard to data management practices.

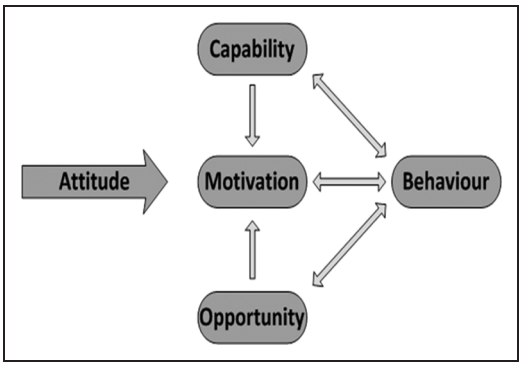

Figure 1. A-COM-B framework for understanding behaviour.

Methodology

The present study has used the descriptive research method, based on surveying selected participants through interviews. The basis for the survey design was the behavioural framework described below.

A-COM-B Framework

In investigating interventions which could be used to improve researchers' data management practices, Wolski and Richardson (2015) examined a number of behaviour and behaviour change theories and models. Ultimately they chose the COM-B system developed by Michie et al. (2011) as the simplest, yet most comprehensive framework on which to base their approach. In the COM-B system, C = Capability; O = Opportunity; and M = Motivation, all of which interact to generate behaviour (B).

However, influenced by the work of Piderit (2000), Wolski and Richardson modified the COM-B system by incorporating attitude as a key component in understanding researcher behaviour. Figure 1 offers a diagrammatic representation of this concept, now presented as A-COM-B. The single-headed and double-headed arrows represent the potential for influence between the various elements.

Attitude is defined as 'an individual's evaluation or belief about something' (Wolski and Richardson, 2015: 6); capability is the 'the psychological or physical ability to enact the behaviour' (Michie et al., 2011: 4). Motivation is defined as 'all those brain processes that energize and direct behaviour, not just goals and conscious decision-making. It includes habitual processes, emotional responding, as well as analytical decision-making' (Michie et al., 2011: 4). Opportunity is defined as 'all the factors that lie outside the individual that make the behaviour possible or prompt it' (Michie et al., 2011: 4). Behaviour is the result of the interaction between these four key elements.

In applying the framework to a situation in which one wishes to modify a current practice, the starting point is to (1) identify the underlying elemental behaviours that make up the practice and (2) identify which of these needs to be changed. Importantly the next step is to:

> . . . identify current attitudes to the desired change in behaviour. However a challenge is that unlike behaviour, attitudes are more difficult to observe, measure and quantify. Therefore attention may need to be paid to employing techniques such as qualitative interviewing coupled with good listening skills. This is an important step as understanding the nature of attitudes will normally provide insights into the other elements of the framework, i.e. capability, motivation and opportunity. An understanding of all these elements creates a foundation for developing an intervention plan. (Wolski and Richardson, 2015: 8)

In developing an intervention plan with specific reference to data management, Wolski and Richardson strongly advocated the need for research support teams to understand individual behaviours within the context of their local cohort level rather than at the larger faculty or institutional level. They suggested that librarians could trial the strategy with a research centre. In addition, any intervention should be 'a multi-pronged approach which targets the different elements of the framework' (Wolski and Richardson, 2015: 8).

Applying the framework

The research team, comprising of three of the authors, selected a small but high-profile social sciences research centre to test the theories in Wolski and Richardson's A-COM-B model. This centre was chosen because of the existing relationship which the authors had already built with its researchers and because this choice would align with Wolski and Richardson's (2015: 8) theory that 'to understand the current attitudes and to plan an 'intervention' plan, local service delivery teams may need to understand their local cohort to develop an effective response'.

To identify underlying attitudes and behaviour, the research project team designed a series of interview questions, categorised according to five broad headings:

- How and where researchers store their data;
- Criteria for selecting the solutions they use;
- Backup methods and approaches;
- Data management planning; and
- Additional aspects of their research habits.

Table 1. Distribution of invited participants by career stage.

Early – mid career (Within 15 years of being awarded PhD)	Late career (15+ years since awarded PhD)	Total
10	14	24

There are 23 questions in the survey, comprised of a mixture of multiple-choice and free-text formats.

The interviews were conversationally led by one research project team member, while a second team member noted responses to the questions in a Google form. Each interview was allocated 30 minutes. The interviews were also recorded and later transcribed.

It is estimated that the research project will take 12 months to complete. The initial phase of the project as described in this paper was carried out over a six-month period. It is expected that the next phase of the project will include devising and implementing intervention strategies, follow-up surveys and analysis, and the creation of a toolkit for service providers. It is anticipated that these activities will take a further six months.

Demographics of the research centre

The centre chosen for the study is a high-profile, interdisciplinary social science research group. It comprises researchers from several departments across the University, whose research focuses on human resource management, industrial relations and organisational behaviour. A recent benchmarking exercise, conducted by librarians in Library and Learning Services, placed the research centre among the top national and international research centres with a similar research focus.

For the purpose of the study, adjunct members and those on extended leave were excluded. The number of researchers invited to participate was 24.

As shown in Table 1, there is a fairly even distribution among early/mid-career and late career researchers.

Table 2 indicates that 20 (83%) of the participants have an academic rank of senior lecturer or above.

At the time of writing this paper, 12 (50%) of the invited participants had been interviewed.

Initial findings

Types of data

Researchers were asked to describe their research areas and the types of data they collect. Most researchers reported collecting both qualitative and

Table 2. Distribution of invited participants by academic rank.

Research Assistant	Research Fellow	Lecturer	Senior Lecturer	Associate Professor	Professor	Total
1	1	2	6	6	8	24

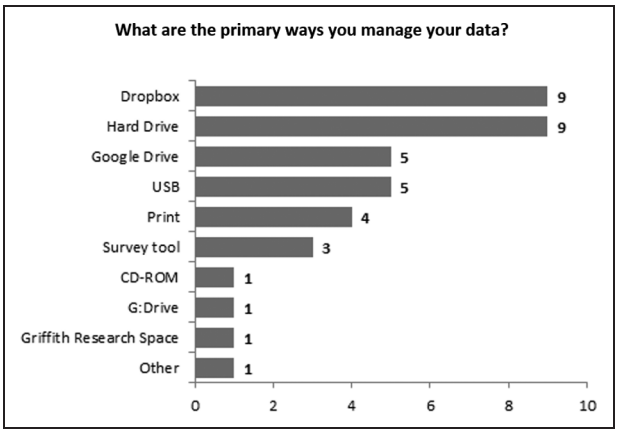

Figure 2. Solutions for managing research data.

quantitative data, with only two respondents (16.7%) reporting that they collected qualitative data only. The most common method of data collection was audio recording interviews and transcribing to text-based format. Only two respondents (16.7%) used online surveys to collect data.

Managing data

Dropbox and hard drives were the most common ways of managing research data. Of the nine researchers (75% of total participants) who responded that they used Dropbox, five used the free Dropbox product and four used the paid Dropbox product.

All 12 of the researchers reported using more than one solution for managing their data, with most using Dropbox as well as their hard drive, Google Drive and a USB device. Only one participant (8.3%) used one of the University's data storage services, i.e. Griffith Research Space (Figure 2). However, most researchers reported, in a separate section of the survey, that the size of their research data was less than 5GB, with only two (16.7%) reporting that it was more than 5GB.

Participants were asked to indicate the criteria which influenced them to choose the solutions identified in Figure 2 for managing their data. See Figure 3.

The most common reason given for their choice in data management options was ease of use. One researcher described Dropbox as 'ridiculously easy', and another cited difficulties using Google Drive that do not arise when using Dropbox.

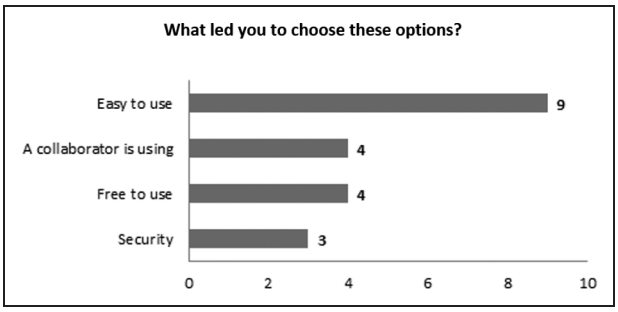

Figure 3. Criteria for selecting data management solutions.

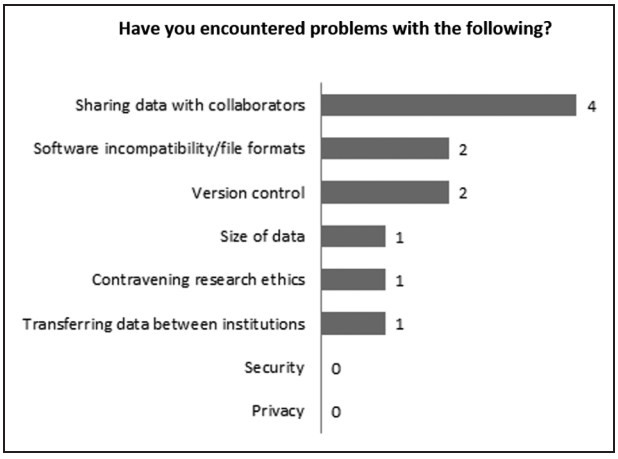

Figure 4. Issues encountered with data management.

Sharing data with collaborators was the most commonly reported problem with the way researchers manage their data (Figure 4). Some of the reported issues included complications arising from collaborators working between the free and the paid versions of Dropbox, and working with collaborators who used different systems. Size of data being shared was an issue for only one researcher (8.3%).

Participants were asked to identify all methods used for backing up their data (if applicable) (Figure 5). Using an external hard drive was the most popular way to back up research data, with 5 participants (41.6%) selecting this solution. Two researchers who considered Dropbox to be their data backup solution explained that they relied on Dropbox to do the backup automatically, and one researcher who used Google Drive 'assumed that going to the inconvenience of using Google Drive [sic], it will be backed up'.

The one researcher (8.3%) who indicated that they did not back up their data cited 'time constraints,

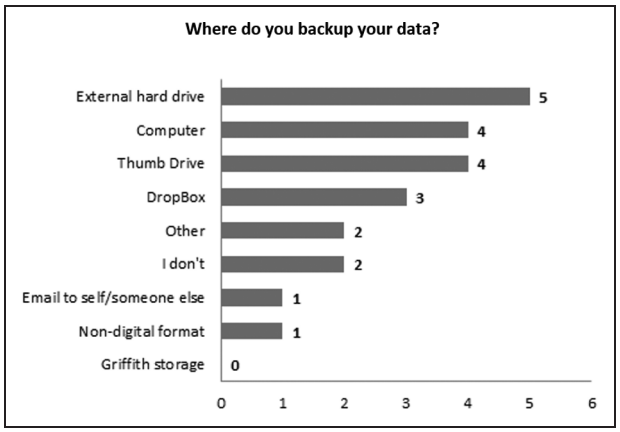

Figure 5. Solutions for backing up data.

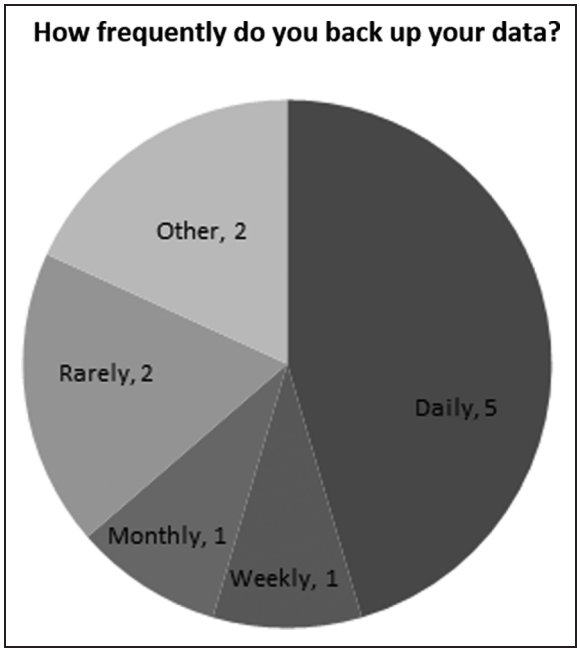

Figure 6. Frequency of data backups.

laziness and naive trust in the systems along with optimism that all is good until the first time something goes wrong' as the rationale.

Figure 6 shows that one-third of the participants (5) back up their data as frequently as on a daily basis. However, this result should be understood in the context that most of the researchers who reported backing up their data daily, were relying on it being backed up systematically by the storage products they used.

Collaboration

Most researchers reported using Dropbox to share data with collaborators (see Figure 7). One researcher explained that it was the chief investigator's responsibility to ensure that the data is stored and shared appropriately. Another researcher who spoke about

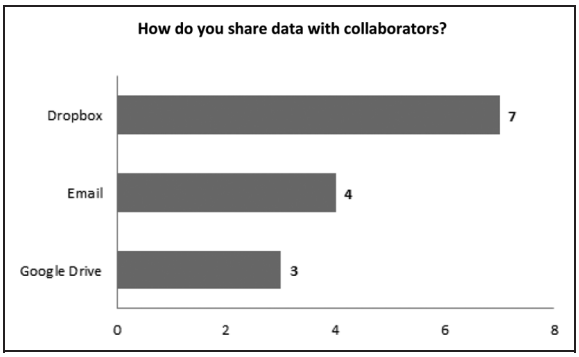

Figure 7. Methods for sharing data with collaborators.

sharing passwords and login information with collaborators to access confidential material said: 'it's get this done now, there isn't time to work through the better way to do things'.

In response to a question as to who owns research data in collaborative research projects, the responses from researchers suggested that although they had not previously thought about data ownership, they had clear perceptions on where ownership should lie. Some of the responses included that it was 'jointly owned' for collaboratively collected data and 'assume that it's mine'. Others believed it was 'wherever the grant sits' or that the University owned it.

Data management plans

When asked if they had a data management plan, all of the researchers interviewed responded that they did not. One researcher referred to 'an informal, unwritten one' and another explained, 'people just decide what to do as they go along'. Another researcher said it was 'instinctive'.

Data sharing

When asked whether they share or intend to share data publicly, there was an overwhelmingly negative response from the researchers. Two researchers recounted hearing stories of data being put online, and others subsequently re-using the data to write papers. This practice was perceived as contentious. One researcher described the UK requirement to publish data from publicly funded research projects as 'controversial' and another researcher responded, 'I would be horrified being forced to do that'. Others suggested that they could not see that their data was useful to anybody else. Only one researcher reported that they had thought about making their data open and that they did not perceive any problems in doing that.

An interesting finding, which corroborates surveys reported in the literature (Federer et al., 2015; Tenopir et al., 2011), was that a number of these same

researchers reported re-using data from research projects or requesting data from researchers with whom they had worked previously.

Griffith research storage services

While most researchers answered that they were aware of Griffith's research storage solution, only one researcher said they used it. The most common reason for not using the service was that they did not know how. When researchers were asked what would motivate them to use the storage solution, several referred to ease of use and accessibility as motivators. One researcher stated that they might start using the service 'if it's the right thing to do' and added 'it's just another thing, it may be well and good but you have to learn it and it contributes to a sense of overload'. Another researcher commented: 'if I had time to do it. I feel I am competent to do it, I just don't have the time available'.

Discussion

Wolski and Richardson's (2015) A-COM-B model was used to analyse the responses from a small group of targeted researchers to test whether researchers' data management behaviour can be understood in the context of attitude, capability, motivation and opportunity. The initial findings have been analysed to ascertain the attitude of the researchers, how it is reflected in behaviour, and whether the model can be used to plan intervention. Wolski and Richardson (2015: 6) assert that 'attitude is an individual's evaluation or belief about something' and 'When planning to initiate a plan to change behaviours of staff, consideration should first be given to the attitudes of staff and having an appreciation of their views and attitudes in relation to the change being considered'. Wolski and Richardson (2015: 8) also describe attitude as 'difficult to observe, measure and quantify'. By analysing the open-ended responses obtained as part of the initial stage of the research project, a picture of the researchers' attitude towards data management and how this influences their behaviour has become clearer.

Wolski and Richardson (2015: 6) also state that 'Understanding the nature of the attitude (more often than not ambivalence in relation to how researchers regard data management) should provide insights into the most appropriate responses that will garner the desired attitudinal change'. The authors will undertake subsequent research to further test the applicability of the A-COM-B model, by designing and implementing intervention strategies that address capability, opportunity and motivation factors in order to influence researcher attitudes and analyse whether this translates into behavioural change.

Managing data

From the interviews to date, it is apparent that researchers' attitude is to choose the 'easiest-to-use' products for storage and moving data. Some researchers reported frustrations with using varied and differing systems and having to make choices regarding the use of them. As one respondent said: 'I might use Google Drive because I know how to use it . . . and to go and use R drive and Research Storage Service and the Vault, I have to go and find out about it'. Another stated: '[there] isn't time to work through the better way to do things'. This has translated into a behaviour of using their preferred solution, mainly Dropbox and a hard drive.

By using the A-COM-B model to design intervention strategies, the research project team could create intervention plans that will target researchers' attitude towards managing data and address capability, motivation and opportunity factors.

If it can be shown that Griffith's preferred data management solution is as easy to use as Dropbox, the research project team could target motivators and opportunities to trigger a change in behaviour. However, if Griffith's preferred solution is not as easy to use as Dropbox, then rather than focus on motivation or capability, intervention strategies could target instead the researchers' attitude that 'easy-to-use' is the best reason to choose a product.

Data backup

While only one researcher reported not backing up their data, most used an external hard drive and relied on automated backup utilities. While backing up data is good data management practice, using an external hard drive may expose the data to security risk through theft or damage.

The researchers' responses demonstrate a casual and indifferent attitude towards data backup. While a few consciously and regularly backed up their data, many left it to the 'system' to do. As nothing had gone wrong in the past, they did not perceive any threat and therefore were not motivated to change their behaviour. One researcher, who regularly and deliberately backed up their data, explained that, although they had not suffered any loss of data, it was witnessing the consequences when close colleagues lost data that had made them act carefully with data backup. This is an example of a motivator that has triggered a behavioural change.

By applying the A-COM-B model to intervention strategies, the research project team could target the underlying attitudes surrounding data backup to reframe the task as being important enough to take considered, planned and careful steps to ensure data is backed up securely. Citing the example above might even be a useful intervention strategy, i.e. communicating the impact of actual cases of lost data. As the well-known axiom says, 'Facts tell, stories sell'.

Collaboration

When asked what led them to choose their data management options, 25% of the researchers said that it was because a collaborator was using that option. Many explained that either whoever was in charge or the chief investigator on a research project normally decided on where and how data was stored and managed. The researchers' attitude towards collaborative data management was that assumptions and unspoken understandings were sufficient and not necessarily in need of more accurate and specific definition.

Although many researchers believed that the chief investigator took the lead in data management, one senior researcher explained that they left it up to their research assistant to establish the data management systems of choice. They explained further that the reason Dropbox was used was because the research assistant had made that decision.

In targeting the attitude of researchers regarding collaborative data management, the research project team could design interventions to reframe researchers' attitudes towards becoming more careful and considered when planning collaborative projects. Interventions targeted at senior levels might focus on establishing clear and uniform collaborative data management guidelines, whereas interventions targeted at more junior levels might focus on the selection of data management solutions.

Data management plans

Another finding was that many of the researchers considered that there was a direct correlation between the size of a dataset and the importance of managing that dataset. Their attitude was that their datasets were not large enough or important enough to warrant more conscientious planning. One researcher, who said they did not see themselves as a data manager, suggested that anyone who made the effort to manage their data was a 'scientist with oodles of scientific data'. Another agreed that 'we need to get it right', but 'it may not be seen as important'. With only relatively small datasets, the centre's researchers may not have felt motivated to change their behaviour and conduct more robust data management planning.

When planning interventions, the research project team could target the researchers' attitude that their data is unimportant or too small to manage effectively. One motivator could be ensuring data planning requirements from grant funders are addressed; providing data management plan templates and support could help to address issues around capability and opportunity.

Data sharing

The attitude of most of the interviewed researchers towards data sharing was that it was contentious, that their data would not be of interest to other researchers, and that sharing data might raise some methodological issues. As one researcher stated:

> The data you collect is so project specific and so raises some general methodological questions you have to ask yourself about the merits or otherwise of using data collected for one purpose, or for one question, or one set of hypothesis, to address another one.

And another responded: 'I don't think it would be of value to other people. They don't have the conceptual framework to make any sense of them'.

Although most of the researchers could not perceive the benefits of sharing their data with others, many had themselves benefited from re-using data from historical research projects.

An intervention strategy might leverage the attitude that re-using data from historical research projects or data previously collected by collaborators was acceptable practice. This could be reframed as data sharing can be beneficial. Motivators might include the citation advantage of data publication and the publication and re-use of data as a measure of research impact. Capability and opportunity might be enabled through support and education around platforms and systems used to publish and share data.

Griffith University's research storage services

Although most of the researchers were aware of Griffith's research storage solutions, only one said they had used it. The researcher who reported using it expressed frustration with the service: 'I think I just found it difficult to use and I tried to get some assistance and I gave up . . . '

The attitude of the researchers' towards Griffith's research storage services reflects their attitude toward choosing data management solutions. Their current

practices work effectively, they are easy to use, and as a result the researchers feel unmotivated to change.

Many researchers cited time constraints as a reason for not changing their data management behaviour. With increasing pressure on researchers to increase their research outputs, demonstrate impact and adhere to local and national policies and mandates, data management planning can be seen as an additional burden.

Interestingly, although Wolski and Richardson (2015) suggest that the mandated policies and guidelines will not guarantee improvement to data management practices, several researchers said that they would be motivated by 'mandates'. One said 'being told to do it' would motivate them to change their data management behaviour and another said: 'if it was the right thing to do'. One researcher, who reluctantly used the University's Google Drive for research data storage, described attending a seminar where they were told they must use Google Drive and said that 'put the fear of god into me about using Dropbox'. They continued, 'Colleagues didn't attend that seminar and they all use Dropbox because it's easier and I agree with them, it's much easier'.

Implications for service providers

As service providers grapple with the challenges of getting researchers to manage their data effectively, especially as the imperative from national and local drivers increases, many studies have focused on the behaviours of researchers towards data management. This research project aims to provide an insight into researcher attitudes and whether attitude change is the key to changing behaviour (World Bank, 2010). By testing the A-COM-B framework for understanding data management behaviour, the authors hope to provide an approach for service providers wanting to design intervention strategies to change researchers' data management behaviour.

From the research undertaken so far, there have been two key findings: (1) the size of the datasets produced by a social science research centre of this type is quite small, which has implications for data management and storage; and (2) attitude is the biggest challenge when seeking to modify the target group's data management practices. As a consequence, the research project team will investigate more thoroughly attitude change strategies, as identified in the literature.

The research team has been influenced by the many insights into researchers' attitudes and behaviours. By analysing behaviours through the A-COM-B lens, the research project team now has a unique perspective on how to approach the challenge of changing data management behaviour. The A-COM-B model has proved a powerful tool with which to analyse researcher attitude and behaviour. Using the framework has allowed the research team to map how they can tailor and directly address the interactions of attitude, capability, opportunity and motivation in order to influence behavioural change.

As the team embarks on the second phase of the research project, it is worth noting that the experience with the initial data collection and analysis has led to a refinement in conversational interviewing and data collection techniques as the research project team members attempt to elicit nuanced responses from researchers in order to obtain a more detailed understanding of their attitudes.

The research project team has benefited not only from a wider understanding of data management issues, but also a broader understanding of how the framework can be applied to changing behaviours in other areas. The research project has additionally influenced the way the team approaches researchers and academics in the wider university environment about data management issues.

Next steps and challenges

This paper reports on the initial stages of the project and tests whether the A-COM-B model can be applied to data management behaviours. During the initial research phase, a tailored response was created for each of the researchers interviewed thus far. By selecting appropriate interventions from a suite of tools and products, the research team was able to deliver a response to the researchers quickly. In the next stage, the remaining researchers in the group will be interviewed, and the data collected and analysed. Responses will be analysed using the A-COM-B model and intervention strategies will be planned and implemented. Each researcher will receive a tailored response that is formulated through addressing specific capability, opportunity and motivational issues, with the aim of influencing attitude and the corresponding behaviours.

Additional intervention methods will take the form of workshops, consultations and drop-in sessions for researchers and support staff. A follow-up interview with the researchers will take place in approximately six months from the initial interview to determine whether attitude and behaviour towards data management have altered and whether engagement with the suggested solutions has occurred.

Given that the A-COM-B framework is assisted by service providers having insights into individual attitudes and behaviours, the challenge of translating the

framework into a model which can be applied to the wider University community will be explored in the next phase of the research project. Mechanisms such as a simplified online survey to capture individual data management practices, filtered through the A-COM-B categories, and coding of responses in order to select intervention strategies quickly will be considered.

Based on the findings from the second phase of the research project, the research team expects to implement a four-part plan:

1. The research team will create a generic toolkit for use by other research support service providers within the Division.
2. The research team will train this cohort in the methodology – and any relevant tools – as outlined in this paper and subsequently refined by feedback from the second phase of the research project.
3. Research support service providers for the four main academic areas within the University will be encouraged to initially work with a research centre with which they have already established a good relationship, so as to encourage open and honest feedback from the participants.
4. Using the data from these later projects, the research team will then explore whether commonalities in data management behaviour that can be inferred through categories such as a researcher's career stage, academic rank and research area, could be used to scale up intervention strategies across the University as a whole.

Based on a continuous improvement cycle, the plan relies on evaluation and adjustment to fine-tune its application. Assuming positive feedback in general throughout the cycle, this would be viewed as a multi-year initiative until such time as critical mass was achieved.

The data collected from this study to date comprises audio interview files, transcripts, interview notes and data analysis. The data will be restricted because the nature of the information means that participants could potentially be identified. The authors intend to create a record about the data in the institutional data repository.

Conclusion

This study has found that in understanding and applying a theoretical framework to the data management practices of researchers in a small research centre, library and information specialists are better equipped to identify underlying behaviours that influence decision making. Although this paper reports on the initial stages of a larger research project, the preliminary findings suggest that attitude is the predominant deterrent to good data management behaviour. By using this framework, practitioners can design intervention strategies that are aligned to individual need, and that lead researchers to using safe and secure institutional solutions and services.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Australian National Data Service (2016a) *Capability Maturity*. Available at: http://www.ands.org.au/guides/capability-maturity (accessed 21 April 2016).

Australian National Data Service (2016b) *Creating a Data Management Framework*. Available at: http://www.ands.org.au/guides/creating-a-data-management-framework (accessed 21 April 2016).

Australian Research Council (2014) *Funding Rules for Schemes under the Discovery Program for the Years 2015 and 2016 – Australian Laureate Fellowships, Discovery Projects, Discovery Early Career Researcher Award and Discovery Indigenous*. Canberra: Australian Research Council.

Brown RA, Wolski M and Richardson J (2015) Developing new skills for research support librarians. *Australian Library Journal* 64(3): 224–234.

European Commission (2016) *Guidelines on Data Management in Horizon*. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 3 May 2016).

Fear K (2011) 'You made it, you take care of it': Data management as personal information management. *International Journal of Digital Curation* 6(2): 53–77.

Federer LM, Lu YL, Joubert DJ, et al. (2015) Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. *PLoS ONE* 10(6): e0129506.

Henty M, Weaver B, Bradbury SJ, et al. (2008) *Investigating Data Management Practices in Australian Universities*. Australia: APSR.

Jahnke L, Asher A and Keralis SD (2012) *The Problem of Data*. Council on Library and Information Resources (CLIR) Report, pub. #154. Available at: https://www.clir.org/pubs/reports/pub154/pub154.pdf (accessed 30 September 2016).

Kennan MA and Markauskaite L (2015) Research data management practices: A snapshot in time. *International Journal of Digital Curation* 10(2): 69–95.

Martinez-Uribe L and Macdonald S (2009) User engagement in research data curation. In: *Research and advanced technology for digital libraries: 13th European conference, ECDL 2009* (eds M Agosti, J Borbinha, S Kapidakis S, et al.), Corfu, Greece, 27 September–2 October 2009, pp. 309–314. Berlin, Heidelberg: Springer.

Michie S, van Stralen MM and West R (2011) The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science* 6(42): 1–12.

National Health and Medical Research Council (2014) *NHMRC's Policy on the Dissemination of Research Findings*. Available at: https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy (accessed 21 April 2016).

National Health and Medical Research Council, Australian Research Council and Universities Australia (2007) *Australian Code for the Responsible Conduct of Research*. Available at: https://www.nhmrc.gov.au/guidelines-publications/r39 (accessed 30 September 2016).

National Science Foundation (2016a) *National Science Foundation History*. Available at: http://www.nsf.gov/about/history/ (accessed 21 April 2016).

National Science Foundation (2016b) *Dissemination and Sharing of Research Results*. Available at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp (accessed 21 April 2016).

O'Reilly K, Johnson J and Sanborn G (2012) Improving university research value: A case study. *SAGE Open* 2(3): 1–13.

Peters C and Dryden AR (2011) Assessing the academic library's role in campus-wide research data management: A first step at the University of Houston. *Science & Technology Libraries* 30(4): 387–403.

Piderit SK (2000) Rethinking resistance and recognizing ambivalence: A multidimensional view of attitudes toward an organizational change. *Academy of Management Review* 25(4): 783–794.

Quacquarelli Symonds (2015) *QS top 50 under 50 2015*. Available at: http://www.topuniversities.com/top-50-under-50 (accessed 10 May 2016).

Research Councils UK (2014a) *Research Outcomes Overview*. Available at: http://www.rcuk.ac.uk/research/researchoutcomes/ (accessed 20 April 2016).

Research Councils UK (2014b) *About researchfish*. Available at: http://www.rcuk.ac.uk/research/researchoutcomes/researchfish/ (accessed 20 April 2016).

Research Councils UK (2015) *RCUK Common Principles on Data Policy*. Available at: http://www.rcuk.ac.uk/research/datapolicy/ (accessed 20 September 2016).

Schumacher J and VandeCreek D (2015) Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation* 10(2): 96–109.

Searle S, Wolski M, Simons N, et al. (2015) Librarians as partners in research data service development at Griffith University. *Program: Electronic Library & Information Systems* 49: 440–460.

Si L, Xing W, Zhuang X, et al. (2015) Investigation and analysis of research data services in university libraries. *The Electronic Library* 33(3): 417–449.

Stanford University Libraries (2013) *Stanford Prize for Innovation in Research Libraries (SPIRL)*. Available at: https://library.stanford.edu/projects/stanford-prize-innovation-research-libraries-spirl/2013-prizes (accessed 20 April 2016).

Tenopir C, Allard S, Douglass K, et al. (2011) Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6(6): e21101.

Tenopir C, Sandusky RJ, Allard S, et al. (2013) Academic librarians and research data services: Preparation and attitudes. *IFLA Journal* 39(1): 70–78.

Wang M and Fong BL (2015) Embedded data librarianship: A case study of providing data management support for a science department. *Science & Technology Libraries* 34(3): 228–240.

Weller T and Monroe-Gulick A (2014) Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech* 32(3): 467–482.

Westra B (2014) Developing data management services for researchers at the University of Oregon. In: Ray J (ed.) *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IN: Purdue University Press, pp. 375–391.

Wolff C, Rod AB and Schonfeld RC (2016) *Ithaka S+R US Faculty Survey 2015*. Available at: http://www.sr.ithaka.org/publications/ithaka-sr-us-faculty-survey-2015/ (accessed 20 April 2016)

Wolski M and Richardson J (2015) Improving data management practices of researchers by using a behavioural framework. In: *THETA 2015*, Gold Coast, Australia, 11–13 May 2015. Available at: http://hdl.handle.net/10072/69141 (accessed 30 September 2016).

World Bank (2010) *Theories of Behavior Change. Communication for Governance and Accountability Program (CommGAP)*. Washington, DC: World Bank.

Author biographies

Susan Hickson is a Library Services Manager (Business) at Griffith University, Australia and a librarian by profession. Sue is responsible for leading a team of professional Librarians, Learning Advisers and Digital Capability Advisers who support academics, researchers and students in the areas of Research, Learning and Teaching. Previously Sue was a Faculty Librarian supporting the Health Group.

Kylie Ann Poulton began her library career as a researcher in investment banking before joining Information Services at Griffith University. During her career at Griffith she has

held roles in Liaison Librarianship, Project Management and Change Management. She has also led Griffith's Higher Education Research Data Collection team and worked as a project officer in the University's Research Office. She is currently Griffith University's Business Librarian.

Maria Connor is the Business Librarian at Griffith University Gold Coast Campus, Australia. She holds an MLIS from Victoria University, Wellington, New Zealand.

Joanna Richardson PhD is Library Strategy Advisor at Griffith University, Australia. Previously she was responsible for scholarly content and discovery services including repositories, research publications and resource discovery. Joanna has also worked as an information technology librarian in university libraries in both North America and Australia, and has been a lecturer in Library and Information Science. Recent publications have been centred around library support for research and research data management frameworks.

Malcolm Wolski is the Director, eResearch Services, at Griffith University. In his role, he is responsible for the development, management and delivery of eResearch services to support the University's research community, which includes the associated information management systems, infrastructure provision, data management services and media production. Malcolm is a part of the senior leadership team providing library, information and IT services at the Griffith University and he works closely with these groups to provide service desk, infrastructure and outreach services to the research community.

Article

International Federation of Library Associations and Institutions
2016, Vol. 42(4) 266–277

DOI: 10.1177/0340035216671963
ifl.sagepub.com

Research data services: An exploration of requirements at two Swedish universities

Monica Lassi
Lund University, Sweden

Maria Johnsson
Lund University, Sweden

Koraljka Golub
Linnaeus University, Sweden

Abstract

The paper reports on an exploratory study of researchers' needs for effective research data management at two Swedish universities, conducted in order to inform the ongoing development of research data services. Twelve researchers from diverse fields have been interviewed, including biology, cultural studies, economics, environmental studies, geography, history, linguistics, media and psychology. The interviews were structured, guided by the Data Curation Profiles Toolkit developed at Purdue University, with added questions regarding subject metadata. The preliminary analysis indicates that the research data management practices vary greatly among the respondents, and therefore so do the implications for research data services. The added questions on subject metadata indicate needs of services guiding researchers in describing their datasets with adequate metadata.

Keywords

Academic libraries, data services, metadata and semantic web, organization of information, services to user populations, types of libraries and information providers

Submitted: 20 May 2016; Accepted: 6 September 2016.

Introduction

Over the recent years a plethora of discussions have been taking place in relation to scholarly research data and the role of academic libraries and research institutions in providing various services to support research data management (see, for example, Borgman, 2015). Libraries that are not yet offering support for research data may be at the stage of developing such services, as is the case of two Swedish universities in focus of this paper. The libraries of Linnaeus University and Lund University have chosen to first identify characteristics, requirements, needs and related issues of managing different types of research data produced by researchers employed by the respective universities, which would then serve as a foundation to create most appropriate research data management services. The study is based on interviews with researchers, following the Data Curation Profiles (DCP) Toolkit (Carlson and Brandt, 2014). Based on the toolkit, a data curation profile is created for each of the interviewed researchers. The profile may be used by the researcher for research data management (RDM), and by libraries in developing support for RDM.

The outline of the paper is as follows. In the next section (Background), the context of research data services provided by libraries as seen in professional discussions, project reports and research literature is

Corresponding author:
Monica Lassi, Department of Scholarly Communication at University Library, Lund University, PO Box 3, Lund, 22100 Sweden.
Email: monica.lassi@ub.lu.se

provided. This is followed by a section (Methodology) on the methods of the study. The results are presented and analysed in the next section (Results). The last section (Discussion) concludes the paper with a discussion on the implications of the results for developing research data services.

Background

Research data management and services in Sweden

During the past years several RDM projects and events took place at Swedish universities; many of which had been initiated by academic libraries and archives. There has been a lot of focus on data sharing and open data, and lately also on preservation and archiving, in particular with projects for e-archiving.

Several projects have investigated researchers' experiences and attitudes regarding research data management. In 2007 libraries and archives at University of Gothenburg, Lund University and the Swedish University of Agricultural Sciences undertook a joint pre-study with the aim of exploring research data in open archives and university archives. The study comprised a survey on researchers' attitudes towards publishing research data and looked specifically at the future roles of open archives and university archives (Björklund and Eriksson, 2007). The Swedish National Data Service (SND) performed a major study in 2008–2009 on practices of open access to and reuse of research data. The study comprised two surveys targeting professors and PhD students working at Swedish universities. The survey results pointed to the following:

- a number of barriers to sharing research data;
- unresolved issues regarding legal and ethical aspects;
- lack of resources to make research data available.

It concluded that researchers need to be trained in relation to research methods, digital research databases and accessible e-tools, as well as that funds should be made available for preparing research data to be shared and archived (Carlhed and Alfredsson, 2009).

SND participated in a similar project also involving university libraries at the University of Gothenburg, Lund University, Linköping University and Malmö University, focusing on researchers within Arts and Humanities. The participating libraries interviewed researchers about their attitudes towards publishing their research data. The project revealed a positive attitude in this respect, and identified the need for quality RDM systems and related professional support (Andersson et al., 2011).

In 2014 the Lund University Library undertook an investigation of the conditions for RDM within Humanities and Social Sciences at the university. It both looked at the organizational aspects for RDM, and brought in researchers' opinions through a small survey (Åhlfeldt and Johnsson, 2014).

These above mentioned projects were all early, exploring projects on RDM, with a focus on data sharing. The conditions for this project with interviews with Data Curation Profiles have been a bit different, as the debate on open data and RDM have been more animated the last year. With the Data Curation Profiles we have had the possibility to study researchers' experiences on a more individual basis and close to their everyday work.

At the national level it is the Swedish Research Council that is the major stakeholder in RDM, active particularly in funding related research infrastructures in Sweden, such as SND, Environment Climate Data Sweden (ECDS), Bioinformatics Infrastructure for Life Sciences (BILS) and the Max IV Laboratory which provides access to synchrotron X-rays. At the request of the Swedish government, in 2015 the Swedish Research Council proposed a national policy for open access to scientific information, including publications and research data (Swedish Research Council, 2015). The Swedish government's stance on the proposal is expected to be made clear in the fall of 2016 when the research bill is anticipated. The policy aims to ensure open access to all Swedish scientific information that has been fully or partially funded by public funds, by the year 2025. In Sweden, research data are owned by the academic or research institution at which the researcher is employed (see Bohlin, 1997), and not by the researchers themselves as is the case in many countries, such as Finland. However, this is not very well known by Swedish researchers; they regard the data they have collected as their own (Swedish Research Council, 2015). The proposal suggests a model in which the responsibility for archiving and providing access would lie with different organizations. Academic and research institutions would have responsibility for archiving research data, and a national facility would be responsible for making the research data available by linking to the research data archived at the academic and research institutions.

In spite of the absence of an established national policy, there is a relatively strong spirit of professional development in RDM at Swedish universities and research infrastructures, as is particularly evident from the considerable number of events aimed at

RDM research and training. The primary player is the Swedish National Data Service (SND). Apart from arranging training events in research data management, SND is also coordinating pilot projects on RDM at several Swedish universities during 2016.

Development of research data services – Linnaeus University

Research data services at Linnaeus University (LNU) have been in the planning stage since 2015, when a librarian working as a research strategist at the university library was chosen to serve as the representative for SND. Apart from taking part in SND's training events, the LNU library liaised with the Digital Humanities Initiative at LNU (Linnaeus University, 2016) beginning in February 2016. As one of the original eight pilot projects planned as part of this initiative, the Humanities Data Curation pilot project was envisioned. It set out to be conducted by the Digital Humanities researchers who would collaborate with LNU Library SND representative as well as the Lund University Library, with a particular focus on the survey reported here. The plan for further steps is under development.

Development of research data services – Lund University Library

At the Lund University Library, the Department of Scholarly Communication addresses RDM and development of related services. Lund University Library is one of 26 libraries in the network of Lund University Libraries. Whereas most libraries provide a service to a particular faculty, department or centre, the University Library provides a service to the entire University (Lund University Libraries, 2016). There are a number of RDM-related projects coming up at Lund University, and in several of these library staff is involved. The university hosts several research infrastructures which are working in different ways with questions of data sharing, data preservation, etc. During the fall 2015 Lund University Library conducted a survey of the university faculties regarding their experiences of and attitudes to RDM (Johnsson and Lassi, 2016). The survey, in which faculty managers and library staff were interviewed, showed that the respondents expect RDM to become more important in the future, and that it will become a part of researchers' work tasks in a more structured way than at present. The results of the survey also showed the diversity of research data generated within a full university, and verified the need to investigate the characteristics of different types of research data within different fields. The Lund University Library is also involved in projects of research infrastructures at the university, such as the ICOS Carbon Portal (ICOS Carbon Portal, 2016).

Subject metadata in research data

Naming and organizing data and relationships among them can have 'profound effects on the ability to discover, exchange, and curate data' (Borgman 2015: 65). Standardized metadata schemes increase these abilities. While in many scientific areas metadata schemes are well used (although there are still gaps in a number of them), the question is raised here about the degree to which subject metadata are standardized. Subject searching (searching by topic or theme) is one of the most common and at the same time the most challenging type of searching in information systems. Subject index terms taken from standardized knowledge organization systems (KOS), like classification systems and subject headings systems, provide numerous benefits compared to free-text indexing: consistency through uniformity in term format and the assignment of terms, provision of semantic relationships among terms, support of browsing by provision of consistent and clear hierarchies (for a detailed overview see, for example, Lancaster 2003). In terms of cross-searching based on different metadata schemas using subject terms from different KOS, challenges like mapping across the KOS need to be addressed in order to meet the established objectives of quality controlled information retrieval systems like those provided by libraries.

Libraries are providing quality subject access of resources described in library catalogues through, for example, classification schemes and subject headings; however, when it comes to research data, there seems to be a lack of controlled subject terms in metadata schemes. An exploratory analysis of 36 disciplinary metadata standards from the list provided by the Digital Curation Centre (2016) in the UK shows that 18 (50%) of them to do not provide any subject metadata field. Of those that do, only a few offer clear guidelines on the controlled vocabulary use, while others leave it to be a freely added keyword.

Several examples of the former include the following:

- SDAC (Standard for Documentation of Astronomical Catalogues) which provides three categories of keywords: (1) the ones as in the printed publication, (2) controlled ADC keywords (take from controlled sets), and (3) mission name, a header which precedes the

satellite name for data originating from the satellite mission.
- FGDC/CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata) supports two categories of subject metadata: (1) theme keyword thesaurus, from a formally registered thesaurus or a similar authoritative source of theme keywords, (2) theme keyword, a common-use word or phrase used to describe the subject of the data set.
- ClinicalTrials.gov Protocol Data Element Definitions for its keyword element advises use of words or phrases that best describe the pro tocol with a note to use Medical Subject Head ings (MeSH) where appropriate.

Several of the others use Dublin Core's dct: subject element, which as in the previous examples allows free keywords, while recommending the use of controlled vocabulary.

Methodology

The research questions guiding the study are:

1. What are the respondents' current practices of research data management?
2. What are the respondents' current practices of using subject metadata to describe their data?
3. What are the implications for developing research data services?

The research questions have been investigated through structured interviews with researchers, during which they were asked to fill out the interview worksheet of the Data Curation Profiles Toolkit, while the session was audio recorded. The researcher was asked to respond to the questions in the interview worksheet with a specific project in mind on themes such as sharing, ingestion to a repository and organization and description of data. The study design is further elaborated below.

Data Curation Profiles Toolkit

In this exploratory study the Data Curation Profiles (DCP) Toolkit was used, developed through a research project conducted by the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois Urbana-Champaign (Witt et al., 2009). The DCP Toolkit is envisioned as an aid to starting discussions between librarians and faculty and in the planning of data services that directly address the needs of researchers. The profiles are supposed to give information about a particular data set and the researcher's doings in terms of curating that data set.

The method consists of an interview template with questions concerning RDM of a specific data set. The interviewee is the researcher who has created the data set and the interviewer is a librarian or from a related profession. Also included in the DCP Toolkit are guidelines and instructions to the interviewer as well as instructions to the interviewee. The interviews are to result in data curation profiles for specific research areas or projects, and could be useful both to research support staff as well as to the researchers themselves.

The DCP template comprises 13 modules on RDM of a specific data set. The questions cover the nature of the specific data set, such as form, format, size, and bring up aspects like sharing, archiving, discovery, organization and description, linking, interoperability, and measuring impact. They are formulated in a generic way in order to allow their use in all kinds of scientific areas. In order to address metadata and subject access in particular, four questions were added to the interview worksheet in Module 6 – Organization and Description of Data, each followed by a 5-point Likert scale ranging from 'not a priority' to 'high priority', followed by the 'I do not know' option (please see Appendix for the modified version of the list of questions in Module 6):

- The ability to apply standardized subject classification to the data set. (Please list relevant controlled vocabularies next to this table);
- The ability for automatic suggestions of keywords for subject classification;
- The ability to add your own tags to the data set; and
- The ability to connect the data set to the keywords of the publications that are based on it.

After the development and launch of the DCP Toolkit, in 2011 about a dozen workshops on how to use the toolkit were held across the USA. As a result, several libraries have constructed data curation profiles which have been published in a public directory of the DCP (Carlson and Brandt, 2015).

In the years since 2011 several projects have used the DCP in the development of RDM services (Bracke, 2011; Brandt and Kim, 2014; Carlson and Bracke, 2013; McLure et al., 2014; Wright et al., 2013). The Cornell University Library used DCP when developing a research data registry at the university (Wright et al., 2013). The library performed eight interviews with researchers in a wide range of subject areas. In spite of considerable variety in researchers' priorities, the project team was able to

discern similarities among their needs. In another project at Purdue University, DCP was used to investigate the needs for data curation for researchers within agricultural science (Bracke, 2011). The project was focused on establishing the role which the subject librarian in Agricultural Science could take in data curation. The project performed interviews based on the DCP, and the identified data sets were put in a prototype data repository.

One example where they used selected parts of the DCP was a project at the Library of Colorado State University (McLure et al., 2014). They conducted a number of focus group interviews with researchers in order to explore what kind of data sets researchers had, how they managed their data, and what kind of support they would need in terms of RDM. Methodologically, they also investigated the feasibility of adapting DCP to focus groups. The findings showed that focus groups may be very useful when investigating general conditions for RDM and to spot trends and behaviours among researchers. On the other hand, conducting individual interviews using the DCP gives more specific, granular detail on researchers' data management (McLure et al., 2014).

Data collection

Data was collected through structured interviews which were guided by the DCP Toolkit's interview worksheet, described above. The DCP interview template was translated into Swedish to fit the Swedish-speaking researchers at Lund University. At Linnaeus University all interviews were held in English, but allowing the respondents to reply in Swedish if that was their preference – this was because one of the interviewers there was not a native Swedish speaker. The respondents were recruited through email advertisements, personal contacts with researchers and a faculty survey (the latter only at Lund University).

In total 12 interviews were conducted – five at Linnaeus University and seven at Lund University. The interviews were conducted between 28 January and 28 April 2016 and lasted between 46 and 119 minutes. Respondents were at different stages of their research career, ranging from PhD candidate to professor. Further, the respondents were active in a wide range of research areas, such as archaeology, biology, business administration, film studies, library and information science, and media and journalism. In this exploratory study we could not cover all disciplines, e.g. informatics and chemistry. A follow-up study with a larger sample size would be able to cover more disciplines.

At the start of each interview, the respondents were presented with information about the study and the interview session, and were asked to read through and sign an informed consent form. Following the research ethics guidelines of the Swedish Research Council (2011) the informed consent form stated that all answers would be anonymized and that no risks of participating in the study could be predicted. Other information in the consent form included the aims of the study and the explanation that the data collected might be used for scientific publishing and in other studies, in which case all data would be anonymized.

The respondents were asked to fill out the DCP interview worksheet during the interview session, which was also audio recorded. The audio recordings served as an aid for the interviewers to capture any explanations or discussions that arose from the interview worksheet questions that might not have been written down in the worksheet. The interview worksheets were scanned and stored along with consent forms and audio recordings in an online collaborative environment, with access allowed only to the three authors of the paper.

Note that the DCP Toolkit recommends that two sessions are conducted with each researcher in order to allow for time to learn from the discussions that arise from the interviews. In the reported study only one session per respondent was scheduled in order to avoid the risk of getting fewer respondents because it would take too much of their time for anyone to accept the invitation to take part in the study.

Data processing

The responses from the interview worksheet were transferred into MS Excel. The respondents were anonymized in the MS Excel spreadsheet in order to facilitate the reuse of the data set. The audio recordings were used to take notes of statements that could be of interest in relation to the research questions, the statements that further explain something from the interview worksheet or those that were interesting examples of situations concerning RDM.

For each of the respondents at Lund University, a data curation profile was created based on the instructions provided by the DCP Toolkit. The audio recordings contributed to creating a set of recommendations tailored to each respondent concerning their current and future RDM practices. The recommendations section is a modification of the original data curation profile. The data curation profile was delivered to the researcher with an invitation to a follow-up meeting on the DCP. We followed the detailed guidelines and instructions for constructing the data curation profiles

from the collected data. Together with the audio recordings, the paper-based interview worksheets provided rich material for creating the data curation profiles.

Data analysis

Guided by the research questions, the data analysis was started by a simple numeric analysis of the number of responses to each question. These results were complemented by analysing the relevant modules of the audio recordings, aiming to find any explanatory statements or comments to the questions and their responses. Whereas the numerical data analysis was conducted for all the data, in this paper we focused on Modules 3 (Sharing), 5 (Ingestion), 6 (Metadata) and 7 (Discovery) for the qualitative analysis, as these were deemed to in conjunction cover many, but not all, of the aspects of RDM.

Results

As stated, the respondents conduct research in a wide range of scientific research areas, using a wide range of research methods to collect, process and analyse data. Also, the study is based on a small number of respondents, 12 persons in total. Therefore, the results are presented to show this broad array of experiences and descriptions.

Research data sets

The tools used by the respondents to generate data varied greatly and included audio recorder (3), camera (3), questionnaires (2), pen and paper (2), sensors, field notes and a plethora of software tools. Most of the tools required to utilize the respondents' data are proprietary software that requires a purchase to use, such as Excel (4/12), SPSS (2), NVivo (1), Word (1) and MATLAB (1). The R software environment and programming language and the Genome browser are two examples of tools used, available under licences allowing free use for academic purposes.

The DCP interview guide asked the respondents to identify the stages that their data had passed through during their research project. All respondents identified at least two stages, commonly three. The first stage typically concerns raw data collected or obtained by the respondents. The second stage typically concerns some type of processing, e.g. quality checking or cleaning the data, or preliminary analysis, e.g. first coding. In total 11 out of 12 persons declared they had a third data stage. In the third data stage the data are often in a second form of analysis format, and in this phase the researchers start posing their research questions to the data, or start with the deeper form of analysis. Of the 12 data sets four were bound by privacy or confidentiality agreements, whereas five were not, and one respondent was unsure.

Attitudes towards, and incentives for, sharing data

Out of the 12 respondents two have deposited the data in focus in the interview in a repository. Among the other 10 respondents, eight stated that they were willing to deposit their data, while two were hesitant to share their data. The hesitancy was motivated by the fact that the data was sensitive, comprising interviews, and that the researcher had promised their respondents confidentiality. Further, 11 of the 12 respondents stated that their data would probably be of interest to others, suggesting, for example, public libraries, media companies, teachers, study participants, activists and non-government organizations. When speculating about the use that others could have of their data, respondents suggested addressing new research questions, developing educational programmes, conducting statistical analysis, and serving for informed policy making. As to citation, nine of the 12 respondents would require a citation or an acknowledgement when others used their data.

The first data stage of the data management cycle was commonly seen as the best stage to share, by 10 out of 12; of these, five would share the data with anyone, and the other five would prefer to share with immediate collaborators or others in their field. Whether the respondents would share the data at the second data stage differed: three respondents stated that they would not share at that stage, while nine stated that they would share, but predominantly with immediate collaborators or researchers in the same field (5). The hesitancy to share data at this stage could perhaps be explained by the processing or initial analysis done to the data, having moved from raw data to another state which could possibly be more difficult for others to understand or use. Of the 12 respondents eight indicated that they would share their data after they had published results.

As stated above, in Sweden research data are owned by the academic or research institution at which the researcher was employed at the time of the data collection. Of the 12 respondents five reported knowing that their employers owned the data, whereas the answers to this question among the other seven respondents varied to include the researcher (4), the research community (1) and the general public (1). The research funders of the data collected in the respondents' studies generally do not require data to be shared or deposited in a repository (11/12), while

one person stated that there was such a requirement but that it would be illegal to share the data due to privacy protection.

The ability to see usage statistics of their data sets had a high priority (5/12) or medium priority (5) for the majority of the respondents. The majority of the respondents indicated that the ability to gather information about the people who have used the data set would be of high (6/12) or medium priority (3), and a few that it would be of low priority (3). Suggestions of other metrics or analytics that could be of interest to the respondents include citations (4), seeing publications based on the data set, and the country, academic or research institution, and time at which a data set has been used.

Metadata

The respondents organized and described their data sets in a wide variety of ways, including metadata schemes, codebooks, a paper filing system using plastic folders ordered geographically and by topic, and tables in a MS Excel file. Two respondents used standardized metadata schemes: one scheme is provided by a Swedish national infrastructure, and one is a domain-specific taxonomy. One respondent reported using a particularly wide range of different descriptions to cover the needs of others who may be interested in the data, including: a standardized metadata scheme coupled with their own classifications to describe instruments and parameters measured, technical specifications of, for example, how often measurements were conducted, and a verbal description of how the data had been collected and processed.

The majority (7/12) considered that the existing organization and description are sufficient for others to understand and use the data. One respondent noted that there is a risk that codebooks are too detailed and complex for someone else to read, which could impede the understanding and correct use of the data.

The ability to apply standardized metadata from the respondents' own fields was considered important by most (high priority by 8/11, medium priority by 2, low priority by 1). The ability to apply standardized subject classification to the data set was also considered of medium to high importance (high priority 5/11, medium priority 5, low priority 1).

Most (10) respondents requested more information about what was meant by the question '[t]he ability to apply standardized subject classification to the data set'. This could indicate that the question is designed with library and information science jargon; while researchers may be required to add keywords and other subject classification when submitting publications, they might not know the name for this activity. Respondents who elaborated on their responses in the worksheet indicated that subject classification is important for others to find their data sets and that standardization of the classification is needed to ensure findability. Also, a respondent described the importance of standardized KOS in archaeology, as places have changed names throughout history.

The ability for automatic suggestions of keywords for subject classification was also deemed as high to medium priority (high priority 6/11, medium priority 3, low priority 2). Out of the six respondents who rated this as having a high priority, four stressed the importance of them being able to make the final decision as to which suggestions to add to the data set. One respondent noted that automatic suggestions of keywords would be a good idea to start to think about classification; they generally do not think about keywords before it is required to add them to a publication.

Ability to add one's own tags was considered the top priority (high priority 9/11, medium priority 2) regarding subject metadata features. One respondent justified the high priority by explaining that the development of the research area moves quicker than the classification systems, so tagging as a complement to standardized subject classification would increase findability and show the data set's uniqueness. Another respondent suggested that tags could be useful to inform others about anomalies that might be interpreted as errors but are actually just outliers.

Ability to connect the dataset to the keywords of the publications based on it was also considered rather important (high priority 5/11, medium priority 3, low priority 1, N/A or I do not know 1). Among the more hesitant respondents, one responded that it would just be confusing, and another one stated that everything that can be automated does not have to be automated in its own right; choosing keywords oneself could add a level of reflection to the process. Having a more positive attitude towards this, one respondent stated that the importance of this ability is more or less self-evident.

Data ingestion into a repository

All (12/12) respondents reported on the need to prepare the data for their ingestion into a data repository to some extent, ranging from what the respondents viewed as hardly any work at all (2) for raw data, to a lot of work (1). Activities needed were reported to be related to making the data set understandable to others by preparing codebooks and metadata; filling out gaps or fixing errors in the data; digitizing

materials that are on paper; and anonymizing identifying information from interviews. One respondent expanded on the process of making the data set understandable through adding sufficient metadata to them, stating that it takes a lot of time and effort, and needs to be balanced with how much work is needed for the data to be understandable and usable.

Most respondents (7/12) expressed high priority for the ability to submit the data by themselves, while a few (4) thought it was low priority. Two respondents stated that preparing and submitting the data should be done by experts, one of them suggesting the library as a potential service provider. A respondent who expressed high priority to submitting the data themselves explained that they wanted control of the ingestion process.

As to the automatized submission process, the opinions ranged from 'it is not possible' (1) and 'not a priority' (5), to 'medium priority' (2) and 'high priority' (4). Two respondents stated that they would appreciate it as a time-saving factor and the fact that it would be possible only for some types of data such as those that do not need to be anonymized, or particularly secured for privacy reasons. One respondent stated that they were generally in favour of automating processes, but for these purposes it seems a bit risky in case there are errors in the data set that they have not yet had time to fix before the automated submission process starts.

The possibility of batch upload to repository was considered high priority by six of 12 respondents, medium priority by a few (2), and low and no priority by one person each; in addition, two people chose the N/A/I do not know option. As above, the time factor was reported as an important factor for batch upload, as well as creating a collection of all the data (thousands of documents, photos and notes) in one set.

Discovery and use of the data

The ability for researchers within the same scientific area to easily find the data was ranked as high priority by all respondents, although one referred to metadata about the data, rather than data itself due to high confidentiality. Possibility for researchers outside the scientific area at hand to easily find the data was also considered a high priority (10 high priority, 2 medium; of the high 1 answered for metadata in the same context as for the previous question). The possibility for the general public to easily find the data and the possibility to discover them via a search service was also considered important; on average it was less important than the first two possibilities to make the data available to researches. For the general public the distribution was three high priority, three medium, four low, two not a priority; for discovery via a search service like Google the distribution of answers was five high priority, two medium, three low, two not a priority.

The replies varied greatly as to how the researchers envision the users would use the data, and included various types of discovery services, ranging from Google, databases like repositories or library catalogues, integrated cross-database search, to defining interfaces such as access through metadata, datasets divided into subject areas, search box with keywords, classification-based browsing interface.

The ability to connect data sets to visualization or analytical tools was highly desired by most respondents (7/12), but had a low (2) or no priority (1) for a few others. Allowing others to comment on or annotate their data sets had varying importance from high (4/12) and medium priority (3), to low (2) or no priority (3).

Regarding the ability to connect data with publications and other research output, most respondents reported a high (5/12) or medium priority (4), with a few giving this a low priority (2), or did not know (1). Support for web service APIs to give access to their data was seen as being either of high priority (4/12) or being of low priority (3). Three respondents responded that they did not know or that this was not applicable for their data. The ability to connect or merge their data with other data sets was again reported as either a high (4/12) or medium priority (5), or not a priority at all (3).

Data preservation

Which parts of the data set that would be most important to preserve over time varied greatly. Some respondents (4 out of 12) responded that everything is equally important, including documents from funders, project descriptions, data at different stages of the research process and publications. Other respondents (4) indicated that it would be sufficient to preserve the raw data. Lastly, some respondents (4) responded that one or a few stages of the processed data would be most important to preserve. The time period for which the respondents estimated that their data would be useful or valuable to others ranged from indefinitely (5) to 10–20 years (3), 5–10 years (1), 3–5 years (2). One person indicated that they did not know.

The ability to audit datasets to ensure structural integrity long-term was indicated to be of either high (6/12) or medium priority (3). Two respondents responded that this had a low priority, and one that they did not know.

Migrating datasets into new formats over time had a high (8/12) or medium priority (3), whereas only one respondent indicated that it had a low priority. Similarly, the need for a secondary storage site for datasets was indicated to be of high (8/12) or medium (2) priority, one respondent responding that it was not a priority, and one respondent indicating that they did not know. Having the secondary storage site at a different geographic location was not as prioritized, responses varying from high (5) to medium priority (4), one respondent giving it a low priority, and one responding that they did not know.

Concluding discussion

During the interviews, the respondents shared their experiences with collecting, processing, managing and storing data in their everyday work. They could easily give examples of situations that arise in their RDM practices and procedures that they take for their data. However, they had not always reflected upon why they worked in a certain manner and if there were other ways of managing their data, for example concerning going about long-term preservation and sharing of data. As expected, some respondents had more elaborate and thorough protocols for RDM, predominately in areas that require meticulous protocols such as environmental research based on observational data.

Generally the respondents are positive towards the idea of sharing research data, and they show a clear awareness about it. For most of the researchers, sharing and exchanging data with research colleagues is very common and natural, in particular when collaborating with others. This could possibly include sharing data for educational purposes, as open educational resources. If there are sensitive aspects to the research data, the researchers are also aware of this, not least if they have had to get permissions from an ethics board to collect the data. Concerning which types of research data to share, it seems least problematic for the respondents to share data from the first data stage, e.g. raw data. It seems the respondents are more hesitant about sharing data which has undergone some initial processing or analysis, corresponding to the second or third data stage. Those who are willing to share data from the second or third data stage would prefer to share with immediate collaborators or researchers in the same field only. There is a concern that the 'early processed' data may not be properly analysed or interpreted by an external researcher. In these context libraries and other research support services may play an active role in providing guidance on how to share data in trustful and secure ways, including setting embargos on data sets when metadata could be available, though the data themselves are not.

As to data ingestion, the fact that most respondents need to prepare the data for their ingestion into a data repository means that in order to save the time of the researcher, a data management plan would need to be in place, which could help the researcher plan for the ingestion already in the planning phase of a project. Similarly, services from the library could be provided to support both the planning and the ingestion, which was suggested by a few respondents. Combining the possibility for researchers to submit data to a repository themselves with a service in which librarians deposit data on the researcher's behalf would be useful, to facilitate sharing for researchers who want control over the ingestion phase, as well as those who do not have the time to do this themselves. By providing training in data ingestion, researchers could become more and more self-sufficient in sharing their data themselves.

As to data organization and description, services to support organization and description of data in order for others to use them, need to be provided. Availability in multiple formats should also be supported. Standardized metadata, controlled vocabularies, automated suggestions of keywords in general and automated assignment of keywords from the related publication(s) in a repository, ability to add own keywords should all be supported. Providing training to researchers in how to use metadata to correctly describe data also seems to be a valuable service. Such training could include the different characteristics and uses of subject metadata derived from controlled vocabularies compared to add user-generated tags, and how to use the correct terms for the intended audience of the data set.

As to funders, there still does not seem to be a policy requiring them to draft a data management plan, to share publish or deposit data in a repository, and in a majority of cases to preserve the data beyond the life of the funding. However, this may change soon depending on the Swedish government's decisions regarding the policy suggested by the Swedish National Research Council, so it is a good time to start planning for these services. Provision of authorized access only would also need to be ensured, as a considerable portion of datasets may be bound by privacy or confidentiality concerns.

A long-term preservation policy, whether local or national, needs to take into consideration the resources required to archive vast amounts of data created at universities in Sweden. Although archives typically require some selection of resources to be

archived, the respondents saw great value in all their materials being archived. It may also be relevant to preserve auxiliary resources related to a research data set, including data collection instruments, software and databases used and created during different stages of the research process. This could be addressed in future research.

Using the DCP Toolkit to gather data about researchers' RDM and needs for service development was a valuable experience. Based on the interviews we created data curation profiles that can provide valuable guidance for the researchers and the results of the interviews will be considered in our further research data service developments. In some of the interviews, the work sheet responses were quite brief, and thus hard to translate into something meaningful in the data curation profile. The audio recordings of the interviews provided extra information what helped to make a richer profile. Also, we added an extra section to the data curation profile with personalized recommendations focusing on the research data management of future research projects. This was a particularly appreciated part of the data curation profile, according to communication and follow-up meetings with the respondents. As the construction of data curation profiles was a rather cumbersome process, we would want to simplify the process if we were to introduce data curation profiles as a regular service. This would require further investigation into the benefits and challenges of the DCP Toolkit from the researchers' perspective as well as the library's perspective. According to the website, the DCP Toolkit is in the planning stages of redesign, so the profile construction process may be made simpler in the future.

The next step of this project is to continue to analyse the qualitative data, i.e. the recordings of the interview sessions, which will provide complementary information to the DCP Toolkit questions, including explanations to the responses to the worksheet questions. Lund University Library is in the planning stages of a study evaluating a tool for data management plans, which will complement the results from this study concerning, for example, file types, volumes of data and needs for storage solutions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Andersson U, Alfredsson A, Arvidsson S, et al. (2011) *Projektet Forskningsdata inom humaniora och konstnärliga vetenskaper - Open access? Projektrapport till Kungl. biblioteket, Programmet för OpenAccess.se* [The Project Research Data in the Humanities and Arts – Open Access? *Project Report to the National Library of Sweden, Programme for OpenAccess.se*]. Available at: https://gupea.ub.gu.se/bitstream/2077/29112/1/gupea_2077_29112_1.pdf (accessed 9 September 2016).

Björklund C and Eriksson J (2007) *Forskningsdata i öppna arkiv och universitetsarkiv: en förstudie vid Göteborgs universitet, Lunds universitet och Sveriges Lantbruksuniversitet.Projektrapport till Kungliga biblioteket, Programmet för OpenAccess.se* [Research Data in Open Repositories and University Repositories: A Pilot Study at Gothenburg University, Lund University and Swedish University of Agricultural Sciences. *Project Report to the National Library of Sweden, Programme for OpenAccess.se*]. Available at: https://gupea.ub.gu.se/bitstream/2077/7379/1/forskningsdata_rapportKB.pdf (accessed 9 September 2016).

Bohlin A (1997) *Offentlighet & sekretess i myndighets forskningsverksamhet* [Public Disclosure and Secrecy in Government Research]. Stockholm: Riksarkivet.

Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Bracke M (2011) Emerging data curation roles for librarians: A case study of agricultural data. *Journal of Agricultural and Food Information* 12(1): 65–74. DOI: 10.1080/10496505.2011.539158.

Brandt SD and Kim E (2014) Data curation profiles as a means to explore managing, sharing, disseminating or preserving digital outcomes. *International Journal of Performance Arts and Digital Media* 10(1): 21–34. DOI: 10.1080/14794713.2014.912498.

Carlhed C and Alfredsson I (2009) Swedish National Data Service's strategy for sharing and mediating data: Practices of open access to and reuse of research data - the state of the art in Sweden 2009. In: *IASSIST/IFDO 2009, Mobile data and life cycle. IASSIST's 35th annual conference*, Tampere, Finland, 26–29 May 2009. Available at: http://uu.diva-portal.org/smash/get/diva2:396306/FULLTEXT01.pdf (accessed 9 September 2016).

Carlson J and Bracke MS (2013) Data management and sharing from the perspective of graduate students: An examination of the culture and practice at the Water Quality Field Station. *portal: Libraries and the Academy* 13(4): 343–361. DOI: 10.1353/pla.2013.0034.

Carlson J and Brandt SD (2014) *Data Curation Profiles*. Available at: http://datacurationprofiles.org/ (accessed 9 September 2016).

Carlson J and Brandt SD (eds) (2015) *Data Curation Profiles Directory*. Available at: http://docs.lib.purdue.edu/dcp/ (accessed 9 September 2016).

Digital Curation Centre (2016) *Disciplinary Metadata*. Available at http://www.dcc.ac.uk/resources/metadata-standards (accessed 9 September 2016).

ICOS Carbon Portal (2016) *ICOS Carbon Portal*. Available at: https://www.icos-cp.eu (accessed 9 September 2016).

Johnsson M and Lassi M (2016) *Hantering av forskningsdata vid fakulteterna inom Lunds universitet – en lägesbeskrivning hösten 2015: Rapport av ett projekt utfört vid Universitetsbiblioteket* [Research Data Management at the Lund University Faculties – *A Description of the Current Status, Fall 2015: Report of a Project at the Lund University Library*]. Lund University Library. Available at: http://portal.research.lu.se/portal/files/5616081/8725586.pdf (accessed 9 September 2016).

Lancaster FW (2003) *Indexing and Abstracting in Theory and Practice*. 3rd edn. Champaign, IL: University of Illinois.

Linnaeus University (2016) *Digital Humanities*. Available at: https://lnu.se/en/digihum (accessed 9 September 2016).

Lund University Libraries (2016) *Organisation*. Available at: http://www.lub.lu.se/en/about-the-library-network/organisation (accessed 9 September 2016).

McLure M, Level AV, Cranston CL, et al. (2014) Data curation: A study of researcher practices and needs. *Libraries and the Academy* 14(2): 139–164. DOI: 10.1353/pla.2014.0009.

Swedish Research Council (2011) *God forskningssed* [Research Ethics]. Available at: https://publikationer.vr.se/produkt/god-forskningssed/ (accessed 9 September 2016).

Swedish Research Council (2015) *Förslag till nationella riktlinjer för öppen tillgång till vetenskaplig information* [Proposal for National Guidelines for Open Access to Scientific Information]. Available at: https://publikationer.vr.se/produkt/forslag-till-nationella-riktlinjer-for-oppen-tillgang-till-vetenskaplig-information/ (accessed 9 September 2016).

Witt M, Carlson JD, Brandt SD, et al. (2009) Constructing data curation profiles. *International Journal of Digital Curation* 4(3): 93–103. DOI: 10.2218/ijdc.v4i3.117.

Wright SJ, Kozlowski WA, Dietrich D, et al. (2013) Using data curation profiles to design the datastar dataset registry. *D-Lib Magazine* 19(7/8): 37–49. Available at: http://www.dlib.org/dlib/july13/wright/07wright.html (accessed 9 September 2016).

Åhlfeldt J and Johnsson M (2015) *Research Libraries and Research Data Management within the Humanities and Social Sciences: Project Report*. Lund: Lund University Library. Available at: http://portal.research.lu.se/portal/files/6286782/5050466.pdf (accessed 9 September 2016).

Author biographies

Koraljka Golub is a researcher in the field of digital libraries and information retrieval. Her research has in particular focused on topics related to knowledge organization, integrating existing knowledge organization systems with social tagging and/or automated subject indexing, and evaluating resulting end-user information retrieval. She is currently running a Digital Humanities initiative at Linnaeus University, and is also exploring establishment of an iSchool at the same university. She works as an Associate Professor at the Department of Library and Information Science, School of Cultural Sciences, Linnaeus University. She also serves as an Adjunct at the School of Information Studies, Charles Sturt University, Australia. For over five years in the recent past she worked as an information science researcher at UKOLN, University of Bath, UK. Her educational background covers technology, social sciences and humanities. In 2007 she obtained her doctoral degree from the Department of Information Technology, Lund University, Sweden, on the topic of automated subject classification. Her earlier degrees are from the Department of Information Sciences and the Department of English, University of Zagreb, Croatia.

Maria Johnsson is a librarian specializing in research support services in the Section of Scholarly Communication at Lund University Library. She has a special focus on research data management and e-science, and on how libraries may develop services within research data management. Before joining the University Library she had a position at the Library of Faculty of Engineering, Lund University. She also has experience in working with library and information services at different companies. She has a Master's in Library and Information Science, combined with studies in Modern Languages.

Monica Lassi is a librarian focusing on research data management and research infrastructures, working at the Department of Scholarly Communication, Lund University Library. She also works for the research infrastructure ICOS Carbon Portal, a pan-European research infrastructure collecting and disseminating quality-controlled observational data related to greenhouse gas. She is a member of the research group Information Practices: Communication, Culture and Society at the Department of Arts and Cultural Sciences, Lund University. Before joining the Lund University Library in 2014, Monica worked as a lecturer at the Swedish School of Library and Information Science for 14 years, teaching knowledge organization, content management and interaction design. Monica's educational background includes library and information science,

informatics and language technology. She holds a PhD in Library and Information Science from the University of Gothenburg, Sweden. The title of her doctoral thesis, defended in 2014, is 'Facilitating collaboration: Exploring a socio-technical approach to the design of a collaboratory for Library and Information Science'. She also holds an MSc in Library and Information Science, which includes studies in systems architecture.

Appendix

Questions added to the original interview worksheet are indicated with an asterisk (*)

Module 6 – Organization and Description of Data
3. Please prioritize your need for the following types of services for your data.

	Not a priority	Low priority	Medium priority	High priority	I don't know or N/A
The ability to make the data accessible in multiple formats.					
The ability to apply standardized metadata from your field or discipline to the dataset.					
* The ability to apply standardized subject classification to the data set. (Please list relevant controlled vocabularies next to this table)					
* The ability for automatic suggestions of keywords for subject classification.					
* The ability to add your own tags to the data set.					
* The ability to connect the data set to the keywords of the publications that are based on it.					

Article

International Federation of Library Associations and Institutions
2016, Vol. 42(4) 278–283

DOI: 10.1177/0340035216674027
ifl.sagepub.com

'Essentials 4 Data Support': Five years' experience with data management training

Ellen Verbakel
TU Delft, Netherlands

Marjan Grootveld
Data Archiving and Networked Services (DANS), Netherlands

Abstract

This article describes a research data management course for support staff such as librarians and IT staff. The authors, who coach the participants, introduce the three course formats and describe the training in more detail. In the last years over 170 persons have participated in this training. It combines a wealth of online information with face-to-face meetings. The aim of the course is to support the participants in strengthening various skills and acquiring knowledge so they feel confident to support, advise and train researchers. Interaction among the students is embedded in the structure of the training, because we regard it as a valuable instrument to develop a professional network. Recently the course has taken on a new challenge: in addition to the regular courses a couple of in house trainings have been delivered on request. The paper ends with a description of the key group assignments for such compact trainings.

Keywords

Blended learning, data education, data literacy, data support, information skills, training library staff

Submitted: 15 May 2016; Accepted: 12 September 2016.

Introduction to 'Essentials 4 Data Support'

At the end of 2011 a 'Data Intelligence 4 Librarians' course was developed by 3TU.Datacentrum to provide online resources and training for digital preservation practitioners, specifically for library staff. The preparation, design and mission of this training is described in De Smaele et al. (2013: 218): 'The course objectives are to transfer and exchange knowledge about data management, and to provide participants with the skills required to advise researchers or research groups on efficient and effective ways of adding value to their data'.

Lessons learned during these training courses and developments in the research data management landscape have led to a revision in 2014. By then, the training had become the flagship service of Research Data Netherlands (RDNL). RDNL is a coalition in the field of long-term research data archiving, consisting of 4TU.Centre for Research Data, Data Archiving and Networked Services (DANS) and SURFsara. The goal of the training was redefined as: 'The Essentials 4 Data Support course aims to contribute to professionalization of data supporters and coordination between them. Data Supporters are people who support researchers in storing, managing, archiving and sharing their research data' (Essentials 4 Data Support, 2014). The training setup as well as all training content was thoroughly revised, as described in Grootveld and Verbakel, (2015).

As of May 2016, more than 170 data supporters from Dutch universities, Higher Vocational Education organisations, academic hospitals and other knowledge institutes have participated in the face-to-face training that Research Data Netherlands provides.

We are aware that the term 'data supporter' is unusual, but exactly this makes it inclusive: The

Corresponding author:
Ellen Verbakel, TU Delft Library, PO Box 98, NL-2600, MG Delft, Netherlands.
Email: p.m.verbakel@tudelft.nl

course appeals to (aspiring) information professionals, data stewards, data librarians, research support officers and others. The intentionally unusual term saves a lot of discussion about job titles and definitions in a highly dynamic field.

In this paper the coaches present 'Essentials 4 Data Support', one of the few courses on data management that explicitly focus on data supporters. In early 2016 Knowledge Exchange organised a survey and workshop to collect and share information on current practices around RDM training. One of the conclusions in the survey report is: 'It is striking that a near totality of respondents provides training for PhD students (. . .) librarians, data curators and IT departments/computer centres are catered for by a third or less of respondents' (Goldstein, 2016: 7).

In the next section the three training variants are introduced. This is followed by sections on the setup of the blended learning course and on the intended competencies and learning goals. After a sketch of the course website the authors describe how the regular blended learning course enables them to also accept invitations for tailor-made in-house training.

Training variants

Since the first training in 2011 the course materials have been publicly available on the website, both in English and in Dutch. The current offering consists of three variants:

- The full – blended – course consists of two face-to-face days, with fellow students, experts and coaches. In the six weeks between the first and the second day students familiarise themselves with the content of the online learning environment, do assignments, and give feedback on the assignments of their fellow students. For this they use the group's private forum. Students pay a fee for the full course.
- In the Online+ course, registered students have access to the online course materials. They are not entitled to support from coaches, but are welcome to contribute to discussions in the open forum.
- Online-Only students can take the course at their own pace, without access to social features.

The next section describes the set-up of the blended course. Recently, we have accepted a few requests for in-house training. Such training is bespoke and reduced variations on the full course and will be described in the section 'Recent developments'.

Set-up of the blended 'Essentials' course

During the six weeks of a full course the coaches and the students have a very intensive contact. The official start is on the first face-to-face day. Here the students meet their fellow students and the coaches. The coaches inform the students about the course, what will be asked from them and what the coaches will deliver. In a game setting the students discuss several aspects of data support – differences in the maturity of data management in their respective organisations start to show. Also on the agenda are the first expert presentations. The coaches have a broad network of specialists working in the field of research data management. These experts are invited to give a presentation on their daily work in the field: what problems occur, what worked well, what did not? What experiences can they share with the students? In general, the students appreciate these presentations very much; they ask many questions which feed the discussion and the subsequent assignments. Data management planning is a standard topic on the first day, to prepare the participants for their first assignment, viz. to write a data management plan.

A very important part of the course is the communication among the students and with the coaches. A private forum is provided for the group on the course website. Students have to upload their assignments and are required to give feedback on the work of their fellow students. This helps them to explore different angles and approaches. Furthermore, questions can be posted on the forum about the course and/or about the content. Because the forum is restricted to the group the students usually feel safe to do so. This can lead to very lively discussions on various aspects of data management. The coaches use the group forum too for their comments on the weekly assignments and to answer questions.

The second face-to-face day again contains one or two guest lectures, but a substantial part of this day is reserved for student presentations. One of the assignments, spanning the full six weeks, is to follow recent developments in one of the fields of research data management close to their personal learning goals and to share some findings with the group. That gives an overview of the subjects they learned about and how they approached the task – varying from using general search engines to subscribing to data and research blogs, mailing lists and twitter streams. From their colleagues' presentations on the information-gathering process and the achieved results every student learns a lot on recent topics.

The final task at the second day, called 'good intentions', aims at the immediate future: The students

must formulate data-related activities that they will carry out once they return home, e.g. design a data support workflow, discuss institutional data policies with their manager, or localise data teaching materials for young researchers. The activities are ideally formulated as small steps towards a concrete goal that can be reached in the next six months. At that time the coaches contact the former students to ask what has been achieved. The students appreciate the interest, as they recognise the purpose to keep the flow of the course going during daily work, and to really put into practice what they found inspiring.

Competencies

In the process of developing as well as during revising the course the following competencies were defined to be essential for the future data supporters. A data supporter:

- Skillfully handles ICT: Efficiently uses available information technology. In the last decades research data became digitally born, and ICT knowledge becomes even more important.
- Shows entrepreneurship: Proactive attitude to improve data services in response to changing needs in the field. Keeps an eye on trends which emerge in the profession, knows where knowledge is available (networks) and disseminates important information to key people in the organisation.
- Sees from the whole: Acknowledges that data are only part of the scientific lifecycle and is aware of the significance research data have or carrying out scientific research. Sees data and information services as part of larger whole in which decisions are made.
- Has consulting skills: Can handle questions skillfully and knows when to address a dedicated expert. Can empathise with customer perceptions.
- Has co-operative skills: Examines how collaboration with others (employees, researchers, institutions) may enhance service provision.

The DigCurV project has developed the DigCurV Curriculum Framework (DigCurV, 2013) which offers a means to identify, evaluate and plan training to meet the skill requirements of staff engaged in digital curation, both now and in the future. This framework has inspired us when we redesigned the training, and the DigCurv game is still part of the face-to-face training (see section: Setup of the blended 'Essentials' course). A more recent initiative, the Interest Group on Education and Training from the Research Data Alliance (2013) is also defining the competencies needed for handling research data. For our purposes it is important to keep in mind that RDNL intends to offer *basic* knowledge and skills to support staff; a project like EDISON (2015), for example, which is developing a curriculum for data scientists, has clearly higher ambitions.

Learning goals

The name for the course – 'Essentials 4 Data Support' – refers to the main goal of the course: teaching the basic knowledge and skills (essentials) to enable a data supporter to take the first steps towards supporting researchers in storing, managing, archiving and sharing their research data. Some learning goals are:

To get an overview of the data supporter's field of work, including the research life cycle and the organisation of data support, e.g. in the form of the RDNL front office – back office model (Dillo and Doorn, 2014):

- To understand the different parts of a data management plan;
- To know about the various ways to store, backup, organize and document research data;
- To know types of archives, data publication and data citation;
- To advise researchers in balancing legislation and practice;
- To be able to engage in a discussion with researchers.

The complete set of learning objectives of the six chapters can be found in (Grootveld and Verbakel, 2015).

An implicit goal, which for us has been dominant from the start, however, is that participants gain self-confidence in discussing research data with researchers. The training has been developed to encourage them in their professional dealings, by providing the knowledge to be a partner to the researcher. Strictly speaking this concerns the participants' attitude rather than their knowledge or skill set.

The 'Essentials' website

In 2014 the website was redesigned by an external developer. The platform for the course is T3Elearning, a Learning Content Management Framework based on Typo3 CMS. The content has been designed and developed with valuable input from subject matter experts within and outside RDNL,

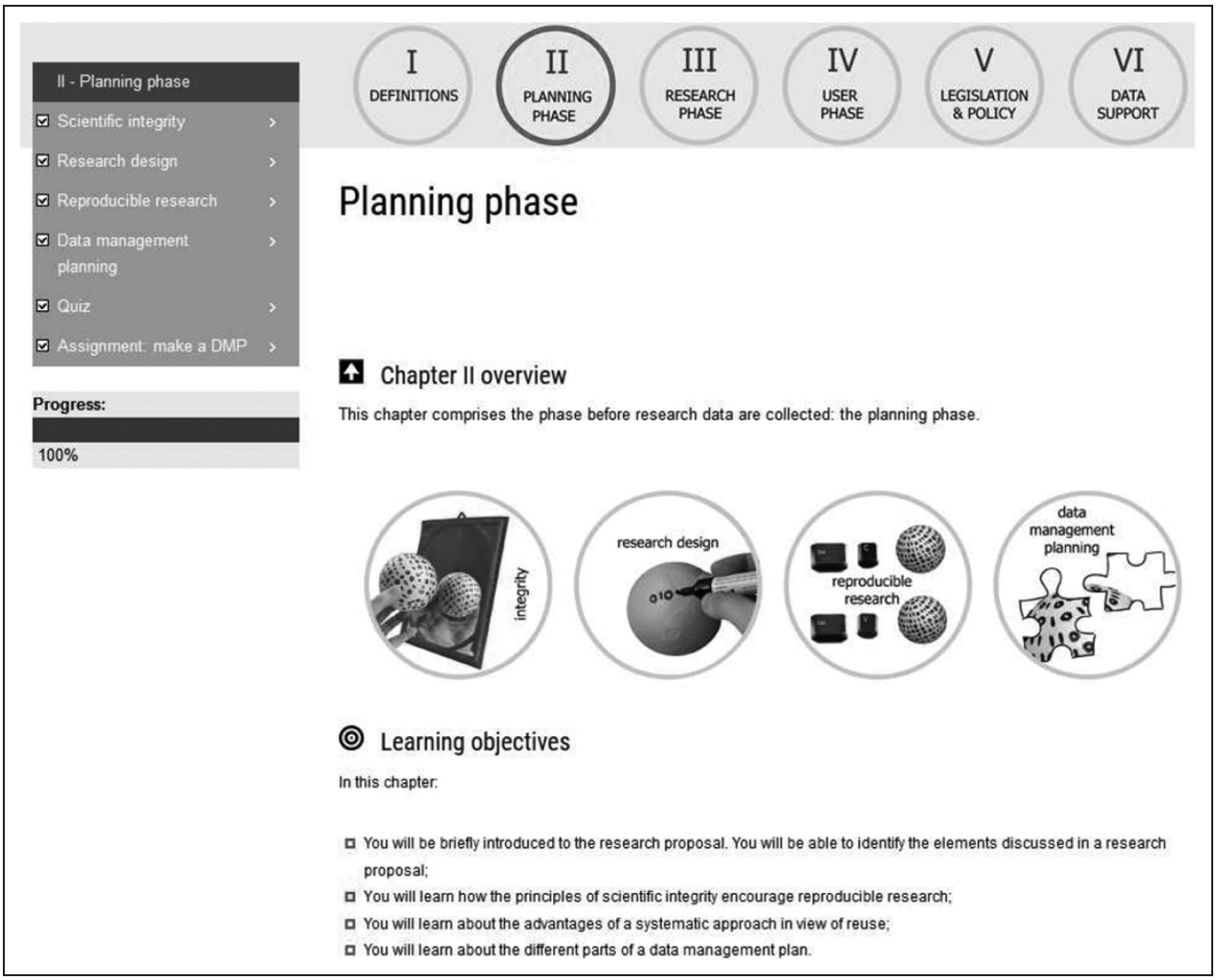

Figure 1. User interface of Essentials 4 Data Support, Chapter II.

of whom several also give a guest lecture at one of the training days.

The website changed into a more user-friendly design. This was achieved by including more images and some videos with brief talks by researchers and 'how to' videos about data management planning, data sharing concerns, reproducible research, data citation and persistent identifiers. Also new in the revised design are the end of chapter quizzes, which allow for a quick recapitulation of the chapter's contents.

Whereas the initial training featured chapters like 'Technical skills' and 'Advisory skills' of the data supporters, in the current version the research cycle is leading, rather than the activities of the data supporters. This can be seen at the top of Figure 1, where the planning phase precedes the research phase, as well as the phase in which researchers make their data available for reuse. Each chapter starts with a short outline of the sections and learning objectives of that particular chapter. In addition to RDNL's own content (available under a CC-BY-SA licence) all chapters contain a large amount of external links to good practices, definitions and reading materials provided by other organisations. This makes the site a valuable work of reference – the video clip about data management planning has been viewed 4500 times – and the students rate the website highly in their evaluations. The assignments and quizzes in the left-hand menu in Figure 1 are only accessible to participants who have enrolled in the full course. The private forum, which these students and the coaches use between the face-to-face meetings for assignments and discussion, is part of this website and remains accessible afterwards for future knowledge sharing.

Recent developments

In the previous year we initiated two new developments, which we describe in this section. The first initiative served to strengthen the contacts with and among former course participants, while the second one can be seen as a spin-off of the regular blended training.

Last autumn the coaches organised a reunion for all participants from 'Data Intelligence 4 Librarians' and 'Essentials 4 Data Support'. The response was quite good: About 20% of the students, from both course iterations, accepted the invitation to discuss their daily work in RDM. They were surprised they had so much in common, although they represented very different institutions. The key speakers at the meeting were former students, who now have become specialists, very able and willing to share their experiences. With regard to our goal to increase the data supporters' professional self-confidence, this meeting convinced us of the longer-term value of our training offer, in a way that even positive evaluations at the end of a training day cannot do.

The second development concerns in-house training, on request. In earlier days RDNL had received some requests about in house training for (prospective) front office staff, but decided against it because we value the networking possibilities of a mixed audience even higher than the intended team-building effect. However, two recent requests did concern mixed audiences from the start. First, a Dutch institute for higher vocational education was looking for RDM training for support staff and researchers, and eventually the institute's policy department was involved as well. Second, the Danish Forum for Data Management, which consists of all Danish universities, two national libraries and the national archive, was looking for a train-the-trainer course. The process in these cases was similar: the customer provided information about the training needs of the group and any wishes regarding the duration and intensity of the training, e.g. with or without assignments ahead of the training day or days. On that basis the coaches drafted a proposal in which they included the relevant parts from the regular face-to-face training, or suggested to develop new modules. Clearly, there cannot be a standard fee for a tailor-made training.

In comparison to the off-the-shelf training, the coaches now became trainers: without guest presenters, it was up to them to convey the contents of the course, in addition to coaching the participants in the practical exercises. Furthermore, for these exercises relatively more time was planned than in the regular training, because all networking interaction and knowledge transfer between the participants has to take place during the meeting. The combination of the customer's requirements, the broad but not always specialist knowledge of the coaches/trainers and the time needed for interaction has led, in both cases, to an abridged and more focused version of the regular course.

In both in-house training courses we have included exercises with writing a data management plan (DMP) and with identifying the stakeholders in data management. With regards to the former, participants have to draft a DMP for a given, fictitious project. This is an assignment that students do in pairs (in the regular training) or in small groups (in abridged trainings), because it is relatively demanding. However, it is usually the most appreciated assignment: all participants know that research funders and research institutes increasingly require DMPs, but they seldom have any practical experience with them at the start of the training. Having written a DMP themselves, they feel more comfortable at the prospect of advising researchers. In the stakeholder exercise students have to chart all staff roles, departments and organisations which also play a role in data management, and to reflect on the quality of all communication and collaboration: 'Do I know who to contact when . . . and how do I make sure that department XYZ informs me about . . . ?' We regard both exercises as crucial stepping stones for exploring the diversity of research data management, and this has worked out well in both in-house training courses. The two in-house courses also featured an item on storing and archiving research data, and for the Danish train-the-trainer course we designed an exercise to analyse and prepare an institutional implementation of the recent Danish Code of Conduct for Research Integrity.

Conclusion

RDNL's training 'Essentials 4 Data Support' results from a clear need for professional data management support training, a need that was identified five years ago. Since then Research Data Netherlands has trained about 170 so-called data supporters, working in research libraries, IT departments, medical centres and elsewhere. As front office staff they play an essential role in research data advocacy and support for gradually realising the international Open Science ambitions. We are very pleased to see a vibrant network of data supporters in The Netherlands and abroad, and to be recognised as a party that contributes to their professional development.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

De Smaele M et al. (2013) Data intelligence training for library staff. *International Journal of Digital Curation* 8(1): 218–228. DOI: 10.2218/ijdc.v8i1.255.

DigCurV (2013) *A Curriculum Framework for Digital Curation*. Available at: http://www.digcurv.gla.ac.uk/ (accessed 22 September 2016).

Dillo I and Doorn P (2014) The front office-back office model: Supporting research data management in the Netherlands. *International Journal of Digital Curation* 9(2): 39–46. DOI: 10.2218/ijdc.v9i2.333.

EDISON (2015) *EDISON: Building the Data Science Profession*. Available at: http://edison-project.eu/ (accessed 22 September 2016).

Essentials 4 Data Support (2014) *Essentials 4 Data Support Course*. Available at: http://datasupport.researchdata.nl/en/ (accessed 22 September 2016).

Goldstein S (2016) *Training for Research Data Management: Comparative European Approaches*. Report, Knowledge Exchange, May 2016. DOI: 10.5281/zenodo.50068.

Grootveld M and Verbakel E (2015) Essentials for data support: Training the front office. *International Journal of Digital Curation* 10(1): 240–248. DOI: 10.2218/ijdc.v10i1.364.

Research Data Alliance (2013) *Interest Group: Education and Training on Handling of Research Data*. Available at: https://rd-alliance.org/groups/education-and-training-handling-research-data.html (accessed 22 September 2016).

Author biographies

Ellen Verbakel is a trained librarian, but moved to 4TU.Centre for Research Data where she is now working as a data librarian. She is situated in the TU Delft Library. She co-developed the course 'Data Intelligence 4 Librarians' and was involved in redesigning the course into 'Essentials 4 Data Support'. Being a coach at the course is a very important part of her daily work.

Marjan Grootveld is senior policy consultant at Data Archiving and Networked Services (DANS). She advises knowledge institutes and research funders on data management policy and practice, both in international research projects and by coaching participants of the Research Data Netherlands training 'Essentials 4 Data Support'.

Article

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 284–291

DOI: 10.1177/0340035216674971
ifl.sagepub.com

Research Data Services at ETH-Bibliothek

Ana Sesartic
ETH-Bibliothek, ETH Zürich, Switzerland

Matthias Töwe
ETH-Bibliothek, ETH Zürich, Switzerland

Abstract

The management of research data throughout its life-cycle is both a key prerequisite for effective data sharing and efficient long-term preservation of data. This article summarizes the data services and the overall approach to data management as currently practised at ETH-Bibliothek, the main library of ETH Zürich, the largest technical university in Switzerland. The services offered by service providers within ETH Zürich cover the entirety of the data life-cycle. The library provides support regarding conceptual questions, offers training and services concerning data publication and long-term preservation. As research data management continues to play a steadily more prominent part in both the requirements of researchers and funders as well as curricula and good scientific practice, ETH-Bibliothek is establishing close collaborations with researchers, in order to promote a mutual learning process and tackle new challenges.

Keywords

Data life-cycle, data management plan, libraries, preservation, research data, research data management

Submitted: 12 May 2016; Accepted: 8 September 2016.

Introduction

The growing volume of data produced in research has created new challenges for its management and curation to ensure continuity, transparency and accountability. Timely and effective management of research data throughout its life-cycle ensures its long-term value and prevents data from falling into digital obsolescence (Corti et al., 2014; Goodman et al., 2014).

Proper data management is a key prerequisite for effective data sharing within a specific scientific community and for data publication beyond any particular target group. This, in turn, increases the visibility of scholarly work and is likely to increase citation rates (Piwowar and Vision, 2013: 25; Piwowar et al., 2007). Managing research data throughout its life-cycle is not only a key prerequisite for effective data sharing but also for efficient long-term preservation because the latter must rely on technical, administrative and rights metadata, as well as sufficient context information being available to make sure that data remains usable and understandable in the long run.

Depending on their respective institutional setting, libraries can contribute to research data management in different ways and likewise, expectations from their patrons can vary widely, e.g. by scientific discipline. Therefore, the following report should be understood as a case study rather than a general recommendation.

Libraries are never the only service providers in a university and, ideally, a range of providers caters for researchers' and students' needs. When it comes to data management in particular, IT services will obviously be a strong player on the technical side whereas research offices must take an interest in how researchers comply with internal and external requirements. In such a landscape, it is important to note that libraries should focus on their strengths such as metadata management, content curation, and support and training of their customers, in this case the researchers. Also, the services offered are never carved in stone and should be adapted to the current needs of science.

Corresponding author:
Ana Sesartic, ETH Zürich, ETH-Bibliothek, Rämistrasse 101, CH-8092 Switzerland.
Email: ana.sesartic@library.ethz.ch

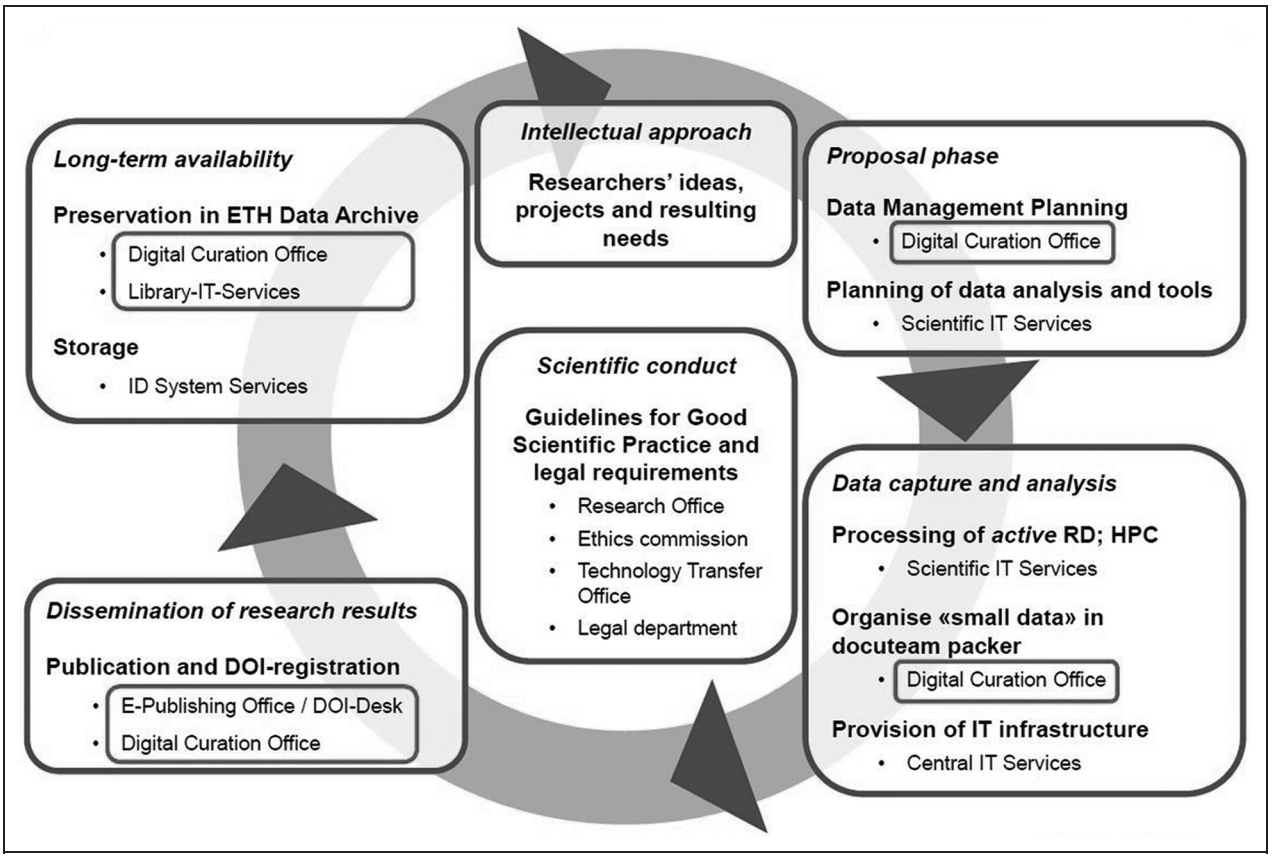

Figure 1. Actors and their tasks at ETH Zürich along phases of the data life-cycle.
Units in red boxes belong to ETH Library.

In the following, we summarize the data services and approach to data management as currently practised at ETH-Bibliothek (ETH Zürich, 2016c), the main library of ETH Zürich (ETH Zürich, 2016b), the largest technical university in Switzerland.

The overall concept

The role of the ETH-Bibliothek within ETH Zürich is to support researchers from early on in the data life-cycle (see Figure 1), starting with general consulting regarding compliance, support through the development of data management plans, the offer of data archiving and publication services, to the creation of DOIs for easier citation and re-use of data. Important phases of the life-cycle mainly in its active research phase (see Figure 1) require competencies beyond what the library can offer. Other actors within the university support these phases and a clarification of the division of tasks between those service providers is needed. At ETH Zürich, a good level of mutual understanding, for example of the activities in storage and preservation, was achieved between library staff and the storage section of the IT services as part of a common small-scale project starting in 2006. Regular meetings have continued ever since. Nevertheless, it can be challenging to communicate the division of labour transparently to customers. Ideally, there should be only a single point of contact for customers to turn to. This has not been established so far. However, it should not matter who customers contact as long as they are redirected appropriately. This, again, requires a reasonable understanding of other units' services, which obviously evolve over time.

With the Digital Curation Office (ETH-Bibliothek, 2016a), the ETH-Bibliothek offers a point of contact for technical and conceptual questions regarding long-term preservation and management of research data. It also offers help and support to the researchers of ETH Zürich in managing and publishing their data, as well as following the requirements as stated in the *Guidelines for Research Integrity* of ETH Zürich (ETH Zürich, 2011: 31). Regarding intellectual property and research ethics issues, the ETH-Bibliothek closely collaborates with the Technology Transfer Office and the Office of Research, respectively.

Figure 1 illustrates the coverage of roughly defined stages of the research data life-cycle by actors and services at ETH Zürich with their respective tasks. Note that Guidelines for Good Scientific Practice and further legal requirements (centre) apply throughout the cycle, but should obviously be considered from early on.

For a first-time or occasional customer, the multitude of actors can obviously be confusing and users should not be left alone with it. From our experience, customers appreciate being able to get in touch with a contact person they can talk to about their needs. In a first instance, it is not even expected that this person can provide an immediate solution, but it is important that someone takes care of the issue and provides guidance as to where to turn. This means that all actors must be aware of tasks not within their own portfolio as well, and we are aware that this remains a challenge and requires an ongoing learning process. As a detail to underline the importance of personal contacts: as one of the first teams within the ETH-Library, the Digital Curation Office put portrait images of its staff on its website to lower the barrier of writing to an otherwise anonymous email box.

Data management training

Research data management (RDM) has gained increasing attention over the last years, due to growing awareness of the value contained in research data and of the risks of losing such data over time. Apart from the need to manage data over the course of a project, there is a need to retain and curate data which one plans to work with in the future. It might be possible and sometimes advisable to repeat an experimental measurement, but in other cases it might either be prohibitively expensive or otherwise ethically unacceptable to do so. For unique observational data, a repetition is not possible at all and accordingly, communities relying on such data have long been aware of the challenges.

While these issues are mainly related to the efficiency and effectiveness of research and its funding, RDM also addresses the accountability of science. One principle of the scientific process is the requirement to be able to justify results by providing underlying data where necessary. There are significant reputational risks involved for individual researchers, principal investigators and institutions who fail to comply with good scientific practice of which RDM is just one part.

To address these questions a workshop within the ETH Critical Thinking Annual Programme (ETH Zürich, 2016e) was developed. The overall aim of this programme is to strengthen critical thinking and a responsible approach to taking actions beyond disciplinary competences. The workshop introduced some services and tools for RDM, as well as encouraging the participants to share both their experiences and the methods and tools they use, during the interactive parts of the workshop. The goal of the workshop within the Critical Thinking programme was to increase the critical thinking skills of undergraduate and graduate students, as well as scientists, regarding RDM. The philosophy behind this approach is the conviction that researchers themselves must be empowered to make informed decisions on their data, as they are the experts with the most intimate knowledge of their own data. The workshop was focused on activating teaching methods, engaging the participants in group work and discussions.

The majority of the participants at this event were working on their doctoral studies with few Master's students and post-docs present. They showed very varied needs and levels of knowledge, but they were well aware of the problems regarding RDM beforehand and were looking for solutions. This is why we also offer tailored training courses in RDM for groups and departments. These can range from so-called Tools & Tricks mini lectures, short 15-minutes inputs over coffee for lunch break, to fully-fledged one-day training workshops. It is important to note that certain departments and institutes already offer similar or overlapping internal training, which is why communication and coordination are key. As of now, there is no dedicated course on RDM within the ETH curriculum; however, some departments offer methodological courses including research ethics and scientific writing which might cover some aspects. With increasing concerns about data management issues, it is to be expected that the topic will figure more prominently in curricula in the future.

Data management plan checklist

Today, the availability of well-managed data is part of good scientific practice and ensures the reproducibility of research results, a key requirement at the core of the research process. Many funding organizations prescribe the use of data management plans and insist on open access publication of the research results they funded.

In some parts of EU's research programme Horizon 2020, DMPs will, for example, be evaluated as part of the impact of a proposal and in the reporting during the course of a funded project (European Commission, 2013: 6). But even if a funding body does not explicitly demand data management, following professional curation and preservation concepts has numerous advantages (DLCM, 2016):

- It greatly facilitates the reuse of research data;
- As a result, this increases the impact of research results;
- It saves precious research funds and ultimately natural and human resources by avoiding unnecessary duplication of work.

As effective and efficient data management becomes more and more challenging for both researchers and information specialists, the question arose how they can best be reached and supported on a national level regarding best innovative practices in digital preservation so as to succeed in this ambitious enterprise. This led to the creation of the Swiss Data Life-Cycle Management (DLCM) project (Blumer and Burgi, 2015: 16), facilitated by *swissuniversities* (swissuniversities, 2016) and involving collaboration between eight Swiss higher education institutions (EPF Lausanne, ETH Zürich, Universities of Basel, Geneva (lead), Lausanne, Zürich, Geneva School of Business Administration at the Western Switzerland University of Applied Sciences and Arts, and SWITCH, the national IT service provider for higher education institutions).

The Data Management Plan (DMP) Checklist (ETH-Bibliothek and EPFL Library, 2016) is among the first tangible deliverables of the DLCM project. It is meant to be an essential tool aiding researchers in the management of their data, thus preparing them for later publication and preservation, as needed. By giving clear guidelines, it should facilitate this task for researchers and eventually save them time and effort. The list has been customized for Switzerland based on pre-existing national and international policies. It covers general planning and the phases of the data life-cycle, from data collection and creation to data sharing and long-term management. Special sections cover documentation and metadata, file formats, storage, ethical and intellectual property issues. Ideally, the list should be used by researchers to critically assess their data management and to gather information they might need to create a data management plan. It can also serve as a starting point for further face-to-face discussions of data management issues within research groups and with support staff if required. The list is static with no further functionality and it will be observed whether a more interactive solution will be required later.

The checklist was created in close collaboration between the Digital Curation Office at ETH-Bibliothek and the Research Data Team at EPFL Library. It is currently available through the ETH-Bibliothek and EPFL Library websites (ETH-Bibliothek and EPFL Library, 2016) and will soon be disseminated on a national portal on DLCM which will be launched in the next few months. The portal itself will touch on many aspects, aggregating and providing further information about data organization, training (in person and online) of end-users, and consulting regarding best practices among many others.

Active data management

Data management is understood as a comprehensive task throughout the data life-cycle. It therefore needs to comprise the handling of research data while the actual research is carried out. We call this active data management to signify that data at this stage is usually not static, but keeps being analysed and worked upon as part of the research. At this stage of the life-cycle, subject-specific tools are employed in data processing which may be implemented and run by a specialized support unit or by research groups themselves.

At ETH Zürich, the section Scientific IT Services of the central IT Services provides this kind of support. Research groups from life sciences are known to be among the most intensive users of their services. Among them figures the data management software platform openBIS (ETH Zürich, 2016d), which has also been extended with components to serve as an electronic laboratory notebook (ELN) and/or as a laboratory information management system (LIMS).

Obviously, this platform and other tools with similar aims must already capture a lot of information, which is relevant for current research. Part of this information might be gathered automatically, while more will be required to be entered by researchers, with quality depending on their willingness to comply and therefore on the ease of the process. Most of this input and possibly additional documentation will be needed in order to use and make sense of the research data at a much later stage and with the active data platform no longer being available.

While such systems themselves are not meant as publishing or preservation tools for research data, they can very well serve as sources for these processes. Ideally, researchers working in such a system should be able to decide which part of the content must be preserved or can be published to trigger an export, e.g. to a long-term preservation solution. Such a process has not yet been implemented, but it is envisaged that an interface from openBIS to the ETH Data Archive will be developed in the current DLCM project.

It should be noted that the exchange between the different systems must not be understood merely as a technical transfer. Data at this stage also crosses a boundary between the research world in a narrow sense and a curation domain (see Figure 2). Such a transition is prone to suffer from misunderstandings between the parties involved due to a lack of common standards or even a uniform vocabulary.

This already highlights the importance of close collaboration not only of researchers and service providers, but also among different providers with

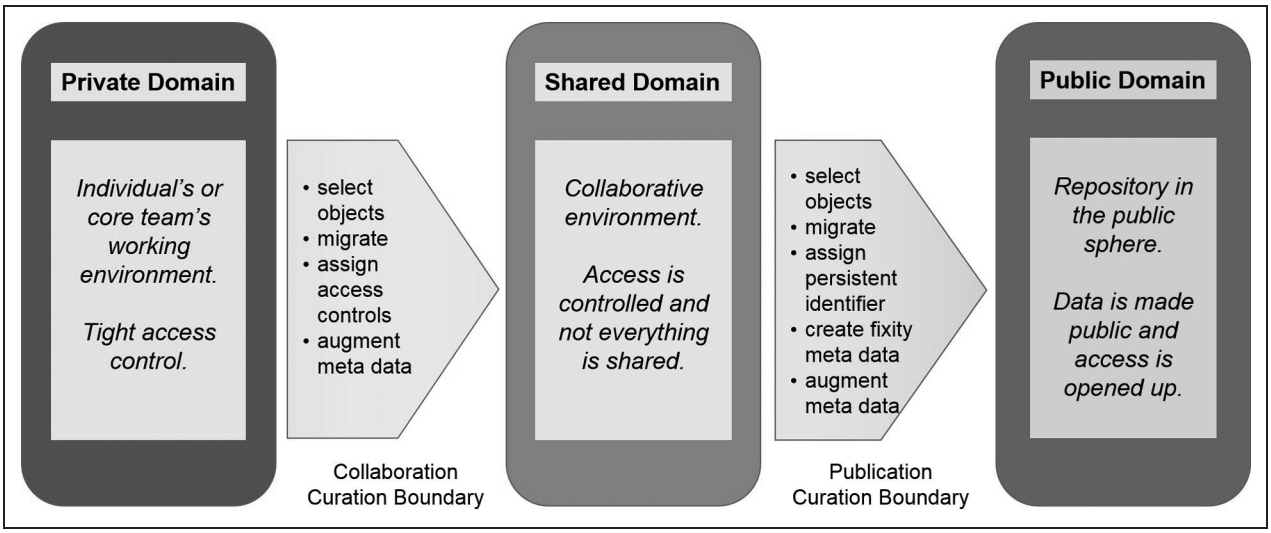

Figure 2. Transitions between curation domains along the research data life-cycle. Adapted from (Treloar, 2012).

complementary competencies. An essential part of the work to achieve this has been a constructive exchange over several years between ETH Library and the Central IT Services. Even before the Digital Curation Team existed, staff from various teams in the library (e.g. University Archives of ETH Zürich, Library IT Services, Consortium of Swiss Academic Libraries) engaged in early pilot projects mainly with the Central IT Services team in charge of storage management. These activities from 2006 onwards served to understand the required functions, to get an idea of the available competencies and not the least to build trust within an, albeit loose, network of players in the field of data management within the university. This helped to achieve a common understanding of the tasks and to raise awareness, e.g. for the need to define more precisely what each stakeholder understands when talking about 'archiving'.

Two rather surprising outcomes arose from the exchange with a retiring professor on how to transfer his well-managed archive to the library. Firstly, one of his post-docs involved in the operation could be hired to work with ETH Library's Digital Curation Team, and secondly, the discussion of the principles underlying this research group's archive recently led to a publication highlighting the general concepts behind their approach which had proven useful for almost three decades of research practice (Sesartić et al., 2016).

Publication and preservation of data

From the start of its activities around research data, ETH-Bibliothek saw a central role in support of the processes for publication and preservation of such data. This was in line with expectations from both researchers and other service providers in the university. At the same time, it was decided to address preservation issues with a view on all kinds of scientific and cultural heritage unique to ETH Zürich. This includes research data from ETH Zürich staff, administrative records or personal bequests to ETH Zürich University Archives (ETH-Bibliothek, 2016c) and digital born and digitized content of ETH-Bibliothek, in particular doctoral theses which must be deposited in digital form and master files from several large scale digitization projects.

Since 2014, ETH-Bibliothek has offered the ETH Data Archive as a productive service. It is based on the commercial long-term preservation system Rosetta (Ex-Libris Group, 2016b). The application itself is maintained by the library's own IT services, while both virtual servers and different types of storage at two different sites of the university are provided by ETH Zürich's central IT services.

ETH researchers can deposit content into the ETH Data Archive and define the appropriate access rights. This can be done manually via a web client, e.g. for supplementary material belonging to an article. Depositors are quite free to define on which level they want to or can provide metadata: an archival package or intellectual entity may contain just a single file with its individual description or a ZIP- or TAR-container with thousands of files included under only one metadata set. The latter is sometimes made use of when a collection of files belonging to a concluded thesis or another publication needs to be retained for a defined period of time (10 years minimum). In this case, it is often assumed that the thesis itself is the most comprehensive documentation of the data

package and should suffice for the professor to answer any inquiries. This might not be an ideal arrangement, but it represents considerable progress compared to some previously existing group archives held exclusively on CD-ROM and DVD. When the focus is on safeguarding data for a limited period of time only, no strict requirements on the longevity of file formats are enforced. If requested by researchers, support is provided with identifying suitable formats.

In other cases, metadata is routinely gathered for research data on the level of individual folders and files in research groups who have already seen a need to operate a managed archive. They may use the open source editor and viewer docuteam packer (docuteam, 2016; ETH-Bibliothek, 2016e) locally to organize files in a defined structure and add metadata as required. To support this potentially laborious task, certain metadata can be pre-defined or inherited from the top-most level. A researcher can then decide when to submit the whole or part of the structure she or he created to the ETH Data Archive. For data in docuteam packer, DOIs (Digital Object Identifiers) can already be reserved which will be registered to become active after submission to the ETH Data Archive. The advantage of this method is that the DOI can already be used in a manuscript even before the data has actually been deposited. DOIs are registered with the consortium DataCite (2016). Likewise, access rights and a retention period can be defined which will be enforced after submission.

Submission itself is handled by the tool docuteam feeder, which processes the Submission Information Package (SIP) put out by docuteam packer into the according Archival Information Packages (AIP) to be submitted to the ETH Data Archive. Docuteam packer will also be used for submissions to ETH Zürich University Archives with a modified configuration and a more interactive workflow between depositors and University Archives staff.

Obviously, there are limits to which kind of submissions can be comfortably handled via a web user interface and via docuteam packer. Automated processes with submission applications relying on existing sources for metadata and content are currently in use only for library content from the institutional repository ETH E-Collection, for master files from the digitization of rare books from ETH-Bibliothek in e-rara (ETH-Bibliothek, 2016b), and for digitized material from ETH Zürich's Archives of Contemporary History (ETH Zürich, 2016a). In the future, such interfaces should also be created with existing sources for research data, such as the platform openBIS mentioned above. Apart from fully automated processes, there might still be others requiring a trigger from the data producer to give them more control of which part of their content is archived when. For example, producers might want to collect, (re-)structure and describe their data before finally deciding to submit an archival package. In other cases, professors want to review datasets from their groups before submission and start the transfer themselves.

Deposit in the ETH Data Archive is not coupled with an immediate publication of the content. While ETH Zürich encourages open access also to research data, it is currently up to the data producer to decide which data they want to make accessible as long as they are observing existing requirements, e.g. from funders. They may opt for an embargo period of e.g. two years or for limited access within the IP-range of ETH Zürich only. Even very restrictive access rights for defined persons only are currently accepted.

Metadata of published content are published in the ETH Knowledge Portal (ETH Zürich, 2016c), in the Primo Central Index (Ex-Libris Group, 2016a) and in the Data Citation Index (Thomson Reuters, 2016).

Currently, the workflow for the deposit of research data is largely separated from the one for publications, e.g. articles to be published via the green road of open access. In the future, a new platform will pull several workflows for publications, bibliographic records and research data together into one service.

For published research data in particular, it is a reasonable expectation that it should remain available in the long term similar to what is expected from formal publications. Whether this will actually be possible in the future depends to a large degree on the file formats being employed. In the heterogeneous environment of a university with numerous contradictory pressures and constraints on researchers, it would not be a realistic approach to admit only a limited number of well-documented open formats to a data archive. Rather, the Digital Curation Office at ETH-Bibliothek offers some guidance on preferred formats and recommends a few of them depending on the expected retention period (ETH-Bibliothek, 2016d). This may limit the chances of actual preservation measures in the future to the extent where only bitstream preservation remains as viable. This is made transparent to researchers submitting data and usually they are fully aware of this serious limitation. However, those depositing data 'just in case' do not consider this as a major problem because they expect a need to invest effort into using such data in the future, anyway. Their perspective then is to postpone this effort to the point where it is actually needed rather than making an upfront investment which might be in vain, given that only a very small part of their data might ever be re-used.

On the other hand, data producers who regularly share and exchange data with colleagues in their own community have usually overcome the barrier of proprietary formats and often rely on open community standards. However, this might only apply to the core of research data, while accompanying material may contain less suitable formats. Given the high level of awareness of these users, they might want to re-consider those formats, as well.

A particular format issue concerns the vast number of research data files in plain text formats, but with a large variety of sometimes misleading file extensions. While text files with documented encoding are actually very suitable for preservation purposes, it can be challenging to identify them in the first place and in many cases, the identification will not be a technical one, but will rely on information from the data producers.

Taking the sections above into account, it is obvious that communication with researchers forms an essential part of providing research data services for publication and preservation. It is in the best cases rewarding for customers and staff alike, but nevertheless time consuming when time might already be pressing, for example when a manuscript is about to be submitted and supplementary material needs to be deposited on time. Obviously, this kind of communication is much facilitated if appropriate skills are available on both sides. It is therefore very helpful to have staff with a scientific background in the Digital Curation Office, although it is obvious that they cannot cover all fields in depth.

Conclusions

With the growing digitization of science and society, a curation gap between research practice and curation needs opens up. However, if there is collaboration and communication between IT services, libraries and researchers, the discrepancy between research practice and research content preservation can be minimized and the curation gap closed. In order to do so, university libraries and data centres must continue to support and educate researchers, which also requires a thorough understanding of researchers' work practices and the challenges they meet. The heterogeneity of their needs limits the possibility to generalize services – or the other way round: in some cases, it may only be possible to serve needs close to the smallest common denominator between various interests. This is the reason why libraries should not aim at serving all communities equally themselves, but rather also keep an eye on subject specific solutions, which are created by third parties to address specific needs of one discipline. A combined and well-integrated landscape of institutional, networked and subject-specific approaches might then cover most needs over time.

While libraries need to build on their strengths to become an active part of the overall landscape for RDM in a university, they must also consider that the services offered are not set in stone but have to be adapted and developed continuously. As scientific practice evolves rapidly, a constant learning and innovation process is needed to keep up with changing requirements. Libraries – and other service providers – need to open up and reach out further towards researchers and collaborate more closely with the researchers, in order to establish a mutual learning process on the part of both libraries and researchers. Requirements of researchers will not only evolve through technical and scientific developments in their field of research, but it can also be expected that RDM will play a more prominent part in curricula and in good scientific practice in the years to come. Constant efforts in information and training on RDM should help to further implement it as an essential task of researchers with each new generation of, for example, doctoral students. A top-down commitment from universities can certainly support this, provided it is appropriately translated into activities, which really reach researchers at their workplace.

Acknowledgments

The authors would like to thank all members of the Swiss DLCM project for insightful discussions. Special thanks go to the Research Data Team at EPFL Library for the excellent collaboration on the Data Management Checklist.

Declaration of Conflicting Interests

The authors are employees of ETH Zürich and participate in the DLCM project.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article beyond their employment by ETH Zürich.

References

Blumer E and Burgi P-Y (2015) Data Life-Cycle Management Project : SUC P2 2015-2018. *Revue électronique suisse de science de l'information* 16: 1–17. Available at: http://archive-ouverte.unige.ch/unige:79346 (accessed 5 October 2016).

Corti L, Van den Eynden V, Bishop L et al. (eds) (2014) *Managing and Sharing Research Data*. London: SAGE. Available at: http://ukdataservice.ac.uk/manage-data/handbook/ (accessed 7 October 2016).

DataCite (2016) *DataCite*. Available at: http://www.datacite.org (accessed 6 May 2016).

DLCM (2016) *Data Management Checklist*. Available at: http://www.dlcm.ch/ressources/data-management-checklist (accessed 22 September 2016).

docuteam (2016) *docuteam packer*. Available at: http://www.docuteam.ch/en/products/it-for-archives/software (accessed 10 May 2016).

ETH Zürich (2011) *Guidelines for Research Integrity*. Zürich: ETH Zürich. Available at: www.vpf.ethz.ch/services/researchethics/Broschure.pdf (accessed 7 October 2016).

ETH Zürich (2016a) *ETH Archives for Contemporary History*. Available at: http://www.afz.ethz.ch (accessed 22 September 2016).

ETH Zürich (2016b) *ETH Zürich*. Available at: http://www.ethz.ch/en.html (accessed 11 May 2016).

ETH Zürich (2016c) *ETH-Bibliothek*. Available at: http://www.library.ethz.ch/en/ (accessed 11 May 2016).

ETH Zürich (2016d) *OpenBIS*. Available at: http://openbis-eln-lims.ethz.ch (accessed 2 September 2016).

ETH Zürich (2016e) *Critical Thinking Initiative*. Available at: https://www.ethz.ch/services/en/teaching/critical-thinking-initiative/ct-annual-programme.html (accessed 22 September 2016).

ETH-Bibliothek (2016a) *Digital Curation Office*. Available at: http://www.library.ethz.ch/Digital-Curation (accessed 11 May 2016).

ETH-Bibliothek (2016b) *E-Rara*. Available at: http://www.e-rara.ch/?lang=en (accessed 6 May 2016).

ETH-Bibliothek (2016c) *ETH Zürich University Archives*. Available at: http://www.library.ethz.ch/en/Resources/Archival-holdings-documentations/ETH-Zurich-University-Archives (accessed 11 May 2016).

ETH-Bibliothek (2016d) *File Formats for Archiving*. Available at: http://www.library.ethz.ch/en/Media/Files/File-formats-for-archiving (accessed 11 May 2016).

ETH-Bibliothek (2016e) *Intended Purpose of Docuteam Packer*. Available at: http://www.library.ethz.ch/en/Media/Files/Intended-purpose-of-docuteam-packer (accessed 10 May 2016).

ETH-Bibliothek and EPFL Library (2016) *Data Management Checklist*. Available at: http://bit.ly/rdmchecklist (accessed 22 September 2016).

European Commission (2013) *Guidelines on Data Management in Horizon 2020*. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 22 September 2016).

Ex-Libris Group (2016a) *Primo Central Index*. Available at: http://www.exlibrisgroup.com/category/PrimoCentral (accessed 6 May 2016).

Ex-Libris Group (2016b) *Rosetta*. Available at: http://knowledge.exlibrisgroup.com/Rosetta (accessed 10 May 2016).

Goodman A, Pepe A, Blocker AW et al. (2014) Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*. 10(4): e1003542.

Piwowar HA and Vision TJ (2013) Data reuse and the open data citation advantage. *PeerJ* 1: e175 Available at: https://doi.org/10.7717/peerj.175 (accessed 5 October 2016).

Piwowar HA, Day RS and Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3). Available at: http://dx.doi.org/10.1371/journal.pone.0000308 (accessed 5 October 2016).

Sesartić A, Fischlin A and Töwe M (2016) Towards narrowing the curation gap: Theoretical considerations and lessons learned from decades of practice. *ISPRS International Journal of Geo-Information* 5(6): 91. DOI: 10.3390/ijgi5060091.

swissuniversities (2016) *Rectors' Conference of Swiss Higher Education Institutions*. Available at: http://www.swissuniversities.ch/en/ (accessed 12 May 2016).

Thomson Reuters (2016) Data Citation Index. *Web of Science*. Available at: http://wokinfo.com/products_tools/multidisciplinary/dci/ (accessed 6 May 2016).

Treloar A (2012) *Private Research, Shared Research, Publication, and the Boundary Transitions*. Available at: http://andrew.treloar.net/research/diagrams/data_curation_continuum.pdf (accessed 22 September 2016).

Author biographies

Ana Sesartic is an environmental scientist with a doctoral degree in Atmospheric Physics. She brings hands-on experience in working with and managing big data sets, and is now bridging the gap between scientists and librarians at the Digital Curation Office. Currently she is working on the Data Life-Cycle Management (DLCM) project.

Matthias Töwe is a chemist with a doctoral degree in Experimental Physics and a trained scientific librarian. Since 2003 he has been working at ETH-Bibliothek in different roles, first for the Consortium of Swiss Academic Libraries and later for the Swiss electronic library: e-lib.ch. Since late 2010 he has been head of the Digital Curation Office of ETH Zürich at ETH-Bibliothek.

Article

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 292–302

DOI: 10.1177/0340035216672870
ifl.sagepub.com

Beyond the matrix: Repository services for qualitative data

Sebastian Karcher
Syracuse University, USA

Dessislava Kirilova
Syracuse University, USA

Nicholas Weber
University of Washington, USA

Abstract

The Qualitative Data Repository (QDR) provides infrastructure and guidance for the sharing and reuse of digital data used in qualitative and multi-method social inquiry. In this paper we describe some of the repository's early experiences providing services developed specifically for the curation of qualitative research data. We focus on QDR's efforts to address two key challenges for qualitative data sharing. The first challenge concerns constraints on data sharing in order to protect human participants and their identities and to comply with copyright laws. The second set of challenges addresses the unique characteristics of qualitative data and their relationship to the published text. We describe a novel method of annotating scholarly publications, resulting in a "transparency appendix" that allows the sharing of such "granular data" (Moravcsik et al., 2013). We conclude by describing the future directions of QDR's services for qualitative data archiving, sharing, and reuse.

Keywords

Collection development, data collections, data services, services to user populations, social science literatures

Submitted: 17 May 2016; Accepted: 6 September 2016.

Introduction

Social science has a rich history of archiving, sharing, and reusing data for secondary analysis. Institutions such as the Inter-university Consortium for Political and Social Research, the Roper Center for Public Opinion Research, UK Data, and the data banks of various international organizations have for decades provided quantitative, matrix-based[1] datasets for a variety of empirical studies across the social sciences. However, a large number of primary data collections created by social scientists still remain invisible to the broader research community (Tenopir et al., 2011: see especially Tables 10 and 11). And, until recently, the absence of dedicated repositories for qualitative materials made the likelihood that qualitative social scientists would either share their own data in a way that increases research transparency or make use of others' for secondary analysis almost nonexistent (see e.g. Medjedović and Witzel, 2010, focusing on Germany; Yoon 2014).

The Qualitative Data Repository (QDR) provides infrastructure for the sharing and reuse of digital data used in qualitative and multi-method social inquiry (Elman et al., 2010). QDR is housed at the Center for Qualitative and Multi-Method Inquiry[2] (a unit of Syracuse University's Maxwell School of Citizenship and Public Affairs[3]), and funded by the National Science Foundation. The repository is guided by four beliefs:

Corresponding author:
Sebastian Karcher, Syracuse University, 346 Eggers Hall, Syracuse, NY 13244-1100, USA.
Email: karcher@u.northwestern.edu

- all data that can be shared and reused should be;
- evidence-based claims should be made transparently;
- teaching is enriched by the use of well-documented data; and
- rigorous social science requires common understandings of its research methods (Qualitative Data Repository, 2015).

QDR operates most explicitly to develop and provide a user-friendly data submission and preservation platform and to publish data projects in the social sciences domain that scholars can use for analytic and pedagogic purposes. Underlying these operations, however, is the deeper mission to supply practical guidance for its user community of qualitative researchers, especially on the challenging issues of legal and ethical sharing; to educate researches in the basics of data management so that they are well positioned to prepare their own projects with the goal of sharing in mind from the early planning stages; and even to promote a different way of thinking about the materials these scholars create and collect as "qualitative data." To achieve these goals, QDR does not just provide technical infrastructure but has dedicated staff working individually with data depositors to curate qualitative data for preservation and reuse. Over the past two years, QDR has made substantial progress in promoting data sharing and research transparency in social science by tackling challenges unique to qualitative data. We have developed strategies for addressing the concrete copyright and human participant constraints that occur when sharing such data and have developed tools that allow for transparent inference accommodating their unique structure.

Copyright and human subjects

An early set of challenges encountered by QDR are constraints placed on data sharing in order to protect human participants and their identities and to comply with copyright laws. Qualitative and mixed methods research often draws upon data—such as archival documents, images, video, interviews, etc.—that have a mix of intellectual property rights and personally identifying information (including rich contextual information). Concerns about privacy and ethics in data sharing figure just as prominently for quantitative research (King, 2011; Lagoze et al., 2013; Lupia and Alter, 2014), but qualitative data pose a unique set of challenges, which tools and strategies borrowed from quantitative traditions cannot solve. Copyright concerns, while shared with other digital archives and repositories, are all but nonexistent for quantitative data (most of which cannot be copyrighted qua data), so QDR has been treading ground unfamiliar for most data repositories.[4]

Annotation for transparency

A second set of challenges in sharing qualitative data concerns the nature of the data. In quantitative, matrix-based data, datasets are analyzed using dedicated software (e.g. R, Stata, SPSS) to produce a set of tables and/or figures. In contrast, qualitative researchers will make a multitude of empirical claims throughout their text. Each claim is backed by a piece of data (e.g. an excerpt from an interview or an archival source), which the author has analyzed or interpreted individually. We refer to this type of data and their analysis as *granular*.[5] Building on the work of Andrew Moravcsik on active citations, a form of author-contributed annotations supplementing a formal publication (Moravcsik, 2010, 2014a, 2014b), QDR has developed a set of pilot studies, working with qualitative researchers wanting to share granular qualitative data in the form of a transparency appendix (TRAX).[6] In the process, we have developed guidelines to help researchers share such data (Moravcsik et al., 2013) and also identified key shortcomings of current technologies and avenues for future improvement, both in methodology and technology.

The rest of the paper will proceed as follows: we begin by reviewing previous work on qualitative data archiving, sharing, and reuse, drawing attention to empirical work that sheds light on the scholarly practices of contemporary social scientists. We then describe curation services developed by QDR to assist researchers in archiving, sharing, and reusing qualitative data. The first set of services describes challenges and solutions in sharing sensitive or copyrighted materials. The second set of services describes an emerging approach to sharing granular qualitative data. We conclude by outlining some of QDR's future activities aimed at improving both the collection and publication of granular data.

Background: Qualitative data archiving, citation, sharing, and reuse

Social scientists have, for many years, used normative arguments for why the reproducibility of their research findings is critical for the viability of the field (King, 1995). As a result, social science data repositories are some of the longest standing and most mature infrastructures for preserving, providing meaningful access to, and representing the myriad research objects produced by social scientists;

including qualitative and quantitative data, statistical software, images, videos, codebooks, and data dictionaries. Studies of scholarly practices—such as citation behaviors, archiving, and reusing data—have similarly had an important impact on the development of policies and services developed by data repositories to serve social scientists. Below, we review some important studies influencing QDR's development of qualitative data curation services.

Data citation

Some of the earliest research into data citation behavior has occurred in the social sciences, including Joan Sieber's work on the norms and practices of data sharing and reuse (e.g. Sieber, 1991; Sieber and Stanley,1988). Sieber and Trumbo (1995) showed that although both qualitative and quantitative social scientists often engage in secondary analysis, only 19% ($n = 168$) of authors explicitly acknowledge the source of their data in a formal publication. In more recent scholarship, Mooney (2011) showed that of 49 studies reusing data archived at ICPSR, 60% ($n = 30$) failed to cite these data in a reference list. Also using ICPSR citation data, Fear (2013) showed that although secondary analysis is often performed with social science data, citation practices are diffuse—many authors use citations to credit both a repository (or source of data) as well as authors who had originally interpreted those data.

Data reuse

Fear's (2013) study also showed that tracking the impact of data reuse in the social sciences is complicated by hostility towards formal archiving policies. Other practical objections to the viability of meaningful reuse of social science data are whether or not the context of data collection can be meaningfully reinterpreted in secondary analysis (Boddy, 2001; Fujii, 2016), the complications of intellectual property rights held by participants and researchers (Mauthner et al., 1998), and the ethical ramifications of preserving human-subject materials (Bishop, 2009; Broom and Cheshire 2009; Carusi and Jirotka, 2009). Depending on epistemological commitments, for some scholars who use ethnography or participatory research, data can be seen as co-produced by both researcher and participants—both with equal claims to ownership and intellectual control (Mauthner et al., 1998). Parry and Mauthner (2004) compare the tradition of oral history scholarship, where ownership of materials (transcripts and recordings) is assumed to be held by respondents, with qualitative social scientists, where ownership is retained by an institution or individual conducting the research. They argue that while some social science archives offer guidance for researchers navigating these issues (e.g. Corti and Thompson, 1997), the various modes of producing qualitative data will require a broader, more malleable intellectual property framework (Parry and Mauthner, 2005). In studying social scientists who had completed secondary analysis of social science data, Faniel et al. (2015) demonstrate that five qualities were significant factors of satisfaction in reuse—data completeness, data accessibility, data ease of operation, data credibility, and data documentation.

Data sharing

Few studies have focused specifically on qualitative data sharing. Yoon's (2014) study of 13 researchers who had published a research article based on secondary analysis of qualitative data found that participants had only engaged in this type of analysis one to two times previously, that sharing happened solely through advisors or co-workers, and that successful reuse relied upon access to and in-depth conversation with the original data collector. As Yoon notes, ethical objections to sharing qualitative data are often the reasons why researchers withhold data from federal archiving mandates. This is the case in the UK where "ethics-related" exemptions are the most frequent waiver used by social scientists (Van den Eynden, 2008). Ethical objections to sharing qualitative data are typically made for the sake of protecting research participants' confidentiality, consent, or privacy. However, Bishop (2009) argues that studies impacted by these issues represent only a narrow subset of qualitative inquiry and the subject requires more careful consideration by archivists and infrastructure developers. Indeed, while there are many tools to facilitate privacy-protection in archiving research data (Fung et al., 2010), there have been few applications that cater to the unique ethical aspects of qualitative data as described above.

What makes the challenge even more complicated is that the ethical constraints that researchers legitimately worry about do not overlap completely with the legal protections put in place with the intent of preventing exploitation and harm of research participants (see Bosk and De Vries, 2004). A researcher wanting to safely observe both sets of considerations, whose only guidance on the issue might come from a local, risk-averse, and tradition-bound institutional review board, will almost always conclude that sharing of the granular data they have collected in interactions with human participants is not a good idea and will thus perpetuate the status quo of putting all these rich materials "under lock" or, even worse, promising

to destroy them at the end of the project (Bishop, 2009: 261).

Intellectual property

Concerns about infringing on intellectual property, which come into play when archival and other records fixed in a tangible medium are used as primary qualitative data, leads to a different dynamic when these materials are thought of as the data which a scholarly project analyzes. Mindful of the need to use representative information in order to arrive at accurate inferences, scholars are often in pursuit of comprehensiveness and breadth when they collect such materials. But the greater the number of items they use, say from an archive or a magazine articles database, and the greater the proportion of each copy they obtain for their purposes, the more problematic further sharing might become according to the stipulations of current US copyright law (Copyright Act of 1976 and subsequent amendments, known collectively as Title 17 of the United States Code). The usual exceptions, primarily the so-called "fair use" allowance for quoting portions of restricted length for the purpose of scholarly analysis, is of little help where the goal is to share complete data, as encouraged by research transparency norms.

Where scholars plan to use a comprehensive selection of materials with third-party ownership of copyright, the best approach may be to request explicit permissions for archiving and sharing with one's wider scholarly community (UK Data Service, n.d.). The permissions obtained will then become part of the administrative documentation of a data project. This can be accomplished best when the researcher is aware of the issue in advance and plans for its resolution from the initial stages of the project.

Qualitative data sharing in practice: Lessons from QDR'S pilot studies

How did the issues explored in the previous session surface during QDR's pilot phase? What were the principal challenges and what some of the solutions encountered by QDR staff? In this section we explore this question focussing on two sets of cases. We first address copyright and privacy concerns that affect nearly all qualitative data. We then explore the experience with sharing granular qualitative data in the form of active citation compilations.

Copyright and privacy concerns

In order to test how some of these principles of resolving the two key categories of concerns might be implemented in the form of practical solutions for qualitative data sharing, QDR began by commissioning a number of pilot projects, largely selected to illustrate in practice some of the thornier aspects of both copyright and human participant safety. Processing these early deposits, we learned that a lot of qualitative data can be shared both legally and ethically but that social scientists need to be aware of the strategies and the repository's technical tools that enable that. Most of all, we became convinced that the repository's role must be pedagogical as much as technological. It needs to impart knowledge of basic data management concepts, as derived from the library and information sciences' long-standing efforts in this field, to its intended user community. Once empowered with an understanding of general principles and best approaches in data management, depositors were in a position to collaborate with QDR's staff to identify applicable ways of sharing their data.

One interesting discovery was that while all researchers who used data collected by interviews or participant observation were acutely aware of the potential for problems regarding the privacy and confidentiality of their interlocutors in the field, few of those who had used archival or secondary literature resources considered the legal implications when they were asked to share such data. In fact, the discussions with QDR's staff and legal counsel on the topic was the first time some researchers encountered the concepts of copyright protection and fair use.

Another general point that united all piloteers was the need for guidance in the creation of useful documentation that contextualized both the data collection/creation process and the resulting qualitative materials. As mentioned above, the extremely heterogeneous nature of granular data makes existing categories such as samples, variables, codebooks, descriptive statistics, and survey instruments, routinely used to document matrix-organized data, unhelpful to many qualitative researchers. Understanding both the logic of documenting the process through which data were obtained and the nature of the process of field research itself enabled QDR's staff to "translate" to researchers in language more relevant to them the goals of contextualizing a deposit in a way that makes the data intellectually accessible and useful for secondary users as well as more discoverable and secure. In fact, we believe that most qualitative researchers would welcome the opportunity to provide in-depth descriptions of the path that started with their initial research designs, passed through the excitement of entering their field sites for the first (or nth) time, proceeded through the vicissitudes of exploring the site, led to the compilation of much more diverse data than could be usefully processed in a single

project, and resulted in the subset of better organized materials that actually underwent analysis for publication and ended up as the coherent data deposit.

In a particularly gratifying instance (by a team who engaged in multi-method work that used information gleaned from interviews and extensive following of the local press's coverage on the topic of interest to develop a rich case study, supplemented by quantified aggregations of some of the same facts), the researchers had prepared what they called "data narratives," which essentially captured all the expected details about the research process, enabling traceability of their whole research process, independent methodological learning, and any secondary reader's understanding of their conclusions.

In dealing with human participants, the majority of pilot project creators—whether active citations or more conventional stand-alone data projects—did resort to the default position of qualitative researchers of not sharing the data. This was where the challenge of working through data management issues only retroactively (i.e. belatedly from the perspective of a data manager) proved most detrimental to transparency. In some cases, the provision of at least some excerpts from interviews proved to be an acceptable technique. Researchers felt comfortable using short extracts (in some cases anonymous, in others not, especially when public figures were the source of information) at discrete points in their publications where they deemed that sufficient to illustrate the empirical contention they were making.

Compared with the alternative of no access to any elements of the primary data, this partial solution seems preferable. But in at least one case the researchers decided to undertake a more ambitious step and carefully anonymize a full set of 100 interview transcripts submitted as a stand-alone data collection. The process required close iterative discussions between their team and QDR's curation staff, combing through the basics of direct identifiers first and subsequently the ever more fine-grained details that taken together could serve as indirect identifiers—an aspect that will inevitably continue to bedevil the richest examples of qualitative data projects created in interaction with participants. The fact that the language of the original data collection was not English (also a characteristic we expect will continue to be common to the type of international fieldwork data QDR attracts) introduced an additional wrinkle in the process. As a general solution, QDR currently requests all documentation in English, regardless of the source language of the data. Despite all these difficulties, the anonymization protocol arrived at by these researchers, and the extensive set of materials processed according to it, are both valuable scholarly products that others can learn from both methodologically and substantively (see Dunning and Camp, 2016).

While the exact anonymization choices had to be agreed through considerable discussion and applied laboriously, other strategies QDR could offer to enhance participant protections were more easily borrowed from existing data management best practices long used for quantitative surveys. In addition to de-identifying the data, the researchers made sure to select only a sample of their full set of interviews[7] and availed themselves of one of the more stringent access control options QDR offers. They also agreed to revisit the last choice within a few years, as QDR's developing practices suggest more precise ways to assess both the level of risk that the anonymized transcripts might somehow be de-identified by others familiar with the context (currently estimated as very low to low for this project), and the severity of any potential repercussions for the participants, should their identity become known (currently estimated as low to medium, with the imagined repercussions being of a reputational and professional nature).

The lessons learned about copyright-related issues can similarly be grouped into "technological" and "sociological." Only one of the early depositors had considered the copyright status of the archival materials employed in their project before being asked about it by QDR. Their initial assessment was that they would not be able to share the pages digitally photographed during data collection. The terms of use of the relevant archive were the principal cause of concern. However, it seemed to the QDR staff that a more differentiated assessment was needed for the different types of materials the depositor had used, since a large portion of them were created by officials of the United States as part of their official duties and so most likely were in the public domain (another one of the exceptions that the copyright law allows for). Through discussions with the copyright counsel QDR uses, we established both that this was true for a subset of the data files and that within the context of an active citation type project which this piloteer was compiling, the fair use exception would apply to all the source materials in any case.

The fair use rule does apply extremely well to the suggested model of annotating one's work as done in active citation projects. This technique checks all the boxes for the four factors considered in establishing fair use: (1) the use of the original materials is transformative and adds original value; (2) it is done for non-commercial, explicitly scholarly purposes; (3) the portion of the material used is, relatively speaking, not substantial; (4) the use does not negatively affect

the existing market for the original work. While active citations were developed primarily with the goals of analytic and production transparency in mind (Moravcsik, 2014a), employing this novel scholarly technique could be one major way of providing at least partial secondary access to textual or audio and visual data, whose direct sharing might otherwise be prohibited by copyright ownership.

As an interesting aside that highlights the fluid nature of the regulatory and technical context within which QDR operates, in the course of curation work for this project, the archive from which the data had originally been collected underwent its own digitization initiative very much driven by motivations similar to some of QDR's broadest goals, i.e. to protect qualitative data in the form of historical assets and facilitate wider discoverability and access for further research. This changed the way the various materials could be linked to the points in the publication where they undergirded particular empirical claims. For a future project that might use the same archive's materials, a scholar will only need to provide the hyperlinks to the items now seamlessly available online.

A second use case is more typical of how QDR's intended depositors carry out their work and the copyright concerns that may commonly arise. A researcher had collected copies of over a thousand video recordings (some digital, others digitized from videotapes, or, in some cases, videotaped from television broadcasts and then digitized), created over almost two decades and in several different countries. He had inventoried all of them diligently (and so had created a lot of the substantive metadata we would need for repository purposes), but had not once wondered about the copyright ownership of any of them. In fact, for the pre-sharing academic uses he put these expansive data to (research, analysis, and citation), he did not need to. Once the question of storing them with QDR and making them more widely available for other researchers came into play, however, that consideration became paramount. Once again, the diversity of sources (and, relatedly, potential copyright claimants) meant that a blanket assessment was neither useful nor feasible. But the amount of work it would take to investigate the situation for each recording was prohibitive, not even taking into account the different unfamiliar national jurisdictions that might need to be considered and the multiple foreign languages in which legal communication would have had to take place.

The solution in this case was two-fold: (1) in the short term, QDR presents only a small number of items (fewer than 50) from the rich collection, which the researcher had obtained directly from the production companies, giving at least implicit permission for further use by the copyright-holders; (2) during a second phase, QDR is planning to make five-to-ten-second previews of all videos available via a dedicated viewer. Anyone wanting to use the full set of materials will need to visit on-site and access the files on a computer disconnected from the Internet (data enclave). In this instance, the intentional "diminishment" of the data could happen purely in quantitative terms, with fewer items or a small sample of each item being made directly available to comply with copyright constraints. Unlike in the anonymization example above, where the diminishment via considered aggregation of qualitative characteristics was substantive and could only be applied by the original researchers familiar with their subject matter, in this case the solution was technological and, once developed, can be applied to future cases that exhibited the same copyright constraint.

Another lesson presents a recurring theme in QDR's work: had the researchers planned for data archiving and sharing from the beginning of their work and been aware of these concerns[8] and the various possible solutions to them, they would have been in a much better position to tackle them during the course of their data collection and found themselves with much less to correct after the fact. This common sense conclusion emerged from QDR's work on every aspect of each pilot project, but it is hardly unique to qualitative data. Archivists and data management professionals across the disciplinary spectrum have made the case for early data management consultations for a significant time.

The challenge for QDR is to create useful guidance texts to prepare the members of our user community, who are not accustomed to think in "data management" terms when embarking on research projects. We believe that the recent requirement by many of the leading funding agencies in the social sciences to submit a data management plan (DMP) along with all grant proposals will publicize the broader need for researchers to seek out such guidance and think through the data management issues most relevant to their type of data collection. And indeed, QDR has received some early inquiries based on grant-required DMPs. Additionally, the creation of a handful of new DMP databases[9] that cover a wide variety of research contexts will advance the opportunities for qualitative researchers in particular to see examples relevant for their approaches to data.

As an encouraging sign, QDR has already started to see a number of inquiries by scholars heading out to the field, asking for consultation exactly on the points of preparing informed consent forms for human

participant protection and template language they can use to request copyright permissions for data sharing via a scholarly digital repository. We believe that thanks to appropriate planning and preparation for such issues, future projects deposited with QDR will involve smoother curation and faster completion to publication.

Working with granular data

Annotating publications with excerpts, additional notes, and access to primary materials to create a transparency appendix (TRAX) is one of the most interesting newer techniques for making qualitative research more transparent. It is endorsed by the flagship journal of the American Political Science Association (2016: Note 5). For qualitative researchers, having a data-sharing methodology that conforms to their actual practice rather than a poorly adapted quantitative template makes active citations particularly attractive. Browsing the active citation compilations currently available on QDR, enriching publications with materials ranging from old Soviet publications (Snyder, 2015), diary entries from John F. Kennedy (Saunders, 2015), to interviews with African judges (Ellett, 2015) provides a sense of the promise of the methodology.

In working with a group of eight piloteers (beyond the five cited in footnote 1, three more are scheduled for publication during 2016), we have collected insights about best practices and possible impediments to adoption of active citations. The two principal obstacles to wider adoption are the mode of publication and the workload active citations impose on depositors.

All active citation compilations published by QDR contain the full text of the article or book chapter in question, with annotations unfolding in the text on click. This allows readers to move seamlessly between data and publication. However, given that so called "gold" open access, allowing for free republication of works, is virtually nonexistent in political science (Atchison and Bull, 2015: 130), this poses a dilemma. For each of the works published on QDR, authors needed to obtain special permission from their publishers, who hold the copyright. Publishers were willing to grant such permission, but undoubtedly they would be less inclined to do so if active citations or similar models become commonly used. Moreover, QDR, as a data repository, is not in a position to serve as a publisher of scientific literature, even where copyright allows. For active citations to become more widely used, they need to work with content published on publisher's websites.

Concerns about the additional workload data sharing places on researchers looms large in the debate about data policies and is not limited to qualitative data or the social sciences. In an early study on data sharing in genetics, Campbell and Bendavide (2003: 242) find that 80% of researchers who had withheld data reported the work required to produce the requested material as a reason. Among skeptics of data-sharing policies for qualitative social science, the fact that such standards "impose a genuinely burdensome amount of new work" (Lynch, 2016: 37) is the most commonly cited objection. Such objections are not limited to "outsiders." Two of QDR's piloteers, reflecting on their experience note that "there is no getting around the basic fact that there remain significant costs to transparency that will be borne by individual scholars," (Saunders, 2014: 694) and that "At times it seemed that a yawning active citation sinkhole was about to open up and swallow all of my free research time and my assistant's" (Snyder, 2014: 714).[10]

A closer look at feedback from piloteers suggests that the effort is due to a combination of factors. Several authors found instructions lengthy and hard to follow. They also encountered occasional technological issues which added to their frustration. Finally, even where technology and instructions worked, compiling the required information into required forms and formats proved time consuming. Even where, for example, an author has a digital image of a document as well as a transcribed excerpt, attaching it to a given passage in an already authored text takes time. QDR has reacted to concerns about unsustainable amounts of work required for active citations by providing a streamlined workflow and continuously testing both technology and instructions so that researchers encounter a minimum of frustrations.[11]

These short-term efforts, however, provide only a partial solution. They provide no solution to the dilemma of publication. Moreover, they still leave a lot of "manual" labor for researchers to annotate their documents, even where they have all data present. In the conclusion of this article, we will discuss our vision forward that will, following the advice of Elizabeth Saunders, leverage "scholars' existing practices for capturing, storing, organizing, and maintaining data" (Saunders, 2014: 697) to achieve "transparency without tears."

Future directions

A lot of what QDR has managed to do on the basis of these pilot projects was to identify ways in which the original vision for tools and strategies could be made more complete and better integrated with solutions being worked on outside the immediate social science milieu that the repository serves. Through institutional and staff membership in organizations such as the Data

Preservation Alliance for Social Science, the Research Data Alliance, the International Association for Social Science Services and Technology, and the Annotating All Knowledge coalition, QDR is pursuing new ideas and new partners on several fronts.

Automating annotations via reference managers

In promoting annotations for transparent inference (ATI), the ''next generation'' of active citations, the single most important issue will be to improve usability and decrease the additional burden annotations place on qualitative researchers. From countless conversations with researchers as well as several published accounts (such as Saunders, 2014: 696–697), we know that researchers already store all information required for annotations in various software products. Any workflow that forces researchers to abandon their preferred tools will face stiff resistance and make adoption unlikely. What are the tools used by qualitative researchers to store data? From simple text documents and spreadsheets, to image organizing software like Picasa, to dedicated database products such as Microsoft Access, the variety of products used by researchers is enormous. Nevertheless, by far the most common tools used by researchers for storing their notes are reference managers such as Mendeley or Zotero.

This is fortuitous, as these tools also interact with text in a way very similar to annotations (Moravcsik, 2014b and Tonnesson, 2012 both make this point independently). In other words, all that is missing is a tool that connects the information users already store in their reference managers, such as notes and source documents, with the references they are inserting into their manuscript using these tools. QDR is currently working on building such integration. While initial prototypes focus on Zotero,[12] any tools and workflows developed by QDR will be sufficiently flexible to extend easily to other tools.

Scholarly annotation tools

In future work, we also hope to leverage a number of emerging standards and tools for creating and representing the content of ''annotations for transparent inference'' in machine readable form. For example, the Open Annotation data model is a standard endorsed by the W3C consortium (Haslhofer et al., 2011). This model allows a number of different domains to—in a standardized way—express the relationship between a user-generated annotation and a web-based text. In the example below, we demonstrate how QDR could use the Open Annotation model to express the relationship between an annotation applied by an author to their writing:

In this case, annotation 9 (anon-9) has a target in the text, and an explanation of the reason why the author has provided this annotation (to further interpret the main text). QDR could then model the motivation of this annotation using Open Annotation model's definition for ''interpreting'' as the reason for the annotation. In the future, we envision creating machine readable knowledge graphs (such as that shown in Figure 1) that can be included alongside new qualitative social science publications or as a supplement to previously published manuscripts.

Similar to the Zotero extension described above, we also hope to provide tools that allow authors to easily apply annotations to existing publications. This will be done by building upon existing browser-based tools like hypothes.is,[13] which allow users to apply annotations to any published text on the Internet (e.g. blogs, scientific articles, newspapers, etc.). QDR strongly believes that in combining these tools with standards like the Open Annotation model we can provide a practical and less burdensome path towards making annotation for transparent inference (ATI) a common practice among qualitative researchers.

Our short time spent learning from and working with the existing data and annotation communities has provided us with confidence that the technological solutions are in an exciting emergent state where QDR's vision of ATI can fruitfully contribute to the development of new digital tools that will facilitate researchers' personal annotating for the purposes of greater transparency. The sociological challenge remains a greater hurdle however, which will take longer to solve. The current training approaches and professional incentives for social scientists do not immediately allow, even for those interested in adopting the available new practices and tools for data sharing, to dedicate the necessary time and attention to prepare conventionally collected data for sharing as a public good. The mandates that journals and funders are starting to create on that front however, might change the current calculations. While necessary, these changes may not be sufficient in the absence of a broader reconsideration of data-sharing activities in hiring, tenure, and promotion decisions.

Outreach and guidance

While this broader academic landscape is beyond QDR's control, the repository is committed, also on the basis of these early lessons learned, to sharpen and apply its own strengths in close interaction with each future depositor. QDR continues to design strategies and services for qualitative data sharing, to

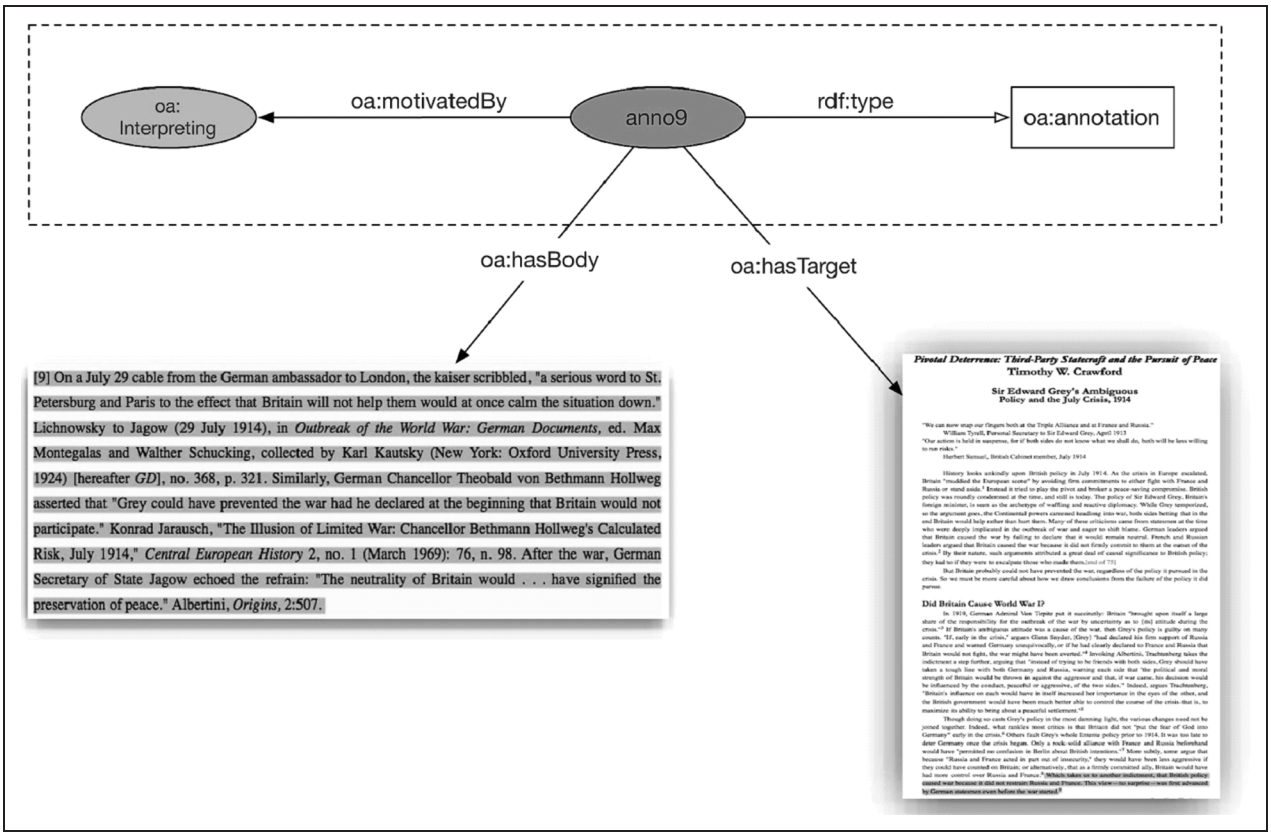

Figure 1. Open Annotations for Transparent Inference.

collaborate with journal publishers, to write up practical guidance materials and present data management classes at appropriate venues,[14] thus educating its user community with the goal of lowering the sociological barriers for an individual researcher. In this ongoing process, QDR's staff are themselves learning from the established data management and information science communities, refracting all new knowledge through the prism of social scientists' needs. Whereas all researchers need to add data management skills to their work and apply relevant techniques for data sharing more broadly and research transparency more specifically, having a dedicated venue for qualitative and multimethod data which provides the relevant expertise for curation tailored to them and advocates on behalf of the scholars who create and use them should make this task easier for this user community.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the National Science Foundation, award no. 1424191. Nic Weber is sponsored in part through a grant from the Alfred P. Sloan Foundation grant #G-2014-13746.

Notes

1. We refer to matrix-based data as data arranged in columns (variable) and rows (observations) as commonly used in statistical analysis. For this purpose, even data that are not rectangular by design, such as event-history data, are arranged in matrix form for analysis.
2. http://www1.maxwell.syr.edu/moynihan_cqrm.aspx
3. http://www.maxwell.syr.edu/
4. QDR is the first archive focusing on qualitative social science data in North America, but is able to draw upon a longer, though still young, tradition in Europe, including UK Data/Qualibank, the Irish Qualitative Data Archive, and Qualiservice in Germany.
5. We realize that the term ''granular'' may be used by information and data science to mean an indivisible unit of raw data. In this context we understand it simply in contrast to matrix data.
6. Finished active citation pilot studies as of mid-2016 include Crawford (2015); Handlin (2015); Herrera (2016); Saunders (2015); Snyder (2015). Given both our experience during those pilots and trends in scholarly annotations, QDR is moving away from the somewhat restrictive ''active citations'' towards ''annotations for transparent inference'' (ATI). Throughout this paper, we refer to our pilot projects

as active citation compilations. The final section then introduces the concept of annotations.

7. In order to not lose information about social connections among actors, they intentionally chose not to remove data completely at random, but to sample from within one of the three localities where interviews had taken place—a choice documented in the data narrative.
8. That is, human participant protection and copyright, for which external legal requirements need to be met as well and data documentation and file naming, where early preparation helps to prepare data for later sharing logistically.
9. Such as http://dmptool.org and http://rio.pensoft.net/browse_user_collection_documents.php? collection_id=3&journal_id=17.
10. In spite of this honest assessment of the challenges, both authors do see significant benefits to active citations—all the more reason to take their concerns seriously.
11. During the time of this writing, we are in a second series of user testing of the simplified instructions, which includes both feedback protocols and debriefing focus groups.
12. The choice for Zotero is mostly pragmatic: it is not only one of the most widely used reference managers, but it is also open source and has a well-documented API, which makes integration into other tools particularly easy.

13 Hypothes.is: https://hypothes.is/about/

14 As an example of a new training initiative, we have published a teaching module and shared it with methodology instructors in Political Science graduate programs, so that they can introduce to their students the practical foundations of data management in a class session.

References

American Political Science Association (2016) *APSR Submission Guidelines 2016 in Brief*. Available at: http://www.apsanet.org/PUBLICATIONS/Journals/APSR-Submission-Guidelines-2016-in-Brief#5 (accessed 16 May 2016).

Atchison A and Bull J (2015) Will open access get me cited? An analysis of the efficacy of open access publishing in political science. *PS: Political Science & Politics* 48(1): 129–137.

Bishop L (2009) Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues* 44(3): 255–272.

Boddy M (2001) *Data Policy and Data Archiving: Report on Consultation for the ESRC Research Resources Board*. Bristol: University of Bristol.

Bosk CL and De Vries RG (2004) Bureaucracies of mass deception: Institutional review boards and the ethics of ethnographic research. *ANNALS of the American Academy of Political and Social Science* 595(1): 249–263.

Broom A and Cheshire L (2009) Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology* 43(6): 1163–1180.

Campbell EG and Bendavid E (2003) Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *Journal of Health Care Law & Policy* 6(2): 241–255.

Carusi A and Jirotka M (2009) From data archive to ethical labyrinth. *Qualitative Research* 9(3): 285–298.

Corti L and Thompson P (1997) Latest news from the ESRC Qualitative Data Archival Resource Centre (QUALIDATA). *Social History* 22(1): 83–86.

Crawford T (2015) *Data for: Pivotal Deterrence and the Chain Gang: Sir Edward Grey's Ambiguous Policy and the July Crisis, 1914*. Active Citation Compilation QDR:10049. Syracuse, NY: Qualitative Data Repository [distributor]. Available at: http://dx.doi.org/10.5064/F6G44N6S (accessed 16 May 2016).

Dunning T and Edwin C (2015) *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Data Collection, QDR:10055. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F6Z60KZB.

Ellett R (2015) *Democratic and Judicial Stagnation*. Active Citation Compilation QDR:10064. Syracuse, NY: Qualitative Data Repository [distributor]. Available at: http://dx.doi.org/10.5064/F6PN93H4 (accessed 16 May 2016).

Elman C, Kapiszewski D and Vinuela L (2010) Qualitative data archiving: Rewards and challenges. *PS: Political Science & Politics* 43(1): 23–27.

Faniel IM, Kriesberg A and Yakel E (2015) Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology* 67(6): 1404–1416.

Fear KM (2013) *Measuring and anticipating the impact of data reuse*. PhD Thesis, University of Michigan, USA. Available at: http://deepblue.lib.umich.edu/handle/2027.42/102481 (accessed 16 May 2016).

Fujii LA (2016) The dark side of DA-RT. *Comparative Politics Newsletter* 26(1): 25–27.

Fung B, Wang K, Chen R, et al. (2010) Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)* 42(4): 14.

Handlin S (2015) *Data for: The Politics of Polarization: Governance Quality, Left Factionalism, and Party Systems in Latin America*. Active Citation Compilation QDR:10065. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F66Q1V52.

Haslhofer B, Simon R, Sanderson R, et al. (2011) The open annotation collaboration (OAC) model. In: *MMWEB'11 Workshop on Multimedia on the Web*, pp. 5–9. Washington, DC: IEEE.

Herrera V (2016) *Data for: Commercialization and Decentralization of Local Services Provision*. Active Citation Compilation QDR:10050. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F6F769GQ.

King G (1995) Replication, replication. *PS: Political Science & Politics* 28(3): 444–452.

King G (2011) *Ensuring the Data-Rich Future of the Social Sciences*. Available at: https://dash.harvard.edu/handle/1/12724029 (accessed 16 May 2016).

Lagoze C, Willliams J and Vilhuber L (2013) Encoding provenance metadata for social science datasets. In: Garoufallou E and Greenberg J (eds) *Metadata and Semantics Research*. Berlin: Springer, pp. 123–134.

Lupia A and Alter G (2014) Data access and research transparency in the quantitative tradition. *PS: Political Science & Politics* 47(1): 54–59.

Lynch M (2016) Area studies and the cost of prematurely implementing DA-RT. *Comparative Politics Newsletter* 26(1): 36–39.

Mauthner NS, Parry O and Backett- K (1998) The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology* 32(4): 733–745.

Medjedović I and Witzel A (2010) Sharing and archiving qualitative and qualitative longitudinal research data in Germany. *IASSIST Quarterly* (Fall/Winter 2010): 42–46. Available at: http://www.iassistdata.org/sites/default/files/iqvol34_35_witzel.pdf (accessed 29 September 2016).

Mooney H (2011) Citing data sources in the social sciences: Do authors do it? *Learned Publishing* 24(2): 99–108.

Moravcsik A (2010) Active citation: A precondition for replicable qualitative research. *PS: Political Science & Politics* 43(1): 29–35.

Moravcsik A (2014a) Transparency: The revolution in qualitative research. *PS: Political Science & Politics* 47(1): 48–53.

Moravcsik A (2014b) Trust, but verify: The transparency revolution and qualitative international relations. *Security Studies* 23(4): 663–688.

Moravcsik A, Elman C and Kapiszewski D (2013) *A Guide to Active Citation*. Syracuse, NY: Qualitative Data Repository. Available at: https://qdr.syr.edu/guidance/acguide (accessed 19 August 2016).

Parry O and Mauthner NS (2004) Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology* 38(1): 139–152.

Parry O and Mauthner N (2005) Back to basics: Who re-uses qualitative data and why? *Sociology* 39(2): 337–342.

Qualitative Data Repository (2015) *Our Mission*. Available at: https://qdr.syr.edu/ (accessed 17 May 2016).

Saunders EN (2014) Transparency without tears: A pragmatic approach to transparent security studies research. *Security Studies* 23(4): 689–698.

Saunders EN (2015) *Data for: John F. Kennedy*. Active Citation Compilation QDR:10048. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F68G8HMM.

Sieber JE (1991) Openness in the social sciences: Sharing data. *Ethics & Behavior* 1(2): 69–86.

Sieber JE and Stanley B (1988) Ethical and professional dimensions of socially sensitive research. *American Psychologist* 43(1): 49–55.

Sieber JE and Trumbo BE (1995) (Not) giving credit where credit is due: Citation of data sets. *Science and Engineering Ethics* 1(1): 11–20.

Snyder J (2014) Active citation: In search of smoking guns or meaningful context? *Security Studies* 23(4): 708–714.

Snyder J (2015) *Data for: Russia: The Politics and Psychology of Overcommitment*. Active Citation Compilation QDR:10047. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F6KW5CXS.

Tenopir C, Allard S, Douglass K, et al. (2011) Data sharing by scientists: Practices and perceptions. *PloS one* 6(6): e21101.

Tonnesson S (2012) Active citation through hyperlinks: The retarded replication revolution. *International Area Studies Review* 15(1): 83–90.

UK Data Service (n.d.) *Copyright Scenarios for Data Sharing*. Available at: https://www.ukdataservice.ac.uk/manage-data/copyright/scenarios (accessed 17 May 2016).

Van Den Eynden V (2008) Sharing research data and confidentiality: Restrictions caused by deficient consent forms. *Research Ethics Review* 4(1): 37–38.

Yoon A (2014) 'Making a square fit into a circle': Researchers' experiences reusing qualitative data. *Proceedings of the Annual Meeting for the Association for Information Science and Technology* 12(1): 1–4.

Author biographies

Sebastian Karcher is the Associate Director of the Qualitative Data Repository. He is an expert in scholarly workflows and an active contributor to academic open source software, in particular the Zotero reference manager and the Citation Style Language. He holds a PhD in Political Science from Northwestern University.

Dessi Kirilova has served as a Fellow and consultant for the Qualitative Data Repository since its establishment in 2012. In that capacity, she has assisted in shaping the repository's policies, acquisitions outreach, curatorial support for depositors, and the processing and publication of all currently available data deposits. Her related interests are in educating social science researchers in good data practices, starting in the early research planning stages. Her substantive work in political science focuses on post-communist foreign policies, European integration and the processes that lead to national identity formation, and uses elite interviewing and archival historical methods.

Nicholas Weber is an Assistant Professor at the University of Washington School of Information, a member of the iSchool DataLab, and the Associate Technical Director of the Qualitative Data Repository. He holds a PhD in Information Science from the University of Illinois, and is currently a Research Data Alliance (RDA) early-career fellow.

Article

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 303–312

DOI: 10.1177/0340035216672238
ifl.sagepub.com

Data governance, data literacy and the management of data quality

Tibor Koltay
Eszterházy Károly University, Hungary

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

Submitted: 11 May 2016; Accepted: 6 September 2016.

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 the 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 et al., 2012).

Against 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 RDSs. The second one is data literacy, about which there is a massive body of literature, among others in the form of review articles (Koltay, 2015a, 2015b; MacMillan, 2014). 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

Corresponding author:
Tibor Koltay, Eszterházy Károly University, Rákóczi út 53, Jászberény, 5100, Hungary.
Email: koltay.tibor@uni-eszterhazy.hu

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* in 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 et al., 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 of academic librarianship's past and present development paradigms, outlined by Martell (2009). The first paradigm, called the 'Ownership' or 'Collections' paradigm evolved after 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 could adequately satisfy the institutions' scholarly and teaching needs. Such support 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 came into being. 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 that are induced by its presence from the influence of 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 in which 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 on it to some degree.

Data governance in detail

As stated 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 clear that the academic sector, librarianship, as well as library and information science also should pay attention to DG, albeit it attracted 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 the adoption of common approaches to data issues. Data governance also helps build standard, repeatable processes, reduce costs and increase effectiveness through coordination of efforts and by enabling transparency of processes. It is governed by the principles of integrity, transparency and 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 (Khatri and

Brown, 2010; Weill and Ross, 2004). Seiner (2014: 2) 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 and 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 horizontally. 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 special attention. Big data governance is especially important in regard to the acceptable use of data (Soares, 2012). In environments where big data plays a 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). In addition to 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 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 and 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 about data governance's nature is foundational for RDSs and well-developed data governance is one of the necessary conditions for open data (Weber et al., 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 of 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 (2014) emphasizes that data governance is a critical component of any data quality management strategy. 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, directed towards data librarians, being knowledgeable about the ways to assess the digital objects' authenticity, integrity and accuracy over time would also 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 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 other things – documenting the provenance of data that ensures its quality is secured (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 $328m. With appropriate attention to data governance principles and to the actual details, and if data literacy skills had been mobilized, 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 (Koltay, 2015b; Qin and D'Ignazio, 2010; Schneider, 2013). 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 to read 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 Fosmire and Miller (2008) spoke simply about information literacy in the data world. Two years later, data literacy was defined plainly as the ability to understand, use and manage data (Qin and D'Ignazio (2010). According to Calzada Prado and Marzal's (2013) definition, data literacy enables individuals to access, interpret, critically assess, manage, handle and ethically use data.

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 to 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 essentially to enable us to efficiently process all types of information content (Badke, 2010) – the community of practice for data librarians differs from that 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 (2004) states that information literate persons obtain, store and disseminate not only text, but data as well. 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 (2011) Seven Pillars of Information Literacy model and the information literacy lens on the Vitae Researcher Development Framework (Vitae, 2011) 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: 126) identified a number of abilities, some of which clearly show their origin in the 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 to identify 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 (ACRL, 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 decision making. Therefore, they bring data literacy up to data governance, recognizing 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 and 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, 2015a).

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 their 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 et al. (2014) shows that information literacy both in the health sciences and the social sciences have their own specific 'personality'. In general, newer approaches to information literacy underline that information is used in 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, 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 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 and 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 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 materializes 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 3 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. At 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 ensure that students, faculty and researchers become data literate science workers. As Carlson and Johnston (2015) underline, 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 et al. (2013) pointed out, faculty members rarely like to hear that they are doing something in the 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 a lot of 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 form data governance can improve data management in the library. Second, data governance is a service that is based on standardized, 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 towards making all our target audiences accept the library's mission to provide research data services and to offer these services to their full satisfaction.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

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 May 2016).

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 May 2016).

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 May 2016).

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

Badke W (2010) Information overload? Maybe not. *Online* 34(5): 52–54.

Bailey CW (2015) *Research Data Curation Bibliography, Version 5*. Houston, TX: Digital Scholarship. Available at: http://digital-scholarship.org/rdcb/rdcb.htm (accessed 2 May 2016).

Bawden D and Robinson L (2015) 'An intensity around information': The changing face of chemical information literacy. *Journal of Information Science*. DOI: 10.1177/0165551515616919.

Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(5): 662–669.

Briney K (2015) *Data Management for Researchers: Organize, Maintain and Share your Data for Research Success*. Exeter: Pelagic.

Bundy A (ed.) (2004) *Australian and New Zealand Information Literacy Framework*. 2nd edn. Adelaide: Australian and New Zealand Institute for Information Literacy.

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 and Johnston LR (2015) *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 (eds) (2015) Planting seeds for data literacy: Lessons learned from a student-centered education program. *International Journal of Digital Curation* 10(1): 95–110.

Carlson J, Fosmire M, Miller C, et al. (2011) Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2): 629–657.

DGI (2015a) *Definitions of Data Governance*. Available at: http://www.datagovernance.com/adg_data_governance_definition/ (accessed 2 May 2016).

DGI (2015b) *Data Governance: The Basic Information*. Data Governance Institute. Available at: http://www.datagovernance.com/adg_data_governance_basics/ (accessed 2 May 2016).

Dong XL and Srivastava D (2013) Big data integration. In: *Data Engineering (ICDE), 2013 IEEE 29th international conference*, Seoul, Korea, 16–19 April 2013, pp. 1245–1248. New York, NY: IEE.

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 May 2016).

Duncan J, Clement K and Rozum B (2013) Teaching our faculty. developing copyright and scholarly communication outreach programs. In: Davis-Kahl S and Hensley MK (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 May 2016).

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. In: *Proceedings of the IATUL conferences*, Paper 16. Available at: http://docs.lib.purdue.edu/iatul/2008/papers/16 (accessed 2 May 2016).

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.

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/iqvol282_3hunt.pdf (accessed 2 May 2016).

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 May 2016).

Information Builders (2014) *Breaking Big: When Big Data Goes Bad: The Importance of Data Quality Management in Big Data Environments*. New York: Information Builders.

Jahnke L, Asher A and Keralis SD (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.

Koltay T (2015a) Data literacy: In search of a name and identity. *Journal of Documentation* 71(2): 401–415.

Koltay T (2015b) Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*. DOI: 10.1177/0961000615616450.

Krier L and Strasser CA (2014) *Data Management for Libraries*. Chicago, IL: American Library Association.

Lenzerini M (2002) Data integration: A theoretical perspective. In: *Twenty-first ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems,* Madison, WI, USA, 3–5 June 2002, pp. 233–246. New York, NY: Association of Computing Machinery.

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. *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 C (2009) sAccess: The social dimension of a new paradigm for academic librarianship. *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: *78th ASIS&T annual meeting: Information science with impact: Research in and for the community*, St Louis, MS, USA, pp. 108–111. Silver Spring, MD: American Society for Information Science.

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: 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: *31st annual IATUL conference*. Available at: http://docs.lib.purdue.edu/iatul2010/conf/day2/5 (accessed 2 May 2016).

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 C, 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 May 2016).

Riley AC (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 May 2016).

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.

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 (accessed 2 May 2016).

Searle S, Wolski M, Simons N, et al. (2015) Librarians as partners in research data service development at Griffith University. *Program* 49(4): 440–460.

Seiner RS (2014) *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success*. Basking Ridge, NJ: Technics Publications.

Si L, Zhuang X, Xing W, et al. (2013) The cultivation of scientific data specialists. *Library Hi Tech* 31(4): 700–724.

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 May 2016).

Soares S (2012) *Big Data Governance: An Emerging Imperative*. Boise, ID: MC Press.

Tenopir C, Birch B and 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, 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. Available at: http://dx.doi.org/10.7191/jeslib.2015.1085 (accessed 28 September 2016).

Tenopir C, Sanduski RJ, 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.

Vitae (2011) *Researcher Development Framework*. Cambridge: Careers Research and Advisory Centre. Available at: https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework-rdf-vitae.pdf (accessed 28 September 2016).

Wang M (2013) Supporting the research process through expanded library data services. *Program* 47(3): 282–303.

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.

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.

Author biography

Tibor Koltay is Professor of Library and Information Science and chairs the Institute of Learning Technologies at Eszterházy Károly University, in Jászberény, Hungary. He received his MA and PhD from Eötvös Loránd University, Budapest, Hungary. He holds a Certificate of Advanced Studies in Library and Information Science from Kent State University, Ohio. His professional interest includes examining the relationship among information literacy, media literacy, digital literacy and data literacy. He has also published papers about information overload.

Article

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 313–327

DOI: 10.1177/0340035216673382
ifl.sagepub.com

Data information literacy instruction in Business and Public Health: Comparative case studies

Katharine V. Macy
Indiana University-Purdue University Indianapolis, USA

Heather L. Coates
Indiana University-Purdue University Indianapolis, USA

Abstract

Employers need a workforce capable of using data to create actionable information. This requires students to develop data information literacy competencies that enable them to navigate and create meaning in an increasingly complex information world. This article examines why data information literacy should be integrated into program curricula, specifically in the instances of business and public health, and offers strategies for how it can be accomplished. We approach this as a comparative case study within undergraduate business and master of public health programs at Indiana University-Purdue University Indianapolis. These case studies reveal several implications for practice that apply across social and health sciences programs.

Keywords

academic libraries, business, data information literacy, data reuse, information literacy, instruction, public health

Submitted: 18 May 2016; Accepted: 14 September 2016.

Introduction

Graduates entering the workforce must be able to navigate personal and professional environments in which data plays an ever increasing role. The volume of available data continues to increase at an exponential rate as the internet of things grows, social media and mobile devices generate vast quantities of data, and the open data movement spreads. Graduates will need to function as critical and ethical data consumers of quantitative and qualitative data originating from organizations and systems over which they have little control. In addition, public health professionals must be capable of generating and managing data about specific communities. Both fields require fluency in identifying the data, contextual, and technical components of their information needs in order to solve problems. We describe these skills as data information literacy (DIL), a critical set of competencies for graduates entering the workforce. To best prepare our students for using data to make decisions, we believe DIL needs to be integrated throughout the curriculum. This article examines initiatives at Indiana University-Purdue University Indianapolis (IUPUI) to integrate DIL instruction into an undergraduate business program and a graduate public health program. In each case study, we describe the disciplinary and professional context for DIL instruction, examine relevant educational standards and professional competencies, describe early instructional activities, and present implications for practice.

Background

In January 2016, the Association of College & Research Libraries (ACRL) adopted a new

Corresponding author:
Heather L. Coates, Indiana University-Purdue University Indianapolis, 755 W Michigan Street, Indianapolis, IN 46202, USA
Email: hcoates@iupui.edu

Framework for Information Literacy for Higher Education, an update from the *Information Literacy Competency Standards for Higher Education* published 15 years prior (ACRL, 2015). While data had been referenced, much of the original standards focused on information literacy skills as they relate to textual sources (ACRL, 2000). The new framework is more inclusive of broader information types and practices including finding, using, and understanding data, creating flexibility to adapt as resources, technology, and pedagogy evolves. However, because the framework is less structured, it does not dive deeply into the specific data competencies students need to develop.

Since student learning outcomes drive instructional design, it is necessary to clearly describe the skills we want students to develop. We reviewed existing frameworks for data literacy and found several distinct approaches (Calzada Prado and Marzal, 2013; Carlson et al., 2011; Koltay, 2015; Mandinach and Gummer, 2012; Qin and D'Ignazio, 2010; Schield, 2004). Some consider data information literacy as a facet of information literacy, while others see it as a separate set of skills. Yet another perspective places DIL under the broad umbrella of digital literacies sharing common elements with numeric and statistical literacy. In practice, there is overlap between many of these skill sets, but DIL offers a useful way of describing these skills to students and instructors.

After reviewing the literature, we selected the framework described by Calzada Prado and Marzal (2013) because it is designed to be a common reference framework for the critical use of data as well as research data management. Thus, it is easily applied to multiple data roles - producer, manager, and consumer. Furthermore, the framework uses language that is general enough to be applied to a variety of disciplines and professions. Calzada Prado and Marzal (2013: 126) define DIL as, ''the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle, and ethically use data.'' Their framework for DIL is based on a careful review of the literature conducted across science and social science disciplines. The five-module framework covers:

1. understanding what data is and how it affects society;
2. finding and/or obtaining data resources;
3. reading, interpreting, and evaluating data;
4. managing data including creation of metadata and collection practices;
5. using data including data handling, data visualization, and ethical use.

There are differences in the practical implementation of DIL as compared to information literacy (Hunt and Birks, 2004). Each case study will describe the professional context that shapes application of the DIL competencies as well as the approach for integrating instruction into the curricula.

Methods

The cases were selected in order to compare DIL integration for two fields in which graduates often work as practitioners rather than academic researchers. They also provide the opportunity to examine differences between DIL instruction for undergraduate and graduate students. While the opportunity to integrate instruction into these programs arose independently, the authors identified a common approach. Educational standards were reviewed to determine how DIL can support program curricula. Reviews of professional standards and literature on employer surveys were also conducted. Particularly for the business case, these sources describe the skills that employers desire and whether recent graduates are meeting those needs. Overall, these standards and surveys helped us identify skills gaps that could be mitigated by DIL instruction. Existing library initiatives to integrate information literacy into program curricula presented an opportunity to build on prior conversations and relationships while extending instructional support to include DIL. In both cases, curriculum mapping was the common critical step.

Curriculum mapping is a process of collaboration and communication that librarians use to determine where and when information literacy instruction should be placed within a particular course of study at the point-of-need (Bullard and Holden, 2006). This systematic process allows librarians to gain knowledge of the needs and language relevant to that course of study. Such knowledge facilitates better communication and creates opportunities for librarians to collaborate with instructors in identifying where information literacy fits within the program. Curriculum mapping allows librarians to position DIL within the context of specific disciplines, creating alignment between education standards and program learning outcomes. In developing curriculum maps for these cases, we incorporated professional standards and employer needs as well as academic learning outcomes.

Program learning outcomes, course syllabi, and assignments were examined to identify specific opportunities for integrating DIL skills. Program learning outcomes provide a high level framework that describes the observable skills students are

expected to exhibit upon graduation. Program documents may also include student learning objectives, which are often more detailed and describe what the program intends students to learn. These outcomes and objectives are used to create course goals. Understanding this relationship between program learning outcomes, student learning objectives, and course goals enables librarians to identify and communicate how DIL instruction will support a course and the program. This knowledge enables more efficient analysis of syllabi. Examining course assignments assists librarians in identifying opportunities to integrate and assess DIL.

Case study: Undergraduate business education (subject librarian Katharine Macy)

Businesses wish to take advantage of the exponential growth of data, which requires the skills of data scientists and functional practitioners in all business areas. Employers are asking workers across their organizations to review, understand, and make decisions using data gathered from internal and external sources and systems. Companies desire that their employees have the ability to find and use data to determine trends, develop meaningful insights, and provide actionable recommendations to improve the business. In 2012, discussions at the Business Intelligence Congress 3 (BIC3) brought forth the recommendation that academic institutions should consider tailoring business intelligence education to generalists and specialists. BIC3 recommended that all business majors in undergraduate and master's-level graduate programs be considered generalists who need to develop DIL skills that will allow them to analyze and synthesize data within their functional area (Wixom et al., 2014).

As business schools move to integrate DIL into the curriculum to meet employer needs, libraries have a prime opportunity to collaborate with faculty and curriculum committees to reshape the curriculum. This case explores the skill sets and competencies desired by employers that can be developed through DIL instruction, as well as how business education standards encourage the integration of DIL into the curriculum. Understanding this context facilitates the conversation about DIL integration.

Help wanted: People who understand data

Data information literacy is relevant to undergraduates who are seeking a bachelor's as a terminal degree, particularly in the area of business (Shorish, 2015). Future business workers will need to be well versed in digital technology, social media, and big data in order to analyze information gathered from information technology systems critically (Bisoux, 2016). Business leaders also desire to spur innovation by taking advantage of the increased access to data available through open data initiatives (Calzada Prado and Marzal, 2013). Gartner, a technology market research firm, has found that the emphasis is moving from technology of big data to application, or in other words, how can data analysis address business problems facing an organization or industry (Burton and Willis, 2015). To maximize the business value of analytics, companies need employees who understand the business and can cross the communication divide that often exists between information technology and other departments (Ransbotham et al., 2015). This requires practitioners such as accountants, brand managers, financial analysts, and buyers to be able to use, analyze, and communicate using data when managing their business.

In 2014, the Association of American Colleges & Universities (AACU) commissioned a survey of employers and college students to determine the learning outcomes that employers most valued in college graduates. The survey indicated that there are a number of learning outcomes that employers rank as very important but feel that recent graduates are not well prepared to address (Hart Research Associates, 2015). See Table 1.

DIL integration into the business curriculum will meet some of employers' top needs and helps address the skills gaps employers are observing in new college hires.

Critical thinking and data information literacy

The desire for new hires to have critical thinking and/or analytical skills is emphasized in the AACU survey and business information literacy literature (Hart Research Associates, 2015; Klusek and Bornstein, 2006; Sokoloff, 2012). Critical thinking involves the process of analyzing, evaluating, and interpreting information (Schield, 2004). In business, it also includes the ability to apply that information to make business recommendations.

Three elements contribute to the development of critical thinking skills in undergraduate business students: context, business skills/knowledge, and literacies. Undergraduate education influences all three elements. Context is constructive and created through an individual's life experience and culture. According to Calzada Prado and Marzal (2013), critical thinking draws from a person's knowledge and values in addition to their mathematical and statistical aptitudes.

Table 1. Summary survey results from AACU employer survey (Hart Research Associates, 2015).

Learning outcome	% Employers view as very important	% Employers who feel graduates are well prepared
The ability to effectively communicate in writing	82	27
Ethical judgement and decision making	81	30
Critical thinking and analytical reasoning skills	81	26
The ability to analyze and solve complex problems	70	24
The ability to locate, organize, and evaluate information from multiple sources	68	29
The ability to work with numbers and understand statistics	56	28

Liberal arts education expands the context of reference for individuals across disciplines, while the business school curriculum provides disciplinary knowledge and skills. The third element needed for developing critical thinking skills, literacies, is developed throughout the entire curriculum, which includes information literacy, numeracy, statistical literacy, and data information literacy (Schield, 2004; Stephenson and Caravello, 2007).

Picking a data information literacy framework

Cunningham (2003) lists a number of information competencies for business students that were aligned with the information literacy standards published by ACRL in 2000. These competencies delve deeper into business education needs in regards to information literacy, statistical literacy, and data information literacy and are still relevant today. However, they were drafted prior to the explosion of big data and the increased desire of employers for data-driven decision making. Additional emphasis should be placed on teaching DIL within academic business programs. Calzada Prado and Marzal's (2013) framework provides guidance on the competencies students need to develop in regards to data, better preparing them for the working world. See Table 2 to understand how DIL competencies from Calzada Prado and Marzal's framework align with learning outcomes that employers value.

Meeting business education standards

The business education curriculum standards published by the Association to Advance Collegiate Schools of Business (AACSB), a global accrediting body for undergraduate and graduate business education, and the National Business Educators Association (NBEA), a professional organization supporting business educators in the United States of America, support the competencies listed in Calzada Prado and Marzal's framework.

The AACSB addresses curriculum content in Standard 9 of their *Eligibility Procedures and Accreditation Standards for Business Accreditation*. The standards state the expectation that students graduating with a Bachelor's Degree or higher develop the following skills (AACSB, 2013):

- Written and oral communication (able to communicate effectively orally and in writing);
- Ethical understanding and reasoning (able to identify ethical issues and address the issues in a socially responsible manner);
- Analytical thinking (able to analyze and frame problems).

Standard 9 also states that curriculum should cover several general business and management knowledge areas including (AACSB, 2013):

- Information technology and statistics/quantitative methods impacts on business practices to include data creation, data sharing, data analytics, data mining, data reporting, and storage between and across organizations including related ethical issues;
- Economic, political, regulatory, legal, technological, and social contexts of organizations in a global society;
- Social responsibility, including sustainability, and ethical behavior and approaches to management.

The skills and knowledge areas in the AACSB standards are interdependent and should be considered holistically when developing a curriculum. The emphasis on data in the AACSB standards in conjunction with employer demand for skilled workers makes it clear that integrating DIL into the curriculum is important.

Table 2. Alignment of employer survey learning outcomes to data information literacy competencies.

AACU learning outcome (Hart Research Associates, 2015)	Data information literacy competency (Calzada Prado and Marzal, 2013)
The ability to effectively communicate in writing	5.2 Producing elements for data synthesis
Ethical judgement and decision making	1.2 Data in society: a tool for knowledge and innovation. 5.3 Ethical use of data
Critical thinking and analytical reasoning skills	1.1 What is data 3.1 Reading and interpreting data 3.2 Evaluating data
The ability to analyze and solve complex problems	1.1 What is data 5.1 Data handling
The ability to locate, organize, and evaluate information from multiple sources	1.1 What is data 2.1 Data sources 2.2 Obtaining data 4.1 Data and metadata collection and management
The ability to work with numbers and understand statistics	3.1 Reading and interpreting data

Mapping to the curriculum and starting the conversation

The complexity of different functional courses of study (e.g. accounting, finance, marketing, etc.) creates a challenge when attempting to determine where to integrate DIL instruction within the business curriculum. The NBEA's (2013) standards provide guidance as to where DIL instruction could naturally fit into college-level business curriculum (Level 4) by functional area, and can be used to identify opportunities for collaboration with faculty. See Table 3.

Successful DIL integration within the business curriculum requires the assimilation of concepts into course content, assignments, and assessment. The business liaison can support this by starting the conversation and collaborating with faculty and curriculum committees. At IUPUI, I prepared to do this by first refreshing the student learning objectives (SLOs) used in library instruction planning for the undergraduate business program. Twenty-six SLOs were drafted supporting the development of information literacy skills, business information literacy skills, and data information literacy skills, informed by Cunningham's (2003) article, the ACRL *Framework for Information Literacy for Higher Education* (2015) and the Calzada Prado and Marzal (2013) data information literacy framework. These SLOs were mapped to specific courses within the undergraduate curriculum to create a guide for incorporating library instruction. The curriculum mapping process identified opportunities where inserting DIL aligns with current course objectives.

When speaking to business faculty at IUPUI, some associated DIL with the skills needed for conducting academic research, less applicable for undergraduate business students. From my experience, discussing data as it relates to critical thinking, as well as using disciplinary terms such as business analytics and business intelligence particularly resonated with faculty. One successful tactic was the use of strategic communications that introduced and reminded faculty of data sources available for research and teaching. At IUPUI, I have sent short emails regarding updates to research guides or to introduce specific data sources, as well as speaking briefly at faculty meetings. These communications created demand for instruction sessions and spurred faculty to promote data sources to students within the classroom.

Integration into I-CORE

After reviewing the undergraduate business curriculum, I found that a natural place to start integrating DIL instruction was through I-CORE, the integrated core curriculum at the Kelley School of Business. When admitted into business school after their sophomore year, students are required to take four upper-level classes their first term: finance, marketing, operations and supply chain management, and team dynamics and leadership. These courses culminate with the I-CORE project when student teams analyze the financial feasibility for either offering a new product/service or entering a new market for a local client. This project provides several opportunities to integrate DIL through library instruction and research consultations.

Building relationships and starting the conversation with I-CORE faculty was the first step. Prior to spring 2016, the library provided support for I-CORE

Table 3. Data information literacy competency mapped to National Standards for Business Education by subject/functional area (National Business Educators Association, 2013).

	Subject/functional area as specified by NBEA National Standards				
Data information literacy competency (Calzada Prado and Marzal, 2013)	Accounting	Communications	Information Technology	Management	Marketing
1. Understanding data					
1.1 What is data					
1.2 Data in society: a tool for knowledge and innovation			×		×
2. Finding and/or obtaining data					
2.1 Data sources			×		×
2.2 Obtaining data			×		×
3. Reading, interpreting, and evaluating data					
3.1 Reading and interpreting data	×	×		×	×
3.2 Evaluating data	×	×		×	×
4. Managing data					
4.1 Data and metadata collection and management	×			×	×
5. Using data					
5.1 Data handling	×	×	×		×
5.2 Producing elements for data synthesis	×	×	×		×
5.3 Ethical use of data	×	×	×		×

Note: See detailed curriculum map at: https://hdl.handle.net/1805/10823.

through a 30-minute resource briefing followed by optional research consultations with individual students and teams. During the fall of 2015, I assessed the I-CORE instruction program to understand if library instruction and research consultations were influencing student performance. I conducted a bibliographic review of student deliverables to determine whether students used business databases and/or online resources. Student performance was also assessed in the context of whether students participated in research consultations. Students who had research consultations showed stronger performance and were more likely to use business databases. I also found that the databases covered and questions asked were similar during these consultations. The faculty were receptive when I shared these findings and invited me to collaborate when updating the spring 2016 project curriculum. Project planning sessions allowed me to introduce DIL. During these discussions, I made the case for a longer instruction workshop that would allow me expand on the information literacy and DIL skills taught during research consultations to a larger base of students.

The I-CORE research workshop helped students develop skills in four of the five competencies in the Calzada Prado and Marzal framework: understanding data; finding and/or obtaining data; reading, interpreting and evaluating data; and using data. The two-hour interactive instruction session was designed to introduce students to business resources while teaching research strategies and evaluation skills. During the workshop, I introduced data through a discussion about the term "big data", following up with questions about how data could affect their project, future work, personal lives, and society (Calzada Prado and Marzal's first competency – understanding data). I reinforced this discussion as I introduced market research databases and while students explored resources.

Students were introduced to multiple databases and sources during the workshop enabling the development of the second competency – finding and/or obtaining data. Using the US Businesses database, a directory available in ReferenceUSA, students learned how to obtain information about competitors within a geographic location, locating detailed information including a sales revenue value, which in the case of private companies is estimated using a proprietary algorithm. I discussed with students the importance of searching for data definitions within databases, as well as the advantages and disadvantages of using data that is calculated through algorithms, developing the third competency – reading, interpreting, and data evaluation skills.

The US Consumers/Lifestyles database within ReferenceUSA illustrated to students how consumer

data can be accessed to generate customer-prospecting lists and analyzed to determine market trends. In this database, users can segment geographic markets by lifestyle and demographic criteria. The consumer lists provide data at the individual level, including demographics data, lifestyle preferences, interests, and contact information. This example brought home the earlier discussion about the impact data has on society and the ethical issues that exist, helping students better understand data (Calzada Prado and Marzal's first competency) and use data (the fifth competency). Students were introduced to the concepts of data producers and data consumers through discussion. They realized that business customers/consumers are data producers with each piece of information they choose to share using social media, including when they take fun personality quizzes, such as "What shoe would you be?" This conversation segued into the responsibility of data consumers, such as businesses, who need to consider ethical issues including privacy when accessing, using, and storing data.

Following the required workshop, students could schedule optional research consultations. The research consultations provided additional opportunity to teach DIL as students searched for marketing, operational, and financial data. During research consultations, I helped students develop the ability to find data by teaching them how to break down research questions into key concepts before assisting them in developing search strategies to locate information and data. Students learned to navigate subscription and free resources, such as Demographics Now and American FactFinder. Through guided discussion students learned how to use data, developing data analysis strategies to create sales forecasts and financial analysis. These discussions included the importance of documenting data collection and assumptions made during data synthesis. Students gained understanding on how this documentation, including citing sources, created credibility for their work while providing their audience the ability to find and use data integrated into their projects.

Next steps

The next area to develop of I-CORE instruction centers on data visualization with the goal of providing guidance to students in regards to tailoring messages to communicate data clearly and ethically. This aspect of using data comes into play towards the end of the project as students are preparing their final presentation and report deliverables. One potential option is the introduction of a self-paced tutorial or instructional video that discusses data visualization best practices. However, it may be more appropriate to map this instruction to an earlier course in the student's curriculum such as business communication.

Efficacy of this library instruction is being assessed in a number of ways. At the end of the research workshop students were asked to complete a short 3-2-1 formative assessment where students reflected on three things they learned, two things they found interesting, and one thing that they still had a question about. This feedback helped me understand the most engaging points of the session as well as areas of confusion. Feedback received clearly showed that students found the discussion around data, particularly the ethical use of data as well as the volume and content of data available for analysis illuminating. Of the responses, 14% of the students in spring 2016 and 23% in the summer 2016 commented on DIL topics as things they found interesting during the workshops.

In addition to workshop feedback, I have reviewed a selection of deliverables to evaluate the variety of data sources, quality of sources used, and the integration of data within the deliverables. Short-term assessments indicate that students appear to be making progress in developing DIL competencies as they work through deliverables during I-CORE. However, additional instruction is needed, particularly in regards to producing elements for data synthesis (Calzada Prado and Marzal's fifth competency). I am collecting longitudinal data on deliverables to draw definitive conclusions in regards to program effectiveness. Measuring DIL through assessment will be a moving target as it is further integrated into the curriculum.

Case study: Graduate public health education (subject librarian Heather Coates)

Opportunities in public health

Public health services affect our daily lives by preventing us from harm in numerous ways – by monitoring the water we drink, the food we eat, the air we breathe, personal care products, and the safety features in our cars, schools, and workplaces. The sheer scope of public health services combined with the diversity of backgrounds, technological systems, and stakeholders poses unique challenges for the public health profession. It is a highly multidisciplinary field, drawing on evidence, theories, and methods from sociology, communication, medicine, environmental science, informatics, and statistics, among

other fields. Data and information used to provide public health services are similarly diverse and are generated at many levels – local, state, regional, national, and international. Public health services rely on complex technological and organizational systems to gather and distribute an amalgamation of structured and unstructured data. In doing so, these systems are shaped by other systems (e.g. education, health care, transportation, etc.) as well as social, political, fiscal, and environmental issues. Finally, public health professionals face incredible diversity with regards to the professionals and communities they serve. This demands fluency in accessing, evaluating, and communicating information. Public health issues increasingly cross boundaries of education, wealth, and location. As these concerns become globalized, so too must our systems for prevention, monitoring, and response. Bolstering the DIL of the workforce has great potential to increase the quality of life for many communities by improving public health services. To that end, this case study will focus on the primary professional degree, the Master of Public Health (MPH), to identify opportunities to develop a new generation of data literate practitioners.

Public health practice

The central premise of public health practice is that people are interdependent. This is expressed in the values and beliefs underlying the Code of Ethical Practice (Public Health Leadership Society, 2002). Solving the public health issues arising from the complex interdependencies between people and their environment requires a workforce that can collaborate to build and continuously improve public health systems. *The 10 Essential Public Health Services* (Centers for Disease Control, 2014) include the following activities:

- Monitor health status to identify and solve community health problems;
- Diagnose and investigate health problems and health hazards to the community;
- Inform, educate, and empower people about health issues;
- Develop policies and plans that support individual and community health efforts; and
- Evaluate the effectiveness, accessibility, and quality of personal and population-based health services.

In a 1988 landmark report, the Institute of Medicine (IOM) noted that decision making in public health is too often driven by "crises, hot issues, and concerns of organized interest groups" (Institute of Medicine, 1988: 4). Jacobs et al. (2012) also noted the tendency for policy decisions to stem from political and media pressures, anecdotal evidence, and tradition. Evidence-based public health (EBPH) provides guidance to counteract this tendency. As in medicine, nursing, and other health professions, evidence-based practice has deeply affected public health services. Funders and national initiatives such as Healthy People 2020 encourage the use of evidence-based interventions and approaches. EBPH integrates science-based interventions with community preferences for improving population health (Jacobs et al., 2012). Public health studies are rarely as clean and controlled as randomized-controlled trials; thus, interpreting results requires keen awareness of the context in which the data were collected, particularly with respect to caveats and limitations resulting from a lack of control groups. EBPH demands that professionals can use data and information systems systematically, make decisions based on the best available peer-reviewed evidence, and disseminate what is learned.

Data, specifically surveillance of infectious diseases, is the cornerstone of public health service and research. Public Health Surveillance (PHS) is a critical tool for understanding a community's health issues. It is also an ongoing set of processes for planning and system design, data collection, data analysis, interpretation of results, dissemination and communication of information, and application of information to public health practice (Hall et al., 2012). Although advances in computing and information sciences have greatly improved surveillance since its origins in the early 20th century, there is a long way to go before PHS systems are interoperable, accessible, and effective on a global scale. Less than a quarter of respondents (22%) of a 2009 CDC survey agree that CDC surveillance systems work well for today's information technology systems (Thacker et al, 2012). Technology can act as both a facilitator and a barrier. Public health professionals will need to understand how technologies and systems shape how public health data are stored, consolidated, and disseminated (Thacker et al, 2012). These conditions – the scope of public health services, an increasing emphasis on EBPH, and public health surveillance – all demand a highly skilled workforce that can gather, handle, and use data responsibly to solve public health problems. Librarians can support the development of such a workforce by integrating DIL instruction into Public Health curricula.

Meeting professional standards for public health practice

The Council of Linkages between Academia and Public Health Practice (CLAPHP) is a collaboration of 20

Table 4. Data information literacy competencies to selected CLAPHP Core Competencies.

	CLAPHP Practice Domains		
DIL competencies	Analytical/assessment skills	Public Health Sciences skills	Policy development/ program planning skills
Understanding data	1A1, 1A2, 1A11	6A1, 6A2, 6A3	none
Finding/obtaining data	1A3, 1A4, 1A5, 1A6, 1A8, 1A11, 1A12	2A1, 2A2, 2A6, 2A10	6A4, 6A7, 6A8
Reading/interpreting/ evaluating data	1A3, 1A4, 1A6, 1A7, 1A11, 1A12	2A1, 2A2, 2A3, 2A5, 2A7	6A5, 6A6, 6A7
Managing data	1A3, 1A4	6A7	none
Using data	1A3, 1A4, 1A6, 1A11	6A3, 6A7, 6A9	2A1, 2A2, 2A3, 2A4, 2A5, 2A8, 2A9, 2A11, 2A12

Note: See full curriculum map at https://hdl.handle.net/1805/10825.

national organizations whose aim is to improve Public Health education and training, practice, and research. The Council fosters, coordinates, and monitors the connections between the academic, public health practice, and healthcare communities. Thus, they are well placed and designed to set forth a ''consensus set of skills for the broad practice of public health''. The competencies, organized by eight practice domains, are valuable for assessing workforce knowledge and skills, identifying training needs, and creating workforce development and training plans. The three practice domains selected for curriculum mapping included analytical/assessment skills, policy development/ program planning skills, and public health sciences skills. Example competencies for each domain are listed below.

- Analytical/Assessment Skills 1A6: Selects comparable data (e.g. data being age-adjusted to the same year, data variables across datasets having similar definitions);
- Policy Development/Program Planning Skills 2A5: Identifies current trends (e.g. health, fiscal, social, political, environmental) affecting the health of a community;
- Public Health Sciences Skills 6A4: Retrieves evidence (e.g. research findings, case reports, community surveys) from print and electronic sources to support decision making.

The remaining five practice domains are more relevant to information literacy instruction due to their focus on textual information. Each practice domain describes competencies at three levels or tiers of skill: entry-level or frontline staff, program management or supervisory level, and senior management or executive tier. For this case study, I focus on the first tier competencies for entry-level or frontline staff.

Mapping data literacy to professional competencies

The process of mapping the Council practice domains to DIL competencies revealed substantial opportunities for instruction to address professional standards. Generally, DIL is deeply embedded in public health practice. The Council practice domains substantially cover four of the DIL competencies. Managing data is the one DIL competency that is not explicitly discussed. It is only mentioned as it relates to ethical principles, laws and guidelines, and using information technology effectively. See Table 4.

Calzada Prado and Marzal (2013) describe understanding data as knowing what is meant by data, what types of data exist, and the role of data in society. There are two relevant Council competencies here. First is the ability to identify quantitative and qualitative data and information to assess the health of a community. The second is applying ethical principles and using information technology (IT) in all aspects of working with data. Understanding public health data is particularly challenging because the sources of evidence are diverse and include data collected in uncontrolled settings and experiments. Students must learn to recognize ways that technological and organizational systems producing public health data embed biases, limitations, and assumptions within the data.

Finding or obtaining data in public health requires professionals to navigate complex systems with variable ethical and legal constraints that guide how data are collected, managed, stored, and shared. The distribution of this DIL competency across the Council practice domains reflects its importance and difficulty

(see curriculum map at: http://hdl.handle.net/1805/10825). Professionals should be able to select or collect valid and reliable quantitative and qualitative data, describe assets and resources available to improve the health of a community, and contribute to the public health evidence base. While doing so, they must also be able to interact ethically with the systems in which data are held.

Perhaps the most important thing we can pass on to students is the value of a critical approach to reading, interpreting, and evaluating data. Again, this DIL competency is pervasive across the three Council domains. Fundamentally, it comes down to knowing the origins and characteristics of the data. Where did it come from? Who produced it? How is it described? What are its limitations and biases? Guidelines for EBPH practice provide substantial guidance for teaching this DIL competency by describing the different forms of evidence, as well as processes and strategies for using data (Brownson et al., 1999). However, EBPH is only one lens for critical evaluation. As public health practice advances, professionals will need to be able to adopt other lenses.

In contrast to the other DIL competencies, only the analytical/assessment practice domain addresses data management. Even here, it is only mentioned generally and in relation to ethical, legal, and technical issues. At the professional and program levels, no specific guidance is provided regarding how data are to be managed. However, particular biostatistics and epidemiology courses in the MPH program provide strategies for data management, usually in the context of a particular statistical program (e.g. SAS or R). This gap corresponds with trends in other disciplines to exclude practical data management skills in developing competencies or training (Carlson and Bracke, 2015; Carlson et al., 2011; Maybee et al, 2015). The lack of coverage by the Council practice domains may make it more difficult to make the case for integrating these skills into program curricula. Fortunately, public health course projects often require students to make use of data, providing an opportunity to demonstrate the relevance of data management practices.

Using data includes tasks such as preparing data for analysis, producing elements for data synthesis, and using data ethically. All three Council practice domains include competencies addressing data use. Professionals must recognize how biases, and limitations, and assumptions intrinsic to the data may affect their analysis and interpretation. Within the framework of EBPH practice, Brownson et al. (1999) provides excellent guidance for making use of data. The authors describe a seven-stage process for new practitioners to apply their knowledge in a practice setting. They articulate the types of public health interventions (emerging, promising, effective, and evidence-based) and offer classifications for evaluating and using them. These classifications lead to three types of actions indicated, depending on the evidence available. This framework makes EBPH approachable to students and professionals. As students are taught evidence-based approaches, incorporating DIL competencies related to understanding, finding and obtaining, and using data have the potential to simultaneously deepen students' DIL and proficiency with EBPH practice.

The curriculum map provides clear evidence for the value of integrating DIL instruction into public health program curricula. There is substantial overlap between the Council practice domains, principles of EBPH, and DIL competencies. While the Council practice domains and MPH program competencies discuss the evaluation and use of data holistically, they do not provide the detailed guidance needed to develop classroom instruction. Thus, mapping DIL competencies to program and professional standards reveals a critical gap in the curriculum. Next, I identified opportunities within the MPH curriculum for DIL instruction.

Integrating data information literacy into the MPH curriculum

The 12 program competencies for the Fairbanks School of Public (FSPH) Health Master of Public Health program at IUPUI are designed to communicate to students the skills that they will develop throughout their coursework. These competencies were mapped to Calzada Prado and Marzal's (2013) DIL competencies to determine alignment with academic learning goals (see curriculum map at: http://hdl.handle.net/1805/10825). In combination with syllabi, program competencies were analyzed to identify opportunities for DIL instruction within the core curriculum, which is composed of five required courses that can be taken in any order. The flexibility in the program does not lend itself to a progressive or coordinated instructional approach. Instead, I identified DIL competencies that could be addressed in each course to create a menu of classroom data activities. For each competency, I developed guiding questions for discussion as well as active learning opportunities and course assignments.

Instructors will be able to select competencies and supporting activities from an instructional menu, an abbreviated sample of which is provided in Table 5. One relatively simple approach is to incorporate discussions about data into existing topical sessions

Table 5. Opportunities for data information literacy instruction within the core MPH courses.

Data literacy competencies	General discussion	General active learning and assignments	Course-specific activities
Understanding data	What types, formats, & sources of data are used in this approach to public health? (e.g. biostatistics/epidemiology/ environmental science/health policy & management/social & behavioral sciences)	Peer training session: Students will explore core public health data topics and present to peers (small group activity).	H501: US Health Care System and Health Policy Students will explore the technical, political, social challenges of conducting research in health policy. The will also identify open sources of data.
Finding/ obtaining data	Who generates or gathers data for [the types of public health services covered in this class]?	Students will find and access public data relevant to the course project/paper, cite that data appropriately, and submit a description of your strategy and sources.	E517: Fundamentals of Epidemiology Students will explore core public health data sources, methods, and instruments/systems and present to peers (small group activity) for epidemiologic questions.
Reading/ interpreting/ evaluating data	How can you tell if data are accurate? Authoritative? Current? Relevant? Useful? How do you recognize biases, assumptions, and limitations in data?	Students will break into small groups to develop documentation for an existing dataset including: the biases, assumptions, and limitations of the data for a particular research question.	A519: Environmental Science in Public Health Students will explore and critically appraise a particular environmental health data set. They will complete an evaluative report on its use for the course project.
Managing data	Why is it important to manage data? What are the key challenges in managing public health data? What specific skills and tools can help you to manage data?	Hands-on workshop demonstrating strategies for: -defining roles & responsibilities for a project -file organization & naming, file versioning -data security & encryption -annotation and documentation (e.g. data dictionaries, codebooks, metadata, etc.)	B551: Biostatistics for Public Health Students will enhance or anonymize an existing dataset through processing, annotation, documentation, recoding, de-identification, masking, etc. OR Students will create a functional data management plan for a proposed study.
Using data	What legal, regulatory, and ethical norms affect the use of data? What are the common practices for sharing data? Reusing data? Are licenses used? Data use agreements? How are data attributed to the creators?	Students will complete the data outcomes mapping exercise. Students will describe the permissions and restrictions associated with the data they are reusing. Students will cite data in their papers and project reports.	S500: Social and Behavioral Science in Public Health Students will synthesize data from multiple sources to create a new dataset. OR Students will conduct analysis to develop evidence-based programs, interventions, policies, etc.

across the core courses. These discussions could then direct students to extra-curricular workshops or campus services for further support. A more active approach is to embed classroom activities that cover both conceptual and procedural skills to be assessed. For example, a discussion about the value of documenting data could be paired with a hands-on exercise to create a data dictionary or annotate a processing or analytical script. These activities would fit well in the core biostatistics course (B551). Given the

disciplinary variety in public health, it is crucial to develop customized examples for each course. Full integration into the curriculum requires instructors to incorporate DIL competencies into course content, assignments, and assessment.

Recent conversations with environmental health science (EHS) faculty have presented another path for DIL instruction. I approached the department chair and a former collaborator to discuss integrating DIL skills into the EHS research methods course. Unexpectedly, they responded by inviting me to redesign and teach the research methods course for their department. Both the redesign and instruction are collaborative endeavors. For example, faculty will provide guest lectures on specific research methods (e.g. interviews, observation, etc.). This collaboration was possible because the department had already identified the need for customized EHS content as well as more comprehensive coverage of research data management and scholarly communication topics. It presents an incredible opportunity to embed DIL within a core research methods course. Rather than relegating DIL to a single class session, it will be woven into the fabric of the course. For example, students will discover best practices for data management specific to each of the research methods concurrent with learning about the method. They will also examine how data practices shape and are shaped by ethical guidelines, EBPH, funding agency guidelines, and publisher requirements. When possible, instructors will identify available public health data to be used in course assignments. Transforming open and shared public health data into open educational resources will provide students with authentic learning experiences while demonstrating the value of data sharing, curation, and reuse. This fills a notable gap in the Council practice competencies. Professionals need to be aware of the common models for data sharing and exchange in public health (e.g. restricted data sets with use agreements, limited data sets, public access data, etc.) and the resulting technical and administrative procedures. Importantly, this course presents an opportunity to assess DIL instruction across multiple authentic assignments. It may also provide a model for integrating DIL into other graduate research methods courses. Currently, the pilot course is being developed for the fall semester of 2017. Though the research methods course is a great opportunity, I will continue to discuss with faculty the value of incorporating DIL into other courses using activities like those described in Table 5. In order for students to develop adequate skills to succeed as professionals, DIL competencies must be integrated throughout the program curriculum.

Beyond the classroom, providing DIL training for the existing workforce is crucial to the development of an international system of interoperable data and compatible networks. In addition to providing professionals with practical strategies and knowledge of good data practices, instruction should introduce data governance and stewardship issues that will affect their practice. This training should be strongly rooted in current practice and technologies and look forward to potential solutions for public health research and practice. Academic public health centers are often providers of continuing education, which presents an opportunity for librarians to offer DIL training to the existing workforce.

Discussion

These case studies have highlighted the importance of DIL in two professions, while demonstrating the value of integrating DIL into program curricula. Such integration can address academic standards while developing a data literate workforce. Despite distinct differences between the practice setting, types and sources of data, and context for application, we were able to identify common approaches. The business case, which largely focuses on data consumers, reveals that there is a particular need for professionals who can translate needs between content experts and information technologists who support and design information systems. In public health, professionals who can function as data creator, manager, and consumer are needed to solve 21st-century public health challenges.

The literature review indicates that employers' desire to make information actionable is driving demand for data skills in the business and public health workforce. Professional and academic standards clearly communicate the expectation that graduates have data skills (i.e. data creation, data sharing, data analytics, data mining, and data reporting). One key difference discovered is that the public health standards are primarily organized by skill domains rather than professional subject/functional area. This type of disciplinary knowledge is valuable for framing the discussions that librarians have with faculty. In both cases, when conversing with faculty it was important to illustrate how DIL instruction prepares students to succeed as professionals while supporting program learning outcomes. The selection of DIL competencies for integration should be informed by knowledge of the data roles that professionals in that particular field will hold. For instance, business professionals are largely data consumers and occasionally data handlers. In general, they are not creating

data unless they work in a specialized field such as market research. However, public health professionals often take on multiple roles with respect to data.

Analytical skills are heavily emphasized in both fields. However, the terminology, examples, and resources used vary greatly. Librarians must develop this disciplinary knowledge. Undergoing the curriculum mapping process enabled us to better understand how employer needs, professional and educational standards, and program learning outcomes intersect with regards to DIL. The knowledge gained through this process facilitated collaboration with faculty to develop course content, assignments, and assessment.

Providing concrete examples that are familiar and relevant is crucial for achieving faculty buy-in and engaging students, particularly when teaching the ethical use of data. Although there is much discussion about the ethical issues associated with business intelligence and public health surveillance data, the educational standards and professional competencies lack practical guidance. This is a clear area of need that could be addressed by collaborating with faculty to develop authentic scenarios that will resonate with students. Similarly, a dearth of practical guidance for data management is apparent in both professions. Beyond mentioning the need for these skills, guidance specific to handling, processing, analyzing, and storing data is not offered. As a result, students feel the least prepared to tackle these tasks. Therefore, data management is another area of need for collaboration between librarian and faculty to develop instruction that integrates practical strategies. However, integrating DIL instruction into course content, assignments, and assessment is challenging in already full curricula. It requires faculty to recognize the value of these skills and make room for DIL in their courses. Despite these challenges, there are options available with low barriers to entry. In business, strategically sharing information about existing resources was an effective strategy. In public health, helping instructors identify the DIL and research data management issues associated with current public health trends is an effective conversation starter. In both cases, it has been faculty interest that enabled the projects to move forward.

Finally, effective strategies for assessing data literacy instruction are still emerging. Can student performance be tied to data information literacy instruction? How can we deliver DIL instruction effectively in face-to-face, online, and hybrid formats? Can we build common start-of and end-of-program assessments? Is it possible to gather employer feedback on the preparedness of interns and new hires? How else can we determine long-term outcomes?

Implications for practice

In comparing these case studies, we identified five common steps for integrating DIL instruction into disciplinary curricula. The first is to ask stakeholders—employers, practitioners, and faculty instructors—to identify the particular skills gaps they see as most important. This could be accomplished by interviewing stakeholders, reviewing employer surveys, or analyzing professional or academic competencies. Second, map the identified skills gaps to DIL competencies and analyze course syllabi to identify content, assignments, and assessment opportunities. The third step is to convey these opportunities to faculty by situating the conversation within disciplinary topics and priorities. Once the initial steps are completed, the fourth step is to design instruction. Consider repurposing or adapting existing DIL materials rather than developing them from scratch. Several resources for materials include:

- New England e-Science Portal – Data Literacy section (http://esciencelibrary.umassmed.edu/data-literacy);
- Data Information Literacy Case Study Directory (http://docs.lib.purdue.edu/dilcs/);
- Teaching with Data (http://www.teachingwithdata.org/).

Finally, develop relevant examples that will resonate with students. We have found that using common everyday examples of data, ethical challenges, and visualizations can help students grasp data concepts, particularly when they have no previous experience in the discipline. It can often be helpful to start the discussions on a personal level, then transition to discipline specific issues.

Conclusion

Researchers in academia, employees at a manufacturing firm, and community health workers will all be tasked with using data to make decisions in their personal and professional lives. The 21st-century worker must be fluent in an increasingly complex and unstructured information world. The ultimate goal of DIL is to enable students to become lifelong learners who actively participate in the use and creation of information. To that end, these case studies examine how DIL can help students achieve both academic and professional goals. In writing up these cases, we were able to describe a tailored yet common approach for integrating DIL. We believe these cases offer useful strategies for identifying opportunities and initiating a programmatic approach to DIL instruction. We look

forward to learning how other librarians and disciplines are incorporating DIL into the classroom.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Association of College & Research Libraries (2000) *Information Literacy Competency Standards for Higher Education*. Available at: http://www.ala.org/acrl/standards/informationliteracycompetency (accessed 8 March 2016).

Association of College & Research Libraries (2015) *Framework for Information Literacy for Higher Education*. Available at: http://www.ala.org/acrl/standards/ilframework (accessed 8 March 2016).

Association to Advance Collegiate Schools of Business (2013) *Eligibility Procedures and Accreditation Standards for Business Accreditation*. Available at: http://www.aacsb.edu/accreditation/standards (accessed 9 March 2016).

Bisoux T (2016) Focus on the future: Are business schools ready for the future of work? *BizEd Magazine*. Available at: http://www.bizedmagazine.com/archives/2016/1/features/focus-on-the-future-are-business-schools-ready-for-future-work (accessed 25 February 2016).

Brownson RC, Gurney JG and Land GH (1999) Evidence-based decision making in public health. *Journal of Public Health Management* 5(5): 86–97.

Bullard KA and Holden DH (2006) Hitting a moving target: Curriculum mapping, information literacy and academe. In: *34th LOEX conference 2006 on moving targets: Understanding our changing landscapes* (eds T Valko and B Seitz), College Park, MD, USA, 4–6 May 2006, pp. 17–21. Ypsilanti, MI: Eastern Michigan University.

Burton B and Willis DA (2015) *Gartner's Hype Cycles for 2015: Five Megatrends Shift the Computing Landscape*. Available at: https://www.gartner.com/doc/3111522/gartners-hype-cycles-megatrends-shift (accessed 1 March 2016).

Calzada Prado J and Marzal MÁ (2013) Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri: International Journal of Libraries & Information Services* 63(2): 123–134.

Carlson J and Bracke M (2015) Planting the seeds for data literacy: Lessons learned from a student-centered education program. *International Journal of Digital Curation* 10(1): 95–110.

Carlson J, Fosmire M, Miller CC, et al. (2011) Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2): 629–657.

Centers for Disease Control and Prevention (2014) *10 Essential Public Health Services. National Public Health Performance Standards Program*. Available at: http://www.cdc.gov/nphpsp/essentialServices.html (accessed 24 April 2016).

Council on Linkages Between Academic and Public Health Practice (2010) *Core Competencies for Public Health Professionals*. Washington, DC: Public Health Foundation.

Cunningham NA (2003) Information competency skills for business students. *Academic BRASS* 1(1). Available at: http://www.ala.org/rusa/sections/brass/brasspubs/academicbrass/acadarchives/volume1number1/academicbrassv1%20 (accessed 8 March 2016).

Hall IH, Correa A, Yoon PW, et al. (2012) Lexicon, definitions, and conceptual framework for public health surveillance. *Morbidity and Mortality Weekly Report* 61: 10–14.

Hart Research Associates (2015) *Falling Short? College Learning and Career Success*. Report for the American Association of Colleges & Universities. Washington, DC: AAC&U. Available at: https://www.aacu.org/leap/public-opinion-research/2015-survey-falling-short (accessed 29 September 2016).

Hunt F and Birks J (2004) Best practices in information literacy. *portal: Libraries and the Academy* 4(1): 27–39.

Institute of Medicine (US) Committee of the Study of the Future of Public Health (1988) *Future of Public Health. Report*. Washington, DC: National Academy of Sciences.

Jacobs JA, Jones E, Gabella BA, et al. (2012) Tools for implementing an evidence-based approach in public health practice. *Preventing Chronic Disease* 9: 110324. DOI: 10.5888/pcd9.110324.

Klusek L and Bornstein J (2006) Information literacy skills for business careers. *Journal of Business & Finance Librarianship* 11(4): 3–21.

Koltay T (2015) Data literacy: In search of a name and identity. *Journal of Documentation* 71(2): 401–415. DOI: 10.1108/JD-02-2014-0026.

Mandinach EB and Gummer ES (2012) *Navigating the Landscape of Data Literacy: It IS Complex*. Washington, DC and Portland, OR: WestEd and Education Northwest.

Maybee C, Carlson J, Slebodnik M, et al. (2015) 'It's in the syllabus': Identifying information literacy and data information literacy opportunities using a grounded theory approach. *Journal of Academic Librarianship* 41(4): 369–376.

National Business Education Association (2013) *National Standards for Business Education: What America's Students Should Know and be Able to do in Business*. Reston, VA: NBEA.

Qin J and D'Ignazio J (2010) The central role of metadata in a science data literacy course. *Journal of Library Metadata* 10(2/3): 188–204. DOI: 10.1080/19386389.2010.506379.

Public Health Leadership Society (2002) *Principles of the Ethical Practice of Public Health*. Available at: https://www.apha.org/~/media/files/pdf/about/ethics_brochure.ashx (accessed 24 April 2016).

Ransbotham S, Kiron D and Prentice PK (2015) The talent dividend. *MIT Sloan Management Review* 56(4): 1–12.

Schield M (2004) Information literacy, statistical literacy and data literacy. *IASSIST Quarterly* 28(2/3): 6–11.

Shorish Y (2015) Data information literacy and undergraduates: A critical competency. *College & Undergraduate Libraries* 22(1): 97–106.

Sokoloff J (2012) Information literacy in the workplace: Employer expectations. *Journal of Business & Finance Librarianship* 17(1): 1–17.

Stephenson E and Caravello PS (2007) Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review* 35(4): 525–540.

Thacker SB, Qualters JR and Lee LM (2012) Public health surveillance in the United States: Evolution and challenges. *Morbidity and Mortality Weekly Report* 61: 3–9.

Wixom B, Ariyachandra T, Douglas D, et al. (2014) The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems* 34(Article 1): 1–13.

Author biographies

Katharine V. Macy is the subject liaison to the Kelley School of Business at IUPUI. Prior to joining IUPUI, she was an assistant librarian at Turchin Business Library at Tulane University. She worked over 10 years in analytical roles in private industry, prior to earning her MLIS at the University of Washington. She also has an MBA from the University of North Carolina in Chapel Hill. Her interest areas include data information literacy, information literacy, peer reference, and assessment.

Heather L. Coates is the Digital Scholarship and Data Management Librarian at the IUPUI University Library Center for Digital Scholarship. She provides research support services for the campus and is the subject liaison to the Richard M. Fairbanks School of Public Health. Her interests include evidence-based library and information practice and examining the ways that sociotechnical systems shape academic research practices, including research data management, data sharing and reuse, and research metrics. In her previous career, Heather was a research coordinator for clinical, cognitive, and behavioral psychology research teams.

International Federation of
Library Associations and Institutions
2016, Vol. 42(4) 328–341

DOI: 10.1177/0340035216678784
ifl.sagepub.com
SAGE

Abstracts

قتطفات

Modifying researchers' data management practices: A behavioural framework for library Practitioners

تعديل مُمارسات الباحثين في إدارة البيانات: إطار سلوكي لأمناء المكتبات:

Susan E Hickson, Kylie Ann Poulton, Maria Connor, Joanna Richardson, Malcolm Wolski

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

إن البيانات هي الكلمة الرائجة الجديدة في المكتبات الأكاديمية، وتفرض سياسات المكتبات تسهيل الحصول على البيانات بينما يطلب الممولون خُطط رسمية لإدارة البيانات، وتتبع المؤسسات إرشادات بشأن أفضل المُمارسات، يوضح هذا البحث مُمارسات الباحثين في إدارة البيانات، والنتائج الأولية لمشروع تم تنفيذه في جامعة جريفيث لتطبيق إطارًا مفاهيميًا (A-COM-B)؛ لدراسة سلوكيات الباحثين، يهدف المشروع إلى تشجيع اللجوء للحلول المؤسسية لإدارة البيانات البحثية، وتُشير النتائج الأولية لاستبيان قام به فريق من المكتبيين في مركز لأبحاث العلوم الاجتماعية إلى أن السلوك هو العنصر الأساسي الذي نحتاج إلى دراسته خلال وضع إستراتيجيات تعديل السلوك، ويُختتم البحث بمُناقشة المراحل القادمة من المشروع والتي تشمل جمع وتحليل أكثر للبيانات، وتطبيق الإسترتيجيات المرجوة، وأنشطة لتقييم مدى التعديل اللازم للمُمارسات الموجودة حاليًا وغير المرغوب فيها.

Research data services: An exploration of requirements at two Swedish universities

الخدمات البحثية: استكشاف المُتطلبات في جامعتين سويديتين:

Monica Lassi, Maria Johnsson, Koraljka Golub

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

يتناول البحث دراسة استكشافية في جامعتين سويديتين لاحتياجات الباحثين ؛ كي يتمكنوا من إدارة البيانات بكفاءة، وقد تمت هذ الدراسة لجمع معلومات من أجل تطوير الخدمات البحثية، وتم إجراء لقاءات مع 12 باحث من مُختلف المجالات، منها الأحياء، والدراسات الثقافية، والاقتصاد، والدراسات البيئية، والجغرافيا، والتاريخ، واللغويات، والإعلام، وعلم النفس، وقد تم بناء محتوى هذه اللقاءات على أساس مجموعة مُعالجة البيانات التي طورتها جامعة بيردو، وشملت أسئلة عن واصفات البيانات ورؤوس الموضوعات، وتُشير النتيجة الأولية لتحليل ناتج هذه اللقاءات أن مُمارسات إدارة البيانات تختلف كثيرًا من شخص لآخر مما يؤثر بالتالي على الخدمات البحثية، وأشارت الإجابات عن واصفات البيانات ورؤوس الموضوعات إلى الحاجة إلى توجيهات تُمكننا من وضع واصفات البيانات الكافية للموضوعات.

'Essentials 4 Data Support': five years' experience with data management training

أساسيات دعم البيانات: خمس سنوات خبرة في التدريب على دعم البيانات:

Ellen Verbakel, Marjan Grootveld

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

يصف هذا البحث دورة عن إدارة البيانات البحثية لدعم العاملين من المكتبيين والقائمين على تكنولوجيا المعلومات، يُعرف أصحاب البحث الأشكال الثلاثة للدورة ويصفون التدريب تفصيلاً، وقد شارك 170 شخص في هذا التدريب على مدار السنوات الماضية، والذي يجمع بين اجتماعات الفعلية وأخرى عبر الإنترنت، وتهدف الدورة إلى دعم المُشاركين في تقوية مُختلف مهاراتهم واكتساب معارف جديدة تُشعرهم بثقة في أنفسهم تُمكنهم من توجيه وتدريب الباحثين. إن التفاعل بين الطلاب جزءًا لا يتجزأ من التدريب، لأننا ننظر له كأداة قيمة لتنمية الشبكة المهنية، وقد وضعت الدورة لنفسها تحديًا جديدًا: وقدمت تدريبين إضافيين بناءً على طلب المُشاركين إلى جانب الدورات المُعتادة، ويُختتم البحث بتوضيح بعض الواجبات اللازمة لمثل هذه الدورات المُكثفة.

Research Data Services at ETH-Bibliothek

خدمات البيانات البحثية في ETH-Bibliothek:

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

حوكمة البيانات، ومحو أمية البيانات وإدارتها:

Tibor Koltay

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

إن إدارة البيانات ومحو الأمية المعلوماتية حجري أساس بالنسبة لمعارف العاملين في مجال المعلومات، القائمين على دعم البحث كثيف البيانات، وكلاهما يتناول جودة البيانات وإدارة البيانات البحثية، يُساعد تطبيق حوكمة البيانات على إدارة البيانات البحثية ومحو الأمية المعلوماتية على تحديد مجالات اتخاذ القرار والمُساءلة بشأنها، إن تبني مبدأ حوكمة البيانات مليء بالمزايا؛ لأنه قائم على عمليات مُكررة ذات معايير مُصممة لضمان شفافية العمل المُرتبط بالبيانات وتقليل التكلفة، كما أنه أيضًا مُفيد لأنه يرجع لقواعد، وسياسات، ومعايير، ومُساءلة، وطرق تنفيذ؛ لذا فقد حظت باهتمام أكبر في الشركات ويمتلك المكتبيون بالفعل بعض المهارات اللازمة، وخاصةً تلك المُرتبطة بخدمات البيانات البحثية والبيانات الكبيرة.

Data information literacy instruction in Business and Public Health: Comparative case studies

تدريس محو الأمية المعلوماتية في مجالي التجارة والصحة العامة: دراسات حالة مُقارنة:

Katharine Macy, Heather Coates

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

يحتاج أصحاب الأعمال إلى قوى عاملة قادرة على استخدام البيانات لخلق معلومات قيمة، وهو ما يحتاج من الطلاب العمل على مهاراتهم في محو الأمية المعلوماتية والتي تُمكنهم البحث والتصفح والإضافة في ظل عالمنا المعلوماتي المُعقد، يبحث هذا المقال أهمية إدراج محو الأمية المعلوماتية في المناهج، وخاصةً في مجالات التجارة والصحة العامة، ويطرح إستراتيجيات لتحقيق ذلك، إن هذه الدراسة هي دراسة حالة مُقارنة بين برامج دراسة التجارة والماجستير في الصحة العامة بجامعتي إنديانا وبيردو، تكشف هذه الدراسات العديد من المُمارسات التي تؤثر على برامج العلوم الإنسانية والصحة.

تُعد إدارة البيانات البحثية خلال دورة حياتها أساسًا لمُشاركة البيانات بكفاءة وحفظها جيدًا على المدى الطويل أيضًا، يُلخص المقال خدمات البيانات والمنهج العام لإدارة البيانات في مكتبة ETH-Bibliothek، مكتبة ETH الأساسية بزيوريخ، أكبر جامعة تقنية في سويسرا. تدعم المكتبة الأسئلة المفاهيمية، وتُغطي الخدمات التي تُقدمها المكتبة دورة حياة البيانات كاملةً، كما تُقدم التدريب والخدمات المُتعلقة بنشر البيانات والحفظ طويل المدى، ومع استمرار دور البيانات البحثية الهام في ما يطلبه الباحثون والممولون وأهميتها للمناهج والمُمارسات العلمية الجيدة، تعمل ETH-Bibliothek على التعاون عن قرب مع الباحثين؛ لدعم عملية التعلُّم المُتبادل ومواجهة مع التحديات الجديدة.

Beyond the Matrix: Repository Services for Qualitative Data

ما وراء المنشأ: خدمات المُستودع للبيانات الكيفية:

Sebastian Karcher, Dessislava Kirilova, Nicholas Weber

العدد رقم 42،2 من مجلة الإفلا المُتخصصة:

المُلخص:

يُقدم مستودع البيانات الكيفية (QDR) البنية التحتية والتوجيه اللازمين؛ لمُشاركة وإعادة استخدام البيانات الرقمية المُستخدمة في الأبحاث الاجتماعية الكيفية مُتعددة المناهج، نوضح في هذا البحث بعض التجارب الأولى للمُستودع في تقديم خدمات مُخصصة لاستعادة البيانات الكيفية البحثية، ونُركز على جهود المستودع في مواجهة تحديين رئيسيين في مُشاركة البيانات الكيفية، يتعلق التحدي الأول بالقيود المفروضة على مُشاركة البيانات من أجل حماية الناس وهوياتهم والامتثال لقوانين حقوق الطبع والنشر، أما المجموعة الثانية من التحديات فتتناول السمات الفريدة للبيانات الكيفية وعلاقتها بالمطبوعات، كما نوضح طريقة مُبتكرة للعناية بالمطبوعات الدراسية، نخرج من البحث بـ"مُلحق الشفافية" والذي يُمكننا من مُشاركة "البيانات الجزئية" (Moravcsik et'al., 2013)، ونختتم البحث بوصف اتجاهات خدمات المستودع المُستقبلية في عمل أرشيف بالبيانات الكيفية، ومُشاركتها، وإعادة استخدامها.

Data governance, data literacy and the management of data quality

摘要

Modifying researchers' data management practices: A behavioural framework for library practitioners

改变研究者的数据管理行为：一个图书馆从业人员的行为框架

Susan E Hickson, Kylie Ann Poulton, Maria Connor, Joanna Richardson, Malcolm Wolski

国际图联杂志, 42-4, 253-265

摘要：

数据是学术图书馆中新的流行语，因为政策规定数据必须是开放的和可访问的，资助者需要正式的数据管理计划，机构正在实施有关最佳实践的指南。考虑到研究人员对数据管理实践的关注，本文报告了在格里菲斯大学(Griffith

University)正在进行的项目的初步结果，以应用概念(A-COM-B)框架来理解研究人员的行为。该项目旨在鼓励使用机构解决方案进行研究数据管理。根据社会科学研究中心一个图书馆员小组的访谈，初步结果表明，态度是在设计干预策略以修改行为时需要解决的关键因素。本文最后讨论了项目的下一阶段涉及进一步的数据收集和分析，目标战略的实施以及评估对当前不良行为的修改程度的活动。

Research data services: An exploration of requirements at two Swedish universities

研究数据服务：探索两所瑞典大学的要求

Monica Lassi, Maria Johnsson, Koraljka Golub

国际图联杂志, 42-4, 266-277

摘要：

本文报告了对两个瑞典大学研究人员有效研究数据管理需求的探索性研究，以便为研究数据服务的持续发展提供信息。 来自不同领域的十二位研究人员接受了采访，包括生物学，文化研究，经济学，环境研究，地理，历史，语言学，媒体和心理学。 访谈是由在普渡大学开发的数据固化概况工具包指导的，并附加了关于主题元数据的问题。 初步分析表明，受访者的研究数据管理实践差异很大，因此对研究数据服务的影响。 关于主题元数据的附加问题表明服务的需要，指导研究人员用足够的元数据描述其数据集。

'Essentials 4 Data Support': Five years' experience with data management training

"Essentials 4数据支持"：五年数据管理培训的经验

Ellen Verbakel, Marjan Grootveld

国际图联杂志, 42-4, 278-283

摘要：

本文介绍了一个针对图书馆员和IT人员等支持人员的研究数据管理课程。 作者教导参与者，介绍三种课程格式，并更详细地描述培训。 在过去的几年里有170多人参加了这次培训。 它结合了丰富的在线信息和面对面的会议。 课程的目的是支持参与者加强各种技能和获得知识，使他们有信心支持，建议和培训研究人员。 学生之间的互动嵌入在培训的结构中，因为我们认为它是开发专业网络的有价值的工具。 最近，该课程面临一个新的挑战：除了常规课程，还根据要求提供了几次内部培训。 本文结束了这种紧凑型培训的关键组分配的描述。

Research Data Services at ETH-Bibliothek

苏黎世联邦理工学院图书馆的研究数据服务

Ana Sesartic, Matthias Töwe

国际图联杂志, 42-4, 284-291

摘要：

研究数据在其整个生命周期的管理是有效数据共享和有效长期保存数据的关键先决条件。本文总结了苏黎世联邦理工学院图书馆(ETH-Bibliothek)的数据服务和数据管理的总体方法，苏黎世联邦理工学院图书馆是瑞士最大的技术大学的主图书馆。苏黎世联邦理工学院的服务提供商提供的服务涵盖了整个数据生命周期。图书馆为概念问题提供支持，并提供培训以及关于数据出版和长期保存的服务。 随着研究数据管理在研究人员和资助者的要求以及课程和良好的科学实践中继续发挥越来越重要的作用，苏黎世联邦理工学院图书馆正在与研究人员建立密切的合作，以促进相互学习和面对新的挑战。

Beyond the matrix: Repository services for qualitative data

超越黑客帝国：定型数据的存储库服务

Sebastian Karcher, Dessislava Kirilova, Nicholas Weber

国际图联杂志, 42-4, 292-302

摘要：

定性数据存储库(QDR)为在定性和多方法社会查询中使用的数字数据的共享和重用提供基础设施和指导。在本文中，我们描述了一些存储库的早期经验，提供专门为定性研究数据的策划而开发

的服务。我们专注于QDR的努力，以解决质量数据共享的两个关键挑战。第一个挑战涉及数据共享的限制，以保护人类参与者及其身份并遵守版权法。第二组挑战涉及定性数据的独特特征及其与已发表文本的关系。我们描述了一种注释学术出版物的新方法，产生了一种允许共享这种“粒度数据”的”透明度附录”(Moravcsik等，2013)。我们最后描述QDR的服务质量数据归档，共享和重用的未来方向。

Data governance, data literacy and the management of data quality

数据治理、数据素养和数据质量管理

Tibor Koltay

国际图联杂志, 42-4, 303-312

摘要：

数据治理和数据素养是信息专业人员知识库中的两个重要组成部分，参与支持数据密集型研究，以及解决数据质量和研究数据管理。

将数据治理应用于研究数据管理流程和数据素养教育有助于描述决策领域和确定决策的问责制。采用数据治理是有利的，因为它是一种基于标准化，可重复的流程的服务，旨在使数据相关流程的透明度和成本降低。它也是有用的，因为它指的是规则，政策，标准;决策权;责任和执行方法。因此，虽然它在企业环境中受到更多的关注，并且与其相关的一些技能已经由图书馆员拥有，但是数据治理的知识是研究数据服务的基础，特别是在所有级别的研究数据服务中都是如此，并且是适用的到大数据。

Data information literacy instruction in Business and Public Health: Comparative case studies

商业与公共卫生领域的数据信息素养指导：比较案例研究

Katharine Macy, Heather Coates

国际图联杂志, 42-4, 313-327

摘要：

雇主需要一个能够使用数据来创建可操作信息的员工。 这要求学生开发数据信息素养能力，使他们能够在日益复杂的信息世界中导航和创造意义。 本文研究了为什么数据信息素养应该被纳入计划课程，特别是在商业和公共卫生的情况下，并提供如何实现它的战略。 我们将此作为本科生的比较案例研究和印第安纳大学 - 普渡大学印第安纳波利斯的公共卫生计划硕士。 这些案例研究揭示了适用于社会和健康科学计划的实践的几个影响。

Sommaires

(Modifying researchers' data management practices: A behavioural framework for library practitioners)

Modifier les pratiques de gestion des données des chercheurs: un cadre comportemental pour les professionnels des bibliothèques

Susan E Hickson, Kylie Ann Poulton, Maria Connor, Joanna Richardson, Malcolm Wolski

IFLA Journal, 42-4, 253-265

Résumé :

Les données sont la dernière notion à la mode dans les bibliothèques universitaires, alors que la politique exige qu'elles soient ouvertes et accessibles, que les bailleurs de fonds demandent des plans formels de gestion des données et que les institutions appliquent des directives relatives aux bonnes pratiques. Étant donné les préoccupations concernant les pratiques de gestion des données des chercheurs, cet article rend compte des constatations initiales d'un projet mené à l'université Griffith, consistant à appliquer un cadre conceptuel (A-COM-B) pour comprendre le comportement des chercheurs. Ce projet vise à encourager l'usage de solutions institutionnelles pour gérer les données de recherche. Les résultats préliminaires, basés sur des interviews menées par une équipe de bibliothécaires dans un centre de recherche en sciences sociales, indiquent que l'attitude est l'élément clé dont il faut tenir compte pour concevoir des stratégies d'intervention visant à modifier le

comportement. L'article conclut par une discussion sur les étapes suivantes du projet, qui comprennent la poursuite de la collecte et de l'analyse de données, la mise en œuvre de stratégies ciblées et des actions pour déterminer l'étendue des modifications à appliquer aux pratiques actuelles indésirables.

(Research data services: An exploration of requirements at two Swedish universities)

Services de données de recherche : étude des critères dans deux universités suédoises

Monica Lassi, Maria Johnsson, Koraljka Golub

IFLA Journal, 42-4, 266-277

Résumé :

L'article rend compte d'une étude exploratoire sur les besoins des chercheurs de disposer d'une gestion efficace des données de recherche dans deux universités suédoises, étude menée afin d'obtenir des informations sur l'évolution constante des services de données de recherche. Douze chercheurs de diverses disciplines, y compris biologie, études culturelles, économie, études environnementales, géographie, histoire, linguistique, médias et psychologie, ont été interviewés. Les interviews étaient structurées sur la base de l'outil Profils de Conservation des Données mis au point à l'université Purdue, et comportaient des questions complémentaires au sujet des métadonnées. L'analyse préliminaire indique que les pratiques de gestion des données de recherche varient beaucoup selon les personnes interrogées, ce qui est par conséquent aussi le cas des implications pour les services de données de recherche. Les questions complémentaires sur les métadonnées indiquent qu'il existe un besoin de services pour guider les chercheurs à décrire leurs ensembles de données à l'aide de métadonnées adéquates.

('Essentials 4 Data Support': five years' experience with data management training)

Cours « Essentials 4 Data Support » : cinq ans d'expérience de la formation à la gestion des données

Ellen Verbakel, Marjan Grootveld

IFLA Journal, 42-4, 278-283

Résrumé :

Cet article décrit un cours de gestion des données de recherche destiné au personnel de soutien tel que bibliothécaires et personnel informatique. Les auteures, qui encadrent les participants, ont conçu ces trois formats de cours et décrivent la formation plus en détails. Ces dernières années, plus de 170 personnes ont suivi cette formation. Elle associe une abondance d'informations en ligne avec des rencontres face à face. L'objectif de ce cours est d'aider les participants à renforcer diverses compétences et à acquérir des connaissances de façon à ce qu'ils se sentent en mesure de soutenir, conseiller et former des chercheurs. L'interaction des étudiants fait partie intégrante de la formation, dans la mesure où elle est considérée comme un instrument précieux pour développer un réseau professionnel. Récemment, le cours s'est attaqué à un nouveau défi : en plus des cours réguliers, quelques formations sur site ont été organisées à la demande. L'article conclut avec une description des travaux de groupe, essentiels dans le cadre de formations compactes de ce type.

(Research Data Services at ETH-Bibliothek)

Services de données de recherche à la Bibliothèque de l'École polytechnique fédérale de Zurich

Ana Sesartic, Matthias Töwe

IFLA Journal, 42-4, 284-291

Résumé :

La gestion des données de recherche tout au long de leur cycle de vie est une condition préalable essentielle à un partage efficace des données ainsi qu'à une conservation efficiente des données sur le long terme. Cet article résume les services de données et l'approche globale de la gestion des données pratiquée actuellement à la Bibliothèque de l'École polytechnique fédérale de Zurich, la plus grande université technique de Suisse. Les services proposés par des prestataires au sein de cette bibliothèque portent sur l'ensemble du cycle de vie des données. La bibliothèque offre un soutien en ce qui concerne les questions d'ordre conceptuel, propose des formations et des services sur la publication des données et leur conservation sur le long terme. Étant donné que la gestion des données de recherche joue un rôle de plus en plus important, aussi bien du point de vue des exigences des chercheurs et des bailleurs de fonds qu'en ce qui concerne les programmes de cours et les bonnes pratiques scientifiques, la Bibliothèque de l'École polytechnique fédérale de Zurich met en place des collaborations étroites avec des chercheurs afin de promouvoir un processus mutuel d'apprentissage et de s'attaquer à de nouveaux défis.

(Beyond the Matrix: Repository Services for Qualitative Data)

Au-delà de la matrice : Services de dépôt numérique de données qualitatives

Sebastian Karcher, Dessislava Kirilova, Nicholas Weber

IFLA Journal, 42-4, 292-302

Résumé :

Le dépôt de données qualitatives offre une infrastructure et des lignes directrices pour le partage et la réutilisation des données numériques utilisées lors d'enquêtes sociales qualitatives basées sur plusieurs méthodes. Dans cet article, nous décrivons quelques-unes des premières expériences du dépôt dans la fourniture de services conçus spécifiquement pour conserver les données qualitatives de recherche. Nous étudions les efforts du dépôt de données qualitatives pour relever deux défis majeurs pour le partage de données qualitatives. Le premier défi porte sur les contraintes en matière de partage des données, afin de protéger les individus participants et leurs identités et de respecter la législation sur le copyright. Le second défi porte sur les caractéristiques uniques des données qualitatives et leur rapport avec le texte publié. Nous décrivons une nouvelle méthode d'annotation des publications savantes, résultant en un « appendice de la transparence » qui permet de partager de telles « données granulaires » (Moravcsik et'al., 2013). L'article conclut en décrivant les orientations futures des services de dépôt de données qualitatives pour archiver, partager et réutiliser ce type de données.

(Data governance, data literacy and the management of data quality)

Gouvernance des données, culture des données et gestion de la qualité des données

Tibor Koltay

IFLA Journal, 42-4, 303-312

Résumé :

La gouvernance des données et la culture des données sont deux éléments constitutifs fondamentaux de la base de connaissances des professionnels de l'information qui sont impliqués dans le soutien aux recherches et utilisent des données de façon intensive, et ces deux aspects portent sur la qualité des données et la gestion des données de recherche.

Appliquer la gouvernance des données aux procédures de gestion des données de recherche et à la formation à la culture des données aide à délimiter des domaines de décision et à déterminer les responsabilités pour la prise de décision. Adopter la gouvernance des données présente des avantages, dans la mesure où c'est un service basé sur des processus standardisés qui peuvent être répétés et qu'elle est conçue pour permettre la transparence des procédures relatives aux données et réduire les coûts. Elle est aussi utile dans la mesure où elle se réfère aux réglementations, politiques et normes, aux droits de décision, aux responsabilités ainsi qu'aux méthodes d'exécution. Par conséquent, bien qu'elle reçoive plus d'attention dans le monde des entreprises et que les bibliothécaires détiennent déjà certaines des compétences qui s'y rapportent, une connaissance de la gouvernance des données est essentielle pour les services de données de recherche, en particulier parce qu'elle se manifeste à tous les niveaux des services de données de recherche et est applicable aux métadonnées.

(Data information literacy instruction in Business and Public Health : Comparative case studies)

Formation à la maîtrise des données dans le secteur des affaires et la santé publique : études de cas comparatives

Katharine Macy, Heather Coates

IFLA Journal, 42-4, 313-327

Résumé :

Les employeurs ont besoin d'un personnel capable d'utiliser les données pour créer des informations exploitables. Cela exige des étudiants qu'ils maîtrisent bien les compétences de gestion des données pour pouvoir évoluer dans un univers de l'information de plus en plus complexe et lui donner un sens. Cet article examine pourquoi la maîtrise des données devrait être intégrée aux programmes de cours, en particulier en ce qui concerne le secteur des affaires et la santé publique, et propose des stratégies sur la façon d'y parvenir. Nous abordons ce sujet comme une étude de cas comparative dans le cadre de programmes de premier cycle universitaire pour le secteur des affaires et programmes de maîtrise pour les formations en santé publique à l'université d'Indiana, université de Purdue à Indianapolis. Ces études de cas révèlent plusieurs conséquences pratiques qui s'appliquent à tous les niveaux des programmes de sciences sociales et sciences de la santé.

Zusammenfassungen

(Modifying researchers' data management practices: A behavioural framework for library practitioners)

Den Umgang mit dem Datenmanangement von Wissenschaftlern verändern: Verhaltensrichtlinien für Bibliothekmitarbeiter

Susan E Hickson, Kylie Ann Poulton, Maria Connor, Joanna Richardson, Malcolm Wolski

IFLA-Journal, 42-4, 253-265

Zusammenfassung:

Daten sind das neue Zauberwort in Universitätsbibliotheken, da nach dem Willen der Politik Daten offen und zugänglich sein müssen, Kostenträger formale Datenmanagementpläne verlangen und Institute Best-Practice-Richtlinien implementieren. Angesichts der Sorgen in Bezug auf den Umgang mit dem Datenmanagement bei

Wissenschaftlern berichtet diese Arbeit von den ersten Ergebnissen eines Projekts der Griffith University über die Anwendung eines konzeptuellen (A-COM-B) Rahmenwerks zum besseren Verständnis der Vorgehensweisen von Wissenschaftlern. Das Projekt möchte für die Anwendung institutioneller Lösungen für das Forschungsdatenmanagement werben. Ausgehend von Interviews, die ein Team von Bibliotheksmitarbeitern in einem sozialwissenschaftlichen Forschungszentrum durchgeführt hat, zeigen die vorläufigen Ergebnisse, dass die persönliche Einstellung das Schlüsselelement für die Entwicklung von Interventionsstrategien ist, die das Verhalten beeinflussen sollen. Die Arbeit schließt mit einer Besprechung der nächsten Projektphasen, in denen weitere Daten gesammelt und analysiert werden sollen und die außerdem die Implementierung gezielter Strategien und einer Maßnahme zur Bewertung des Umfangs vorsehen, in dem sich die bisherigen unerwünschten Vorgehensweisen geändert haben.

(Research data services: An exploration of requirements at two Swedish universities)

Service für Forschungsdaten: Untersuchung der Anforderungen an zwei schwedischen Universitäten

Monica Lassi, Maria Johnsson, Koraljka Golub

IFLA-Journal, 42-4, 266-277

Zusammenfassung:

Die Arbeit befasst sich mit einer Forschungsstudie nach den Bedürfnissen von Wissenschaftlern in Bezug auf ein effektives Forschungsdatenmanagement, die den aktuellen Entwicklungsstand von Dienstleistungen für Forschungsdaten feststellen sollte. Dazu wurden zwölf Wissenschaftler unterschiedlicher Disziplinen, darunter Biologie, Kulturwissenschaften, Wirtschaft, Umweltwissenschaften, Geographie, Geschichte, Sprachwissenschaft, Medien und Psychologie, befragt. Die Interviews wurden anhand des von der Purdue University entwickelten „Data Curation Profiles Toolkit" und zusätzlichen Fragen zum Aspekt Metadaten durchgeführt. Nach den vorläufigen Ergebnissen lässt sich schließen, dass die Respondenten ganz unterschiedlich mit dem Forschungsdatenmanagement umgingen, was in der Folge auch zu unterschiedlichen Implikationen für Dienstleistungen im Bereich Forschungsdaten führt. Die Zusatzfragen zum Thema Metadaten weisen auf den Bedarf nach einem Angebot hin, das die Wissenschaftler bei der Beschreibung ihrer Datensätze mit adäquaten Metadaten unterstützt.

('Essentials 4 Data Support': five years' experience with data management training)

'Essentials 4 Data Support': Fünf Jahre Erfahrung mit Datenmanagement-Schulungen

Ellen Verbakel, Marjan Grootveld

IFLA-Journal, 42-4, 278-283

Zusammenfassung:

Dieser Artikel beschreibt einen Kurs im Bereich Forschungsdatenmanagement für unterstützende Kräfte wie Bibliothekare und IT-Mitarbeiter. Die Autoren, die die Teilnehmer coachen, bieten drei Kursformate an und liefern eine ausführliche Beschreibung der Schulung. In den letzten Jahren haben über 170 an dieser Schulung teilgenommen. Die Schulung ist eine Kombination von umfassenden Online-Informationen und Face-to-Face-Meetings. Ziel der Schulung ist es, die Teilnehmer bei der Entwicklung unterschiedlicher Fähigkeiten zu unterstützen und ihnen Wissen anzureichen, sodass sie für die Unterstützung, Beratung und Schulung von Wissenschaftlern gerüstet sind. Die Schulung setzt auch auf die Interaktion unter den Teilnehmern, da diese als wertvolles Instrument für den Aufbau eines professionellen Netzwerks angesehen wird. Seit kurzem wurde ein neues Angebot mit ins Programm genommen: Zusätzlich zu den regulären Kursen wurden auf Anfrage auch mehrere In-house-Schulungen durchgeführt. Die Arbeit schließt mit einer

Beschreibung der wichtigsten Anforderungen an solche Kompaktkurse.

(Research Data Services at ETH-Bibliothek)

Forschungsdatenservices an der ETH-Bibliothek

Ana Sesartic, Matthias Töwe

IFLA-Journal, 42-4, 284-291

Zusammenfassung:

Das Management von Forschungsdaten im Laufe ihres Lebenszyklus ist sowohl eine wichtige Voraussetzung für den effektiven Datenaustausch als auch für eine effiziente langfristige Aufbewahrung dieser Daten. Dieser Artikel stellt die Datenservices und den allgemeinen Umgang mit dem Datenmanagement vor, wie er zurzeit bei der ETH-Bibliothek, der Hauptbibliothek der ETH Zürich und der größten technischen Universität der Schweiz, praktiziert wird. Das von den Serviceanbietern der ETH Zürich bereitgestellte Angebot umfasst den gesamten Lebenszyklus der Daten. Die Bibliothek bietet Unterstützung in Bezug auf konzeptuelle Fragen, Schulungen und Services in Bezug auf die Datenveröffentlichung und die langfristige Aufbewahrung. Da das Forschungsdatenmanagement eine zunehmend wichtige Rolle bei den Anforderungen von Wissenschaftlern und Geldträgern sowie in Bezug auf Curricula und gute wissenschaftliche Praxis spielt, setzt die ETH-Bibliothek auf eine enge Zusammenarbeit mit Wissenschaftlern, um einen gegenseitigen Lernprozess in Gang zu setzen und neue Herausforderungen zu bewältigen.

(Beyond the Matrix: Repository Services for Qualitative Data)

Jenseits der Matrix: Repositorien-Services für qualitative Daten

Sebastian Karcher, Dessislava Kirilova, Nicholas Weber

IFLA-Journal, 42-4, 292-302

Zusammenfassung:

Das Repositorium für qualitative Daten (QDR) liefert die Infrastruktur und Richtschnur für den Austausch und die Wiederverwendung digitaler Daten, die in der qualitativen und Multi-Methoden-Sozialforschung verwendet wurden. In dieser Arbeit beschreiben wir einige der frühen Erfahrungen der Repositorien mit Services, die speziell für die Aufbewahrung qualitativer Forschungsdaten entwickelt wurden. Dabei legen wir den Schwerpunkt auf den Versuch von QDR, die zwei wichtigsten Herausforderungen für den Austausch von qualitativen Daten zu meistern. Die erste Herausforderung bezieht sich auf Einschränkungen des Datenaustausches im Rahmen des Schutzes menschlicher Teilnehmer und ihrer Identitäten und der Einhaltung von Urheberschutzgesetzen. Die zweite Gruppe der Herausforderungen bezieht sich auf die besonderen Eigenschaften der qualitativen Daten und ihre Beziehung zum veröffentlichten Text. Wir beschreiben ein neues Verfahren zur Kommentierung wissenschaftlicher Publikationen mit dem Ziel, einen „Transparenz-Anhang" zu schaffen, mit dem ein Austausch solcher „granularen Daten" möglich ist (Moravcsik et'al., 2013). Abschließend beschreiben wir künftige Anweisungen für QDR-Services für die Archivierung, den Austausch und die Wiederverwendung qualitativer Daten.

(Data governance, data literacy and the management of data quality)

Data-Governance, Datenkenntnisse und das Management von Datenqualität

Tibor Koltay

IFLA-Journal, 42-4, 303-312

Zusammenfassung:

Data-Governance und Datenkenntnisse sind zwei wichtige Bausteine der Wissensgrundlage von *Information Professionals*, die mit der Unterstützung von datenintensiver Forschung befasst sind, und beziehen sich auf die Datenqualität und das Forschungsdatenmanagement.

Die Anwendung von Data-Governance auf Forschungsdatenmanagement-Prozesse und das Training von Datenkenntnissen ist vorteilhaft für die Abgrenzung von Entscheidungsfeldern und die Definition der Verantwortung für die Entscheidungsfindung. Der Einsatz von Data-Governance zahlt sich aus, das es sich um einen Service handelt, der auf standardisierten, wiederholbaren Prozessen gründet und der die Transparenz datenbezogener Prozesse und die Kostensenkung fördern soll. Ferner ist er sinnvoll, weil er sich auf Regeln, Leitlinien, Standards, Entscheidungsrechte, Verantwortlichkeiten und Durchsetzungsmethoden bezieht. Bislang fand Data-Governance eher in Unternehmen Beachtung und einige wenige Bibliothekaren besitzen bereits ansatzweise Erfahrung mit ihren Aspekten. Doch Kenntnisse zur Data-Governance sind grundlegend für Forschungsdatenservices, besonders, da sie auf allen Ebenen der Forschungsdatenservices auftreten und auch auf Big Data anwendbar sind.

(Data information literacy instruction in Business and Public Health: Comparative case studies)

Anweisungen zu Dateninformationskenntnissen in der Wirtschaft und im Gesundheitswesen: Vergleichende Fallstudien

Katharine Macy, Heather Coates

IFLA-Journal, 42-4, 313-327

Zusammenfassung:

Arbeitgeber brauchen Mitarbeiter, die in der Lage sind, Daten auszuwerten, um daraus Informationen für künftiges Handeln zu beziehen. Studenten sollten deshalb Kompetenzen im Umgang mit Dateninformationen aufbauen, sodass sie durch riesige Datenmengen navigieren und ihnen in einer zunehmend komplexen Welt der Informationen Bedeutung beimessen können. Dieser Artikel untersucht, warum Dateninformationskompetenz in Studienordnungen aufgenommen werden sollte, besonders in den Bereichen Wirtschaft und Gesundheitswesen, und bietet Strategien für die Umsetzung dieser Forderung an. Wir nähern uns dieser Frage in Form einer vergleichenden Fallstudie mit dem Programm für das Grundstudium der Wirtschaftswissenschaften und dem Master-Programm für Gesundheitswissenschaften an der Indiana University – Purdue University Indianapolis, USA. Diese Fallstudien zeigen mehrere Folgen für die Praxis in Studiengängen der Sozial- und Gesundheitswissenschaften.

Рефераты статеи

(Modifying researchers' data management practices: A behavioural framework for library practitioners)

Преобразование используемых исследователями методов управления данными: Модель поведения для практикующих библиотекарей

Сьюзан Е Хиксон, Кайли Энн Поултон, Мария Коннор, Джоанна Ричардсон, Малколм Вольски

IFLA Journal, 42-4, 253-265

Аннотация:

Слово "данные" ныне вошло в моду в академических библиотеках, в условиях, когда в соответствии принципами работы данные должны быть открыты и доступны, когда спонсоры требуют формальных планов управления данными, и учреждения основывают реализацию своих ключевых задач на передовых методиках. В рамках рассмотрения методов управления данными, используемых исследователями, настоящая работа сообщает о первоначальных результатах проекта, начатого в Университете Гриффита, целью которого является использование концептуальной схемы (A-COM-B) для понимания моделей поведения исследователей. Проект направлен на стимулирование использования разработанных учебными заведениями методов управления исследовательскими данными. Предварительные результаты, полученные на основе опросов, проведенных группой библиотекарей в центре исследования обществоведения, показывают, что именно точка зрения является тем ключевым элементом, к которому необходимо апеллировать в рамках разработки стратегий вмешательства, направленных на изменение поведения. В завершение работы излагаются выводы относительно следующих этапов проекта, которые включают в себя дальнейшие сбор и анализ данных, реализацию целевых стратегий, а также деятельность по оценке масштабов изменений текущих нежелательных методов работы.

(Research data services: An exploration of requirements at two Swedish universities)

Услуги в сфере научных данных: Изучение требований в двух университетах Швеции

Моника Ласси, Мария Джонссон, Корайка Голуб

IFLA Journal, 42-4, 266-277

Аннотация:

В данной работе приводится отчет о поисковом изыскании в отношении требований исследователей к эффективному управлению исследовательскими данными в двух университетах Швеции, которое проводилось с целью получения информации в рамках развития услуг в сфере научных данных. Были опрошены двенадцать научных сотрудников из различных областей, включая биологию, культурологию, экономику, экологию, географию, историю, языкознание, средства массовой информации и психологию. Структура опросов соответствовала Набору шаблонов для курирования данных (Data Curation Profiles Toolkit), разработанному в Университете Пердью, были добавлены вопросы относительно

описательной информации о предмете исследования. Предварительный анализ показывает, что методы управления исследовательскими данными существенно различаются среди респондентов, и соответственно различаются предпосылки для использования услуг в сфере научных данных. Дополнительные вопросы относительно информации, касающейся предмета опроса, выявили потребность в услугах, направленных на оказание исследователям помощи в составлении полноценного описания своих массивов данных.

('Essentials 4 Data Support': five years' experience with data management training)

"Основы информационного обеспечения": пятилетний опыт обучения управлению данными

Эллен Вербакел, Мариан Гроотвельд

IFLA Journal, 42-4, 278-283

Аннотация:

В данной статье описан курс управления научными данными для вспомогательного персонала, такого как библиотекари и работники сферы информационных технологий. Авторы, которые занимаются обучением участников, повествуют о трех форматах курса и более детально описывают само обучение. За последние годы в обучении приняли участие более 170 человек. Курс сочетает большой объем интерактивной информации с очными встречами участников. Целью курса является оказание помощи участникам в совершенствовании различных навыков, а также в получении знаний, благодаря которым они с уверенностью смогут оказывать поддержку изыскателям, консультировать и обучать их. Взаимодействие между студентами вплетено в структуру подготовки, поскольку мы рассматриваем его как чрезвычайно важное средство для развития сети профессиональных контактов. С недавнего времени в рамках курса реализуется новая задача: в дополнение к обычным занятиям была проведена пара занятий без отрыва от производства. В заключение приводится описание ключевых групповых заданий для таких занятий в малом составе.

(Research Data Services at ETH-Bibliothek)

Услуги в области управления научными данными в Библиотеке ETH

Ана Сесартич, Маттиас Тёве

IFLA Journal, 42-4, 284-291

Аннотация:

Управление научными данными в ходе их жизненного цикла является как необходимым условием для коллективного использования данных, так и средством эффективного хранения данных. В этой статье кратко описаны услуги в сфере работы с данными, а также общий подход к управлению данными, имеющие место в настоящий момент в Библиотеке ETH (ETH-Bibliothek), главной библиотеке Швейцарской высшей технической школы Цюриха (ETH Zurich), крупнейшего технического ВУЗа Швейцарии. Услуги, предлагаемые поставщиками услуг в рамках ETH Zurich, полностью покрывают жизненный цикл данных. Библиотека обеспечивает поддержку в концептуальных вопросах, предлагает тренинги и услуги в области публикации и длительного хранения данных. Поскольку управление научными данными продолжает играть все более значимую роль как в контексте требований исследователей и спонсоров, так и в рамках учебных программ и стандартов качества научных исследований, Библиотека ETH налаживает тесное сотрудничество с исследователями с целью содействия процессу взаимного обучения, а также энергичного решения новых задач.

(Beyond the Matrix: Repository Services for Qualitative Data)

За рамками Матрицы: Услуги хранения качественных данных в репозитории

Себастиан Кархер, Десислава Кирилова, Николас Вебер

IFLA Journal, 42-4, 292-302

Аннотация:

Репозиторий данных, отобранных по качественному признаку (QDR - Qualitative Data Repository) формирует инфраструктуру и методологические принципы обмена и многократного использования цифровых данных, к которым обращаются в рамках социальных исследований с применением качественного а также комплексного методов. В настоящей работе мы описываем некоторые моменты из раннего опыта репозитория в части оказания услуг, введенных специально для курирования данных из области качественных исследований. Мы обращаем особое внимание на усилия QDR, направленные на решение двух основных задач в рамках обмена качественными данными.

Первая задача касается ограничений в области обмена данными, обусловленных защитой людей, участвующих в данном процессе, сохранением их персональных данных, а также соблюдением законодательства об авторском праве. Второй блок задач связан с уникальными характеристиками качественных данных и их взаимоотношением с опубликованным текстом. Мы описываем новаторский метод снабжения научных публикаций комментариями, в результате которого появляется "прозрачное приложение", позволяющее обмениваться такими "гранулированными данными" (Моравчик и др., 2013). В заключение мы описываем будущие направления развития услуг QDR в области хранения, обмена и многократного использования качественных данных.

(Data governance, data literacy and the management of data quality)

Упорядочение данных, грамотность в сфере обращения с данными и управление качеством данных

Тибор Колтай

IFLA Journal, 42-4, 303-312

Аннотация:

Упорядочение данных и грамотность в сфере обращения с данными являются двумя важными структурными элементами, лежащими в основе знаний профессионалов в области информации, участвующих в обеспечении таких научных исследований, которые требуют обработки большого объема данных, и оба эти фактора влияют на качество данных и на управление данными, полученными в ходе исследований.

Использование упорядочения данных в процессах управления данными, полученными в ходе исследований, а также в обучении грамотности в сфере обращения с данными помогает разграничить области принятия решений и определить сферы ответственности при принятии решений. Применение упорядочения данных эффективно, поскольку эта деятельность основана на стандартизированных и повторяемых процессах и предназначена как для обеспечения прозрачности процессов, связанных с обработкой данных, так и для снижения расходов. Оно также целесообразно, поскольку сопряжено с нормами, процедурами, стандартами, правилами принятия решений, отчетностью и методами контроля за исполнением. Следовательно, несмотря на то, что больше внимания им уделяется в корпоративной среде, и что библиотекари уже владеют некоторыми из относящихся к ним навыков, знания в области упорядочения данных являются основополагающими для услуг в области исследовательских данных, особенно когда они присутствуют на всех уровнях услуг в области исследовательских данных и применимы к большим объемам информации.

(Data information literacy instruction in Business and Public Health: Comparative case studies)

Обучение грамотности в области данных и информации в сфере бизнеса и здравоохранения: Сравнительное исследование на конкретном примере

Катарина Мейси, Хезер Коэйтс

IFLA Journal, 42-4, 313-327

Аннотация:

Работодатели нуждаются в рабочей силе, которая способна использовать данные для создания информации, имеющей действенное применение. Этим обусловлена необходимость развития студентами навыков грамотности в области данных и информации, которые позволят им ориентироваться и формировать смысловое содержание во все более усложняющемся информационном мире. В данной статье рассматриваются причины, по которым грамотность в области данных и информации должна быть включена в программы обучения, в особенности в сфере бизнеса и общественного здравоохранения, а также предлагаются стратегии реализации указанной задачи. Мы рассматриваем эту задачу как сравнительный анализ конкретной ситуации в рамках программы бакалавриата в сфере бизнеса, а также магистратуры в сфере здравоохранения в Университете Индианы - Университете Пердью в Индианаполисе. Данный анализ конкретных ситуаций позволил вывести несколько заключений относительно практических приемов, применимых в рамках как социальных, так и медико-санитарных дисциплин.

Resúmenes

(Modifying researchers' data management practices: A behavioural framework for library practitioners)

Modificación de las prácticas de tratamiento de datos de los investigadores: un marco de conducta para los bibliotecarios

Susan E Hickson, Kylie Ann Poulton, Maria Connor, Joanna Richardson, Malcolm Wolski

IFLA Journal, 42-4, 253-265

Resumen:

Datos es la nueva palabra de moda en las bibliotecas universitarias, ya que la política estipula que estos deben ser abiertos y accesibles, los patrocinadores exigen planes formales de tratamiento de datos y las instituciones están aplicando directrices en torno a las buenas prácticas. Debido a las inquietudes relacionadas con las prácticas de tratamiento

de datos de los investigadores, este artículo recoge los resultados iniciales de un proyecto realizado en la Universidad de Griffith para aplicar un marco conceptual (A-COM-B) destinado a entender la conducta de los investigadores. El proyecto pretende fomentar el uso de soluciones institucionales para el tratamiento de datos de investigación. Según las entrevistas realizadas por un equipo de bibliotecarios en un centro de investigación de ciencias sociales, los resultados preliminares señalan que la actitud es un elemento clave que se habrá de abordar para diseñar estrategias de intervención encaminadas a modificar la conducta. El artículo concluye con un debate de las siguientes fases del proyecto, que abarcan la recopilación y el análisis de los datos, la aplicación de las estrategias identificadas y una actividad para evaluar el alcance de las modificaciones introducidas en las prácticas indeseables actuales.

(Research data services: An exploration of requirements at two Swedish universities)

Servicios de datos de investigación: estudio de las necesidades de dos universidades suecas

Monica Lassi, Maria Johnsson, Koraljka Golub

IFLA Journal, 42-4, 266-277

Resumen:

El artículo explica un estudio exploratorio de las necesidades de los investigadores para lograr un tratamiento eficaz de los datos de investigación en dos universidades suecas, realizado para promover el desarrollo de los servicios de datos de investigación. En el estudio han intervenido doce investigadores de diversos campos, como la biología, los estudios sociales, la economía, las ciencias medioambientales, la geografía, la historia, la lingüística, los medios de comunicación y la psicología. Las entrevistas se estructuraron siguiendo las orientaciones del *Data Curation Profiles Toolkit* (kit de perfiles de conservación de datos) desarrollado en la Universidad de Purdue, y se añadieron preguntas relacionadas con metadatos de sujetos. El análisis preliminar indica que las prácticas de tratamiento de los datos de investigación varían en gran medida entre los encuestados, al igual que lo hacen las implicaciones para los servicios de datos de investigación. Las preguntas añadidas sobre metadatos de sujetos indican la necesidad de servicios que ayuden a los investigadores a describir sus conjuntos de datos con metadatos adecuados.

('Essentials 4 Data Support': five years' experience with data management training)

'Essentials 4 Data Support': cinco años de experiencia en la formación sobre tratamiento de datos

Ellen Verbakel, Marjan Grootveld

IFLA Journal, 42-4, 278-283

Resumen:

Este artículo describe un curso de tratamiento de datos de investigación para el personal de asistencia, como bibliotecarios y personal de TI. Los autores, que forman a los participantes, presentan los tres formatos del curso y describen la formación con mayor detalle. En los últimos años, más de 170 personas han participado en esta formación, que combina gran cantidad de información online con reuniones personales. El objetivo del curso consiste en ayudar a los participantes a fortalecer diversas destrezas y adquirir conocimientos al objeto de reforzar su confianza para ayudar, asesorar y formar a los investigadores. La interacción entre los estudiantes se integra en la estructura de la formación, porque considerarla un instrumento valioso para el desarrollo de una red profesional. Recientemente, el curso ha aceptado un nuevo reto: además de los cursos normales, se han impartido un par de cursos a la carta. El artículo concluye con una descripción de las tareas de grupo claves para estas formaciones.

(Research Data Services at ETH-Bibliothek)

Servicios de datos de investigación en la ETH Bibliothek

Ana Sesartic, Matthias Töwe

IFLA Journal, 42-4, 284-291

Resumen:

El tratamiento de los datos de investigación a lo largo de su ciclo de vida es un requisito previo fundamental para el intercambio eficaz de datos y la conservación eficiente de los datos a largo plazo. Este artículo resume los servicios de datos y el planteamiento general del tratamiento de datos aplicado actualmente en la ETH-Bibliothek, la biblioteca principal de ETH Zurich, la universidad técnica más grande de Suiza. Los servicios ofrecidos por los proveedores de servicios en ETH Zurich abarcan todo el ciclo de vida de los datos. La biblioteca ofrece apoyo relacionado con preguntas conceptuales, así como formación y servicios relacionados con la publicación de datos y su conservación a largo plazo. Dado que el tratamiento de los datos de investigación está adquiriendo cada vez más importancia en los requisitos de los investigadores y los patrocinadores, así como en los programas de estudios y las buenas prácticas científicas, ETH-Bibliothek está estableciendo colaboraciones estrechas con los investigadores, al objeto de promover un proceso de aprendizaje mutuo y abordar nuevos retos.

(Beyond the Matrix: Repository Services for Qualitative Data)

Trascendiendo la matriz: servicios de archivado para datos cualitativos

Sebastian Karcher, Dessislava Kirilova, Nicholas Weber

IFLA Journal, 42-4, 292-302

Resumen:

Qualitative Data Repository (QDR) ofrece una infraestructura y orientación para el intercambio y la utilización de los datos digitales usados en consultas cualitativas y sociales multimétodo. En este artículo se describen algunas de las primeras experiencias de archivo que prestan servicios desarrollados específicamente para la conservación de los contenidos de investigación cualitativos. Nos centramos en iniciativas de QDR para abordar dos desafíos para el intercambio de datos cualitativos. El primer desafío tiene que ver con los límites en el intercambio de datos para proteger a los sujetos y sus identidades y cumplir las leyes de copyright. El segundo conjunto de desafíos aborda las características únicas de los datos cualitativos y su relación con el texto publicado. Describimos un método novedoso de anotación de publicaciones científicas, que genera un «apéndice de transparencia» que permite el intercambio de estos «datos granulares» (Moravcsik et'al., 2013). Concluimos describiendo las futuras direcciones de los servicios de QDR para el archivado, el intercambio y la reutilización de datos.

(Data governance, data literacy and the management of data quality)

Gobernanza de los datos, competencias básicas en materia de datos y gestión de la calidad de los datos

Tibor Koltay

IFLA Journal, 42-4, 303-312

Resumen:

La gobernanza de los datos y las competencias básicas en materia de datos son dos elementos importantes de la base de conocimientos de los profesionales de la información que participan en investigaciones en las que se emplean gran cantidad de datos y que abordan tanto la calidad de los datos como el tratamiento de datos de investigación.

La aplicación de la gobernanza de los datos a los procesos de tratamiento de datos de investigación y la educación en competencias básicas en materia de datos ayuda a delimitar los dominios de decisiones y definir las responsabilidades de dichas decisiones. La adopción de la gobernanza de los datos es ventajosa, porque se trata de un servicio basado en procesos normalizados y reproducibles y está diseñado para propiciar la transparencia de los procesos relacionados con los datos y la reducción de gastos. También resulta muy útil, porque se refiere a reglas, políticas y normas, derechos de decisión, responsabilidades y métodos de ejecución. Por lo tanto, aunque recibieron más atención en entornos corporativos y los bibliotecarios ya poseían algunas de las destrezas relacionadas con él, el conocimiento sobre la gobernanza de los datos es fundamental para los servicios de datos de investigación, especialmente porque aparece en todos los niveles de dichos servicios y es aplicable a *big data*.

(Data information literacy instruction in Business and Public Health: Comparative case studies)

Competencias básicas en materia de información sobre datos en los ámbitos de la empresa y la sanidad pública: casos prácticos comparativos

Katharine Macy, Heather Coates

IFLA Journal, 42-4, 313-327

Resumen:

Los empresarios necesitan una fuerza de trabajo capaz de usar datos para crear información verificable. Ello exige a los estudiantes desarrollar competencias básicas en materia de información sobre datos que les permita avanzar de manera significativa en el complejo mundo de la información. Este artículo examina por qué deben integrarse en los programas curriculares las competencias básicas en materia de información sobre datos, especialmente en el caso de las empresas y la sanidad pública, y ofrece estrategias para adquirirlas. Lo abordamos como un caso práctico comparativo en programas de grado de empresariales y programas de posgrado de sanidad pública en la Universidad de Indiana - Universidad de Purdue, Indianápolis. Estos casos prácticos desvelan algunas implicaciones para la práctica que se aplican en todos los programas de ciencias sociales y de la salud.