Beyond Literacies: The Evolving Landscape of Library Support to Research 2.0

Tibor Koltay
Institute of Learning Technologies
Eszterházy Károly University Jászberény, Hungary

Abstract

Purpose/Thesis: This paper identifies some of the tasks and roles that academic libraries have to perform in order to react to the emergence of Research 2.0.

Approach/Methods: The argument is based on a non-exhaustive review of the recent literature. Results and conclusions: Academic libraries should respond to the emergence of Research 2.0 by filling niches in services provided by other academic units.

Research Limitations: As a rule, only the literature of the second half of the 2010s was taken into consideration.

Practical implications: The tasks identified in this paper may not seem urgent today, but the likelihood that they will become an imperative in the future is high.

Originality/Value: The issues identified in this paper are already a part of everyday best practices in several countries.

Keywords:
Data citation. Data literacy. Data quality. Research data services. Research management.

Received: 29 April 2017. Reviewed: 19 July 2017. Accepted: 30 August 2017.

1. Introduction

Research 2.0 (Science 2.0, or eScience) is understood as research in the sciences, social sciences and the humanities. It is making use of the Internet’s power, which provides the possibility to access and manipulate massive amounts of data. As a result, it brings in a potential to enable different new forms of networking and encourages openness of scientific research, even though many of its aspects remain debated (Pasquetto et al., 2015).

The practical attainment of open science is data sharing that leads to open data, the benefits of which in academia are manifold because it enables data to be reused by other researchers and applied in various other contexts, therefore it fosters transparency in the research process and data collection methods. Nonetheless, one of the most difficult challenges is convincing researchers to accept the idea of data sharing.

---

1 This text is the edited version of the paper presented at the 4th International Scientific Conference: Information Science in the Age of Change, Innovative Information Services, in Warsaw, 15th–16th May 2017. Its preparation was supported by the EFOP-3.6.1–16-2016–00001 project "Complex Development of Research Capacities and Services at Eszterházy Károly University".
The authenticity and trustworthiness of the data is reaffirmed by usage. Data sharing also saves time because the researchers who reuse data do not have to “reinvent” many of the results already attained (Patel, 2016).

Besides the intensive use and openness of research data, Research 2.0 is determined by the growing adoption of social media tools by researchers (Koltay et al., 2015). There is a palpable growth of interest in academic social media sites (social media services, designed for researchers), such as Academia.edu (www.academia.edu), ResearchGate (www.researchgate.net), Mendeley (www.mendeley.com) and Kudos (www.growkudos.com).

The allegedly best known ResearchGate offers services for building, showcasing, and measuring reputation as it provides 10 reputational mechanisms, even though it struggles with the deployment of engagement metrics. In general, it has the potential to become a major deliverer of scholarly reputation, but it did not achieve this yet (Nicholas et al., 2016).

Nonetheless, empirical studies of natural and social science researchers in Britain and the United States show that the reliability and credibility remain central for them and are guaranteed by peer review. As authenticity is ensured by the participants, messages can be transmitted on blogs or social networking sites, but usually they are not cited as publishing articles in peer-reviewed journals (sometimes publishing books) remains the condition of researchers’ advancement (Watkinson et al., 2016).

The examination of 116 early career researchers in seven countries (Britain, China, France, Malaysia, Poland, Spain, and the United States) shows that they have no choice but to adapt to the rules, i.e. to publish in journals that have high prestige. Social media is used mainly to obtain PDF copies of articles, to keep contacts with other researchers and build a digital presence of themselves and of their publications, while active collaboration is not present (CIBER, 2016).

There are several tasks and roles that academic libraries may fulfil in order to react in this environment, last but not least to change the views that see libraries only as dispensaries of books and articles (Jahnke et al., 2012). Some of the possible tasks have been identified by Koltay (2015b; 2016a).

While some teaching staff members and researchers may be reserved against the involvement of librarians, interviews also indicate their willingness to accept data-related library services, provided that they are adequate (McLure et al., 2014).

This paper intends to inform about both theoretical considerations and practical experiences reflected in the literature. As such, it addresses only a selection of issues, thus it does not try to be exhaustive.

Awareness of the core processes of research and the dynamics of change within scholarly communities has to become integral part of researchers’ mid-range strategies. Such awareness should result in the provision of a variety of targeted services within research data services, information literacy and data literacy education, raising awareness of teaching staff members on different issues and providing individual support to them.

In accordance with the above, this paper addresses the following issues and activities:

– Research data services;
– Data literacy education;
– Supporting individual teaching staff members;
– Changing roles for the information professional.
2. Research Data Services

Research data services (RDS) provide an overall service framework that is relevant to the variety of players and related processes of managing and curating research data. RDS should be provided by academic libraries to help researchers to meet the challenges of the data-intensive part of Research 2.0. Some academic libraries are already offering research data services that mainly are extensions of their traditional services, which consist of consulting with teaching staff or students on data management plans and metadata standards; providing reference support for finding and citing data sets; or providing web guides and finding aids for data or data sets. Technical services in the form of providing support for using data repositories, preparing data sets for repositories bear considerable potential, but are provided to a lesser extent than informational ones (Tenopir et al. 2016).

2.1. RDS concepts

Research data management (RDM) is the main constituent of RDS. It is a set of general activities not specifically attached, but potentially performed by the library. RDM means caring for research data, facilitating access to it, preserving and adding value to this data throughout its lifecycle. It is one of the key determinants of open data publication and data curation, bringing the benefits of enabling findability and understandability of data, avoiding unnecessary duplication, and validating results (University of Edinburgh, 2015).

Research data management is not always clearly differentiated from data curation. Its goal is enabling data reuse and preventing data loss. Therefore, it is broader than data management, and involves maintaining, preserving and adding value to digital research data throughout its lifecycle and is embedded into the larger meaning of curation as taking care of resources (DCC, 2015; Partlo et al., 2015; Kouper, 2016; Poole, 2016).

Besides data management and data curation, there is also data stewardship that sets somewhat different but not entirely disparate accents by being directed towards maintaining the integrity of and preserving access to data (York et al., 2016).

Even though data governance is usually a subject of interest for the business sector, being knowledgeable of its principles and practice is crucial also for the management of research data. Data governance can be defined as the exercise of decision-making and authority that comprises a system of decision rights and accountabilities that is based on agreed-upon models, which describe who can take what actions, when and under what circumstances, using what methods. It enables better decision-making because it concerns rules, policies, standards, decision rights, accountabilities, and methods of enforcement (DGI, 2017). It also deals with availability, access, provenance, meaning, and trustworthiness (ECAR, 2015). These qualities make it to be the starting point for managing data with the distinction that data management involves determining the actual criteria employed for data quality, while data government is about designating who should make these decisions (Khatri & Brown, 2010). Therefore, it is not only one of the necessary conditions for providing open data, but brings substantial knowledge to the work of the academic librarian (Koltay, 2016b).

All above processes require distinguished attention to data quality that is determined by trust, authenticity, reliability, understandability, discoverability, accessibility, and integrity (Giarlo, 2013). Besides of these, it is characterised by availability, acceptability, accuracy,
integrity, completeness, and usability (IBM, 2012). Credibility depends – among others – on the coverage of data, authority, format expectations, quality control processes and the opposition between being readable by humans or machines. Examining the quality of research data also should mean considering relevance to the given research question and format that comprises the fields and units used, naming conventions, dates of creation and update (Zilinski & Nelson, 2014). Issues, related to copyright, data licensing, data security, data privacy and to the danger of the erroneous interpretation of data also are tied to quality (Patel, 2016).

2.2. RDS processes and services

Data citation and metadata concern both researchers and librarians. Data citation is a potential source of motivation for researchers to share and publish their data, while they may not have the time and expertise to provide standard metadata (Mooney & Newton, 2012). There are several initiatives and guidelines on providing standardized forms for data citation (e.g. DATACITE, 2016a; ANDS, 2017). A metadata schema is offered, e.g. by DATACITE, (2016b). Having recognized its importance, Clarivate, a commercial information provider offers its Data Citation Index.

Supplying metadata for data sets is one of the most important technical RDS activities. Libraries are experienced in this field of activities as they have been providing “traditional” metadata to books and journal articles. By creating metadata to datasets, academic libraries fill in a niche in the support chain. This is especially true because everyone, including funding bodies, recognises that the generation of metadata by researchers is a time and resource-intensive activity, which nonetheless is vital for the curation of research data. According to Chao (2015), in providing metadata, distinguished attention should be given to documenting information about the implemented method.

When students face the difficulties related to exploring the boundaries of their disciplines and doing cross-disciplinary research, they find it extremely hard to know what datasets are available in a given field of study. Established researchers are in a better situation because they are often familiar with at least the main data sources and providers. In both cases, information professionals can gain insight through data reference interviews (consultations) that are rooted in traditional librarianship, and may be guided by accepting some of the following guidance, originating both from the traditions of reference work and the needs for data-related information. These factors make the data reference interview process much more complicated than the traditional one:

– Data reference interviews may consist of more questions that a traditional one;
– Making assumptions about the user’s requirements, prior knowledge or viewpoint should be avoided;
– Acronyms and jargon used by the user may require explanation;
– It may be impossible to find the perfect source for the user. Nonetheless, the librarian can probably find some useful starting points (Rice & Southall, 2016).

RDS may include data management instruction. In the long term, such education should be integrated into the disciplinary curricula to allow learning within the context of the given fields. Nonetheless, several existing programs are already available, making it unnecessary to create new training programs from scratch (Eaker, 2014). The findings of a survey of
researchers and teaching staff members, working in academic institutions, research organizations, federal agencies and non-profit organizations in the United States, representing hard sciences, social sciences, humanities, and law supported the idea that assistance in data management instruction provided by trained librarians can help science educators, who lack time and expertise (Tenopir et al. 2016).

2.3. Opinions on RDS

According to the League of European Research Universities (LERU), supporting research data management is a new role for academic libraries. Nonetheless, they can provide help in increasing the visibility of research data and they are well placed to advocate best practices in data management and data citation. The main reason of involving libraries into these processes is the need to minimize the time that researchers have to spend on technical and administrative processes. According to the LERU report, successful RDM requires cooperation between librarians and researchers, especially in clarifying what is expected of researchers when describing and citing data (LERU, 2013).

The involvement of libraries is a process that has multiple stages and could include some of the actions recommended by the Association of European Research Libraries (LIBER), for example:

1. Offering research data management support, including assistance in preparing data management plans;
2. Liaising with researchers, data archives and data centres to foster an interoperable infrastructure for data access, discovery and data sharing;
3. Providing services for storage, discovery and permanent access;
4. Promoting data citation (Christensen-Dalsgaard, 2012).

As indicated also in the above recommendations, there is no need for every institution to implement the same RDS components. Services should depend on several factors, including the size of the institution and the vision of its management. In any case, engaging in strategic planning is an ideal start either to establishing RDS or further developing the existing services (Briney, 2016).

Another type of guidance is offered by Erway at al. (2015) who outline the building blocks, needed for starting data services in libraries that have very little, if anything in place, and have limited resources at their disposal. They also advise on how to build on these foundations as needs become evident and as resources allow.

Besides of this, librarians who want to educate themselves in RDM can consult a selection of relevant resources, available in the webliography, complied by Goben and Raszewski (2015).

On the other hand, it is necessary to understand that RDM needs to be a researcher-owned endeavour. Having recognised this, the University of Cambridge made efforts to create a community of Data Champions among researchers that provides training and advocacy for good RDM practice (Higman et al., 2017).
3. Data literacy education

Data literacy itself is not without precedents, while providing data literacy education to different groups of (graduate and undergraduate) students, as well as to researchers (including teaching staff members, non-teaching researchers, postdoctoral researchers, and doctoral students) is in many regards new task for academic librarians.

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). Further definitions are reviewed by Crusoe (2016).

Data literacy education should incorporate both the social and technical aspects of data (Sharma & Qin, 2014). This is demonstrated by the list of abilities that define data literacy, conceived by Calzada, Prado & Marzal (2013), as follows:

– Knowing how to select and synthesize data and combine them with other information sources and prior knowledge;
– Identifying the context in which data is produced and reused;
– Recognizing source data value, types and formats;
– Determining when data is needed;
– Accessing data sources appropriate to the information needed;
– Critically assessing data and their sources;
– Determining and using suitable research methods;
– Handling and analysing data;
– Presenting quantitative information;
– Applying results to learning, decision making or problem-solving;
– Planning, organizing and assessing ourselves throughout the process.

Data quality, as well as data citation and metadata also remain foundational for data literacy education (Ridsdale et al., 2015).

Data literacy is clearly tied to information literacy education that is constantly changing. One of the most evident signs of this transformation is the fact that Information Literacy Competency Standards for Higher Education, conceived in the United States of America by the Association of College and Research Libraries (ACRL, 2000) is being supplanted by the Framework for Information Literacy for Higher Education (ACRL, 2015), albeit there are already opinions that the Standards should remain as the principal foundation for the understandings of information literacy practice (Jackman, & Weiner, 2017).

As to the possible methods of data literacy education Maybee and Zilinski (2015) introduced data informed learning as an approach that – instead of teaching generic data-related skills – focuses on how to use data in disciplinary contexts and is based on the idea of informed learning (Bruce, 2008).

4. Supporting individual researchers

Although supporting individual researchers is different from general RDS activities, this kind of services should be strengthened. As individual’s needs are often difficult to articulate, to achieve success in this field, two questions should be asked and answered.
(1) What is the perception of librarians about their own role in relation to the research activities of teaching staff and what do they want to achieve?

(2) What is the perception of teaching staff members about the supporting role of librarians in their research?

The involvement in the support of individual research can be grouped into the following categories:

– Providing informal alerting services;
– Purchasing requested resources;
– Answering in-depth reference questions;
– Creating visual representations of data;
– Co-researching about scholarly publishing in a specific area of knowledge;
– Co-researching the scholarship of teaching;
– Consulting about searching (Brydges & Clarke, 2015).

5. Awareness-raising

There are tasks that academic libraries are expected to address, albeit they are non-traditional, i.e. not offered directly by the libraries themselves, but the librarians are the ones who can make researchers aware of their existence.

Despite problems and controversies around Open Access to scientific publications (OA), it is still important. In particular, there should be ample information about the Diamond, the Gold and the Green routes of OA, not forgetting about issues of article processing charges (APCs). A relatively new and to an extent controversial problem is the presence of predatory journals and publishers. Predatory journals’ main goal is to seize article processing charges that are paid by authors, who publish their papers in the form of Gold Open Access. APCs themselves are the “protagonists”, but the real problem is that predatory journals tend to publish substandard or low-quality papers as they display false identities that mimic established journals, without offering substantial added value. This manifest itself – among others – in the fact that these journals offer unrealistically fast turnaround times of publication, partly because the level of the peer-review process is usually extremely low or it is simply non-existent (Nelson & Huffman, 2015). At the moment, the best tool to sift out predatory journals, is the Directory of Open Access Journals (DOAJ, http://doaj.org) that is a whitelist of quality open-access publishers and journals.

Alternative metrics (often called altmetrics) of scientific output also pertain to Research 2.0 and exhibit great variety (Peters & Bar-Ilan, 2014). Recently, social media sites, first of all ResearchGate, have advanced in showcasing and measuring reputation by using alternative metrics (Nicholas et al., 2016). This is one more reason why librarians have to raise awareness of academic social media sites, even if they have relatively restricted impact on researcher’s perception and behaviour.
6. Changing roles for the information professional

Maintaining datasets requires knowledge of both the given domain and advanced technological knowledge. This raises questions whether academic librarians possess the necessary skills and qualifications for this task. Behind this question, there is also a more profound one, if there are conceptual differences between managing data and managing documents. If the answer is yes, librarians have to fulfil new roles, for which they have to be well-prepared (Nielsen, & Hjørland, 2014).

Research 2.0 and especially open data are connected to big data. Approaches to big data are often accompanied by a rhetoric that is both utopian and dystopian, i.e. big data is seen as a panacea to many problems and a tool that enables decreasing privacy and civil freedoms, while the changes that is causes are more subtle shifts (Boyd & Crawford, 2012). However, debates seem to respond more to the popular media’s hype about these issues, while practicing data analysts are neither proponents, nor critics of these (Carter & Sholler, 2016).

If there is a relative lack of commonality even in the requirements between the roles of data librarians, data stewards or data curators, the disparity is more sizeable between librarians and data scientists. At the same time, there are intersections between them in data management (Lyon et al., 2015). In addition to this, the roles within this domain are often distributed among professionals because they are different, but still show features of interdependence. Having varied names, data analysts, data journalists or data engineers require high quality, curated data because they work with it. Data librarians show their value by having the data in their care and presenting it to be used and analysed (Lyon & Mattern, 2016). The strongest commonality between these roles and the most powerful connection to data science is the significant attention to data quality and to creativity, intelligence in exploring the nonstandard, personalized and often not automated environment that has to be discovered for finding evidence and indicators for decision-making (Cao, 2016).

On a deeper, conceptual level it seems to be important to heed the warning of Robinson (2016) who reminds us that academics and practitioners should not to be specialised either in data, i.e. examining pattern and syntax and using quantitative methodology, or in information, i.e. interested in meaning, semantics and using qualitative methods because becoming entrenched in being data scientists or information professionals would be harmful for both professions.

7. Conclusion

This paper addressed several issues related to research data services, providing a patchy picture that intended to give insight into several essential and intriguing questions. None of the answers to these is final, but – even if slowly – researchers gradually will have to get accustomed to publishing not only manuscripts, but making research data open (Patel, 2016). Academic libraries may take the lead in this process, or – at least – it is necessary for them to reflect on new developments by filling the niches that appear in services, provided for researchers by other academic units.
References


Crusoe, D. (2016). Data Literacy defined pro populo: To read this article, please provide a little information. The Journal of Community Informatics, 12(3), 27–46.

DATACITE (2016a). Why is it so important to cite data? [online]. DataCite, [09.04.2017], https://www.datacite.org/cite-your-data.html


IBM (2012). Successful information governance through high-quality data. Somers, NY: IBM Corporation


---

Poza edukacją kompetencji: rozwój sektora bibliotecznego wsparcia dla badań 2.0

Abstrakt

Cel/Teza: W artykule wskazano niektóre zadania i role, które biblioteki akademickie muszą spełnić w odpowiedzi na pojawienie się tzw. badań 2.0.

Koncepcja/Metody badań: Argumentacja oparta jest na niewyczerpującym przeglądzie najnowszej literatury.

Wyniki i wnioski: Biblioteki akademickie powinny reagować na pojawienie się badań 2.0, wypełniając nisze w usługach, które temu typowi badań zapewniają inne jednostki akademickie.
Ograniczenia badań: Zasadniczo w przedstawionych rozważaniach wykorzystano jedynie piśmiennictwo, które ukazało się po 2015 r.

Zastosowania praktyczne: Zadania wskazane w niniejszym artykule nie muszą wydawać się pilne dzisiaj, ale prawdopodobieństwo, że w przyszłości staną się imperatywem jest wysokie.

Oryginalność/Wartość poznawcza: Zagadnienia wskazane w niniejszym artykule już dzisiaj są częścią codziennych dobrych praktyk w wielu krajach.

Keywords:

---

Dr. habil. TIBOR KOLTAY, PhD Professor, Chair of the Institute of Learning Technologies Eszterházy Károly University in Jászberény, Hungary. Since 1981 he worked as an information officer, then as director of different academic libraries. From 1992 to 2004 he was also involved in part-time LIS teaching activities in different institutions. From 2004 he has been working as a full-time LIS teaching staff member at Szent István University and (until 2012) the University of West Hungary. Since 2004 he won eight research scholarships, including a Fulbright Research Scholarship to the School of Library and Information Science of Kent State University, Kent, Ohio. Prof. Koltay also served as ERASMUS guest lecturer at five European universities. After a longstanding engagement with the study of abstracting, he turned towards information literacy and other related literacies, including data literacy, still maintaining his interest in the interdisciplinary relationships between linguistics and information science. His most important recent publications are Abstracts and Abstracting. A genre and set of skills for the twenty-first century. Oxford: Chandos Publishing, in 2010; The media and the literacies: media literacy, information literacy, digital literacy. Media Culture & Society, 33(2), 211–221, in 2011, Data literacy: in search of a name and identity. Journal of Documentation, 71(2), 401–415 in 2015, Are you ready? Tasks and roles for academic libraries in supporting Research 2.0. New Library World, 117(1–2), 94–104, Data governance, data literacy and the management of data quality. IFLA Journal, 42(4), 303–312, Research 2.0 and the Future of Information Literacy (with Sonja Špiranec and László Z. Karvalics) Oxford: Chandos Publishing, Library and information science and the digital humanities: perceived and real strengths and weaknesses. Journal of Documentation, 72(4), 781–792 in 2016 and Data literacy for researchers and data librarians. Journal of Librarianship and Information Science, 49(1), 3–14 in 2017.