IDENTIFYING NEW ROLES FOR ACADEMIC LIBRARIES IN SUPPORTING DATA-INTENSIVE RESEARCH¹

© Tibor Koltay, 2019

Eszterházy Károly University, Jászberény, Hungary; koltay.tibor@uni-eszterhazy.hu

Reacting to the appearance of data-intensive research prompts academic libraries to become service providers for scholars, who work with research data. Although this is an imperative for libraries worldwide, due to the differences between countries and institutions, the level of readiness to engage in related activities differs from country to country. While some of the related tasks are fairly novel, others heavily build on librarians' traditional, well-known skills. To identify these tasks, as well as making an inventory of the required skills and abilities, this paper, based on a non-exhaustive review of the recent literature, presents both theoretical and practical issues. It is demonstrated that the most obvious directions of the service development in academic libraries to support data-intensive science are research data management, data curation, data literacy education for users, and data literacy education for librarians.

Keywords: Academic libraries; Data librarians; Data literacy; Data science; Research data management; Data curation

Citation: Koltay T. Identifying new roles for academic libraries in supporting data-intensive research. *Bibliosphere*. 2019. № 4. P. 97–102. DOI: 10.20913/1815-3186-2019-4-97-102.

Определение новых ролей библиотек университетов в поддержке исследований, основанных на больших данных Тибор Колтай

Университет им. Кароя Эстерхази, Ясберень, Венгрия; e-mail: koltay.tibor@uni-eszterhazy.hu

Появление исследований, основанных на больших данных, побуждает научные и университетские библиотеки предоставлять услуги ученым, которые работают с исследовательскими данными. Несмотря на то что это является императивом для библиотек во всем мире, уровень их готовности к участию в соответствующей деятельности отличается от страны к стране. В то время как некоторые из услуг, связанных с этим направлением, являются довольно новыми, другие в значительной степени опираются на традиционные, хорошо известные навыки библиотекарей. В работе на основе обзора новейшей литературы представлены как теоретические, так и практические вопросы, которые помогут в постановке задач и составлении перечня необходимых навыков и умений. Показано, что наиболее очевидными направлениями развития сервиса в университетских библиотеках для поддержки информационно-интенсивной науки являются управление исследовательскими данными, курирование данных, обучение пользователей и библиотекарей информационной грамотности.

Ключевые слова: библиотеки университетов, научные библиотеки; библиотекари, работающие с данными, грамотность по работе с данными; управление научными данными, курирование данных

Для цитирования: Колтай Тибор. Определение новых ролей библиотек университетов в поддержке исследований, основанных на больших данных // Библиосфера. 2019. № 4. С. 97–102. DOI: 10.20913/1815-3186-2019-4-97-102.

¹ The writing of this paper has been supported by the EFOP-3.6.1-16-2016-00001 project "Complex Development of Research Capacities and Services at Eszterházy Károly University".

1. Introduction

Without aiming exhaustive coverage of the literature, this paper focuses on theoretical considerations and practical experiences related to one of the major data-related complex library services, involving research data management (RDM), and data literacy education.

It is not by accident that libraries are required to become service providers for scholars, working with research data. This has been acknowledged by the fact that this need appears among the five trends, declared to impact academic libraries [1].

More recently, in a series of interviews and workshops with stakeholders, supporting effective research data management has been identified as one of the areas of future value [2].

The necessity for the libraries to be involved in such activities has been brought about by the existence and growing expansion of data-intensive research, also called Research 2.0, Science 2.0, or eScience. All these names cover scholarly research, with the extensive use of data, carried out in the natural sciences, social sciences and the humanities, made open for the reuse by other researchers [3].

2. Data-intensive research and the academic library environment

Researchers, including teaching staff members and (doctoral) students do not need more data, but require having the right data [4]. However, data is of no use if not analysed [5].

Researchers need data-related support for the whole research data lifecycle from planning, to organizing, documenting, sharing, and preserving datasets. They also may seek advice on copyright, licensing and intellectual property issues. To satisfy these and other related needs, libraries must not only engage in interaction with researchers themselves, but should identify and cooperate with other support service providers [6, 7].

For libraries and librarians it is crucial to understand that their role is instrumental in solving some of the related problems. However, to find their proper place in this process, all librarians should acquire a conceptual understanding of data, and are also required to learn the skills and abilities “to find, extract, collect, clean, organise, analyse, and present data”, even if they do not serve users, who deal with data [8].

When forming data-related services, it is important to know, how researchers view the library. It is beneficial for librarians to appreciate the role of the library without overestimating it. All these tasks require librarians to change their self-identification and gain an appreciation of technical and socio-ethical issues related to research data [8]. It should

also enable them to demonstrate and communicate the value of services that the library could offer [9]. In this regard, they also should understand that relieving researchers’ technical and administrative burdens [10] is a respectable goal, but transcending this status would require focused and intensive work of library managers, who must understand the advantages of providing services related to research data and enjoy the backing of their whole staff [9].

2.1. Research Data Management (RDM)

One of the main and overarching services, where libraries play a key role is RDM, which is a comprehensive set of activities for the organization, storage, access, and preservation of data [11]. It includes services, tools and infrastructure that support the management of research data across its lifecycle [6].

From among RDM activities, informational services, such as providing reference support for finding and citing data (datasets) are offered with relatively high frequency. Most of these services involve a personal client-librarian relationship, thus are similar to reference services, traditionally supplied by libraries.

Providing consultations on data-management plans (DMPs) is usually the most frequent and usual one of this type of services [12]. Data reference interviews, are similar to reference interviews. Nonetheless, they may consist of more questions than their document-based counterparts [13], but still need to rely on the willingness and ability of both the researcher and the librarian to bring their own unique expertise into the interaction. In this regard, we know that librarians are encouraged to engage their users into making decisions during the interview [14]. This requirement goes back to the idea that libraries should acknowledge the knowledge brought to the reference interaction by the users [15].

Technical/technological services, selecting and preparing datasets for deposit, or deaccessioning them appear less frequently [16].

By its complexity and dependence on the requirements of the given disciplines and data types, creating metadata for datasets seems to fall into both categories.

Though much more intricate than citing research publications, data citation is an unquestionable necessity and can be offered in order to:

- Satisfy the need for connecting scholarly papers with the underlying data;
- Enable identification and retrieval of datasets;
- Devise new usage metrics for determining the impact of data;
- Facilitate data sharing and fostering reproducibility [17].

Even though reproducibility of underlying research has been undermined by the prioritization

of citation counts and journal prestige [18], it has the potential to allow not only the verification of research results, but also may lead to the emergence of new research questions [19]. Let us add that – in the hopefully not so distant future – data citation could become a regular motivating factor for researchers to share and publish their data [20].

Providing help to researchers in retrieving data is also a potential task for academic libraries, even though data retrieval is a less obvious task than retrieving research papers, because the average file size of datasets is usually larger than that of research publication files. In addition to this, the retrieved datasets may consist of different file types and/or file formats and may require downloading in order to be readable or used [21].

In general, librarians see RDM services as challenging, but in the same time interesting, enjoyable, satisfying, and rewarding [22]. On the other hand, it is difficult to scale them up [9], because librarians are accustomed to provide direct, on-demand individual consultancies [23].

2.2. Data curation

Data curation is very much in the limelight in the last few years. It is data curation that particularly highlights the need for abandoning opinions that portray data as something inferior to information [24].

Data curation is sometimes regarded as a new label for the existing activities that have been practiced in libraries, museums, and archives [25]. Indeed, many of the data curation activities show similarity to the existing roles of subject librarians [26].

Data curation should focus on preservation for reusability, because the value of data is exclusively in its extrinsic properties, i. e. their fitness for use [27]. In this sense, research data differs from digitized cultural heritage objects, which hold intrinsic value that justifies preservation regardless of ever being utilized or not. Accordingly, the emphasis of data curation is not on internal storage, but gives priority to extracting data for use, similar to exhibiting collections in museums [28] in order to be used by visitors.

Data curation services provide a technical infrastructure that supports data management by providing persistent storage, assignment of unique identifiers and metadata [29]. By aiming the prevention of data loss and adding value to trusted data assets for current and future use, data curation's main goals overlap with those of digital curation [30].

One of the vital messages that academic libraries need to accentuate when promoting data curation is that they will be responsible stewards of data, without taking ownership of this data. This means that the owners of any material will be able to take their content out of library systems at any point and

that this content will be accessible (to the best of libraries' abilities) in-perpetuity [31].

2.3. Data literacy

Data literacy's overarching nature and importance are undeniable among others by being associated with information literacy. Nonetheless, if defined as a set of skills and abilities, related to accessing, understanding, interpreting, managing, critically evaluating and ethically using research data [32] this concept deliberately excludes varied aspects of our life, where we can enjoy the benefits of data literacy. Such aspects include data in the civil society, open government, community informatics, journalism, business, and teacher education. Instead of these, it highlights the common concept of data literacy as a research skill [33]. Accordingly, being data literate is foundational for all kinds of researchers, students, as well as actual and future academic and special librarians, independently of their level of being involved in data-related services [34].

Teaching data literacy involves explaining the need for collecting data and the various ways to do it. It is also important to consider sampling techniques, sample selections and size, as well as survey design. The ability of distinguishing between correlation and causation, as well as the understanding of how variables influence each other is also have to be touched on. Possessing the skills of predicting and generalizing from available datasets, being able to understand trends and draw inferences is also an indispensable content [35].

Such constituents of data quality as provenance and integrity, as well as the accuracy, consistency, completeness, originality and timeliness of data sources should play a central role in data literacy education [36]. By being largely standardised, also data governance is applicable to teaching data literacy [37].

A basic level of data literacy education might be provided by all those members of library staff, who face the public. These librarians need to have basic understanding of RDM, the research lifecycle, and RDM. They have to be knowledgeable of RDM resources and know, who is to be contacted with further questions. The advanced RDM service provision aims at providing points of contact for the disciplines. The librarians working on this level, should be aware of the relevant funder requirements and of disciplinary research workflows, including the ways and tools of compiling data management plans (DMPs), as well as being familiar with the practices and requirements of subject data repositories. The specialist level requires understanding of the local, national and global RDM landscape and being able to collaborate with several stakeholders [38].

3. Data-related professional roles

Fulfilling data-related professional tasks is increasingly becoming relevant and urgent for a wider circle of librarians than ever before. Therefore, information professionals' education and retraining are urgent tasks [8].

From the varied types of data professionals, we will focus on data librarians and data scientists. The former continue to follow many paths of traditional librarianship. On the other hand, they are bound to create new library services and focus on acting as facilitators in all stages of the scientific research [11]. Having varied educational and professional backgrounds, and being engaged in different types of work. They are required to have blended domain knowledge and a skill-set that involves technical skills, combined with contextual understanding of a subject [9]. Data librarians' posts may be fulfilled by data generalists, who have broad knowledge of how data is used across several subject areas. They usually teach students, engage with teaching staff members and researchers, working in varied disciplines, while subject specialist must have deep expertise and select skills that enables them to work with specific user groups [39].

The relationship between data librarianship and data science is characterised by similarities and differences. Data scientists are tied to business environments, which is characterised by short-term tasks, reliance on statistical tools, and relational data modelling, which have not been preferred by academic libraries. Nonetheless, data sciences' methods and practices may turn out to

be applicable to data librarianship, especially in regard to analytical and statistical services [40], data visualisation [41], programming languages, and their logic [11]. Familiarity with the disciplinary norms and standards of the given field is a must for both groups of professionals [6]. Giving attention to data quality is a similarly shared issue [42]. Moreover, the basic missions of librarianship, LIS, and data science overlap to a considerable extent as all of them invariably focus on the information (communication) chain, including the creation, dissemination, organisation, storage, and use [43]. We should not forget either that data scientists are bound to be data literate [44]. Besides all of this, we have to underline that focusing on information and meaning that characterises qualitative methods, applied by librarians should be harmonised with centring on data, i. e. examining pattern and syntax and using quantitative methodology, often preferred by data scientists [45].

4. Conclusion

This paper addressed some of the opportunities and challenges that academic libraries must face, when aiming to provide services for supporting data-intensive research. This worldwide challenge has already been answered in several countries, but the overall level of its recognition is low, thus there is a need to raise awareness of its importance. If embraced more widely, varied data-related services might demonstrate that academic libraries are able to play an essential role in research processes [46]. ■

References

1. Pinfield S., Cox A. M., Rutter S. *Mapping the future of academic libraries: a report for SCONUL*. London, 2017. 67 p.
2. Boyce G., Greenwood A., Haworth A., Hodgson J., Jones C., Marsh G., Sadler R. Visions of value: leading the development of a view of the university library in the 21st century. *The Journal of Academic Librarianship*, 2019, 45 (5), 1–8. DOI: <https://doi.org/10.1016/j.acalib.2019.102046>.
3. Pasquetto I. V., Sands A. E., Borgman C. L. Exploring openness in data and science: what is “open”, to whom, when, and why? *Proceedings of the 78th American Society for Information Science and Technology annual meeting*. St. Louis, 2015, 141–144. DOI: <https://doi.org/10.1002/pr2.2015.1450520100141>.
4. Borgman Ch. L. *Big data, little data, no data: scholarship in the networked world*. Cambridge, MIT Press, 2015. 416 p.
5. Kirkwood R. J. Collection development or data-driven content curation? *Library Management*, 2016, 37 (4/5), 275–284. DOI: <https://doi.org/10.1108/LM-05-2016-0044>.
6. Schmidt B., Shearer K. Librarians' competencies profile for research data management. *Joint Task Force on librarians' competencies in support of e-research*

and scholarly communication. Gottingen, June 7–9, 2016. URL: https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf (accessed 10.09.2019).

7. Tenopir C., Sandusky R. J., Allard S., Birch B. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 2014, 36(2), 84–90. DOI: <https://doi.org/10.1016/j.lisr.2013.11.003>. 6a.

8. Robinson L., Bawden D. “The story of data”: a socio-technical approach to education for the data librarian role in the CityLIS library school at City, University of London. *Library Management*, 2018, 38 (6/7), 312–322. DOI: <https://doi.org/10.1108/LM-01-2017-0009>.

9. Burton M., Lyon L. Data science in libraries. *Bulletin of the Association for Information Science and Technology*, 2017, 43 (4), 33–35. DOI: <https://doi.org/10.1080/01930826.2019.1583015>.

10. LERU roadmap for research data. Leuven, 2013. 36 p. URL: <https://www.leru.org/publications/leru-roadmap-for-research-data> (accessed 11.09.2019).

11. Semeler A. R., Pinto A. L., Rozados H. B. F. Data science in data librarianship: core competencies of a data

librarian. *Journal of Librarianship and Information Science*, 2019, 51(3), 771–780. DOI: <http://journals.sagepub.com/doi/10.1177/0961000617742465>.

12. Tenopir C., Talja S., Horstmann W., Late E. Hughes D., Pollock D. Schmidt B., Baird L., Sandusky R. J., Allard S. Research data services in European academic research libraries. *LIBER Quarterly*. 2016, 27 (1), 23–44. DOI: <http://doi.org/10.18352/lq.10180>.

13. Rice R, Southall J. The data librarian's handbook. London, Facet Publ., 2016. 192 p.

14. Huling N, Dallas L. J., Kinder J. B., Whitlatch J. B., Woodard B. Professional competencies for reference and user services librarians. *RUSA*. 2017. URL: <http://www.ala.org/rusa/resources/guidelines/professional> (accessed 12.09.2019).

15. Whitlatch J. B., Bodner N. E., Diefenthal M. Z., Huling N, Kluegel K. M. Professional competencies for reference and user services librarians. *Reference User Services Quarterly*, 2003, 42 (4), 290–295.

16. Cox A. M., Kennan M. A., Lyon L., Pinfield S. Developments in research data management in academic libraries: towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 2017, 68 (9), 2182–2200. DOI: <https://doi.org/10.1002/asi.23781>.

17. Silvello G. Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, 2018, 69 (1), 6–20. DOI: <https://doi.org/10.1002/asi.23917>.

18. Sayre F., Riegelman A. Replicable Services for Reproducible Research: A Model for Academic Libraries. *College & Research Libraries*, 2019, 80 (2), 260–272. DOI: <https://doi.org/10.5860/crl.80.2.260>.

19. Molloy L., Hodson S., Goldstein S., Davidson J. Addressing data management training needs: a practice based approach from the UK. *iPres2012: proc. of the 9th Intern. conf. on preservation of digital objects (Toronto, 1–5 Oct. 2012)*. Toronto, 2012, 249–256.

20. Koltay T. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 2017, 49 (1), 3–14. URL: <http://journals.sagepub.com/doi/10.1177/0961000615616450>.

21. Bugaje M., Chowdhury G. Is data retrieval different from text retrieval? An exploratory study. *International Conference on Asian Digital Libraries*. Cham, 2017, 97–103. DOI: https://doi.org/10.1007/978-3-319-70232-2_8.

22. Faniel I. M., Connaway L. S. Librarians' perspectives on the factors influencing research data management programs. *College & Research Libraries*, 2018, 79(1), 100–119. DOI: <https://doi.org/10.5860/crl.79.1.100>.

23. Bryant R., Lavoie B. F., Malpas C. The realities of research data management. Pt. 1: A tour of the Research Data Management (RDM) service space. Dublin, 2017. 20 p. URL: <https://www.oclc.org/research/publications/2017/oclcresearch-rdm-part-one-service-space-tour.html> (accessed 13.09.2019).

24. Cox A. M. Academic librarianship as a data profession: the familiar and unfamiliar in the data role spectrum. *eLucidate*, 2018, 15 (1/2), 7–10.

25. Ray J. The rise of digital curation and cyberinfrastructure: From experimentation to implementation and maybe integration. *Library Hi Tech*, 2013, 30 (4), 604–622. DOI: <https://doi.org/10.1108/07378831211285086>.

26. Bracke M. S. Emerging data curation roles for librarians: a case study of agricultural data. *Journal of Agricultural Food Information*, 2013, 12(1), 65–74. DOI: <https://doi.org/10.1080/10496505.2011.539158>.

27. Sposito F. A. What do data curators care about? Data quality, user trust, and the data reuse plan. *IFLA WLIC2017 satellite meeting*. URL: <http://library.ifla.org/1797> (accessed 14.09.2019).

28. Thomas C. V., Urban R. J. What do data librarians think of the MLIS? Professionals' perceptions of knowledge transfer, trends, and challenges. *College & Research Libraries*, 2018, 79 (3), 401–423. DOI: <https://doi.org/10.5860/crl.79.3.401>.

29. Johnston L. R., Carlson J., Hudson-Vitale C., Imker H., Kozlowski W., Olendorf R., Stewart C. How important are data curation activities to researchers? Gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly Communication*, 2018, 6, eP2198, 1–24. DOI: <https://doi.org/10.7710/2162-3309.2198>.

30. Poole A. H. The conceptual landscape of digital curation. *Journal of Documentation*, 2016, 72(5), 961–986. DOI: <https://doi.org/10.1108/JD-10-2015-0123>.

31. Lucky S., Harkema C. Back to basics: supporting digital humanities and community collaboration using the core strength of the academic library. *Digital Library Perspectives*, 2018, 34 (3), 188–199. DOI: <https://doi.org/10.1108/DLP-03-2018-0009>.

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

33. Corral S. Repositioning data literacy as a mission-critical competence. *ACRL 2019: Recasting the Narrative (Cleveland, 10–13 Apr. 2019)*. Cleveland, 2019. 7 p. URL: <http://d-scholarship.pitt.edu/36975/> (accessed 16.09.2019).

34. Morrison L., Weech T. Reading data: the missing literacy from LIS education. *The power of reading: proc. of the XXVI Bobcatss symp*. Riga, 2018, 75–80.

35. Mason J., Khan K., Smith S. Literate, numerate, discriminate: realigning 21st century skills". *Proceedings of the 24th International Conference on Computers in Education (Bombay, Nov. 28 – Dec. 2, 2016)*. Bombay, 2016, 609–614.

36. Daraio C., Lenzerini M., Leporelli C., Naggar P., Bonaccorsi A., Bartolucci A. The advantages of an ontology-based data management approach: openness, interoperability and data quality. *Scientometrics*, 2016, 108 (1), 441–455. DOI: <https://doi.org/10.1007/s11192-016-1913-6>.

37. Koltay T. Data governance, data literacy and the management of data quality. *IFLA Journal*, 2016, 42 (4), 303–312. DOI: <http://journals.sagepub.com/doi/10.1177/0340035216672238>.

38. Mattern E., Brenner A., Lyon L. Learning by teaching about RDM: an active learning model for internal library education. *International Journal of Digital Curation*, 2016, 11 (2), 27–38. DOI: <http://dx.doi.org/10.2218/ijdc.v11i2.414>.

39. Federer L. Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association*, 2018, 106 (3), 294–303. DOI: <https://doi.org/10.5195/jmla.2018.306>.

40. Maxwell D., Norton H., Wu J. The data science opportunity: crafting a holistic strategy. *Journal of Library Administration*, 2018, 58 (2), 111–127. DOI: <https://www.tandfonline.com/doi/full/10.1080/01930826.2017.1412704>.

41. Ogier A. L., Stamper M. J. Data visualization as a library service: embedding visualization services in the library research lifecycle. *Journal of eScience Librarianship*, 2018, 7 (1), e1126, 1–11. DOI: <https://doi.org/10.7191/jeslib.2018.1126>.

42. Cao L. Data science: nature and pitfalls. *IEEE Intelligent Systems*, 2016, 31 (5), 66–75. DOI: <https://doi.org/10.1109/MIS.2016.86>.

43. Robinson L. Information science: communication chain and domain analysis. *Journal of Documentation*, 2009, 65 (4), 578–591. DOI: <https://doi.org/10.1108/00220410910970267>.

44. Koltay T. Accepted and emerging roles of academic libraries in supporting research 2.0. *The Journal of Academic Librarianship*, 2019, 45 (2), 75–80. DOI: <https://doi.org/10.1016/j.jcalib.2019.01.001>.

45. Robinson L. Between the deluge and the dark age: perspectives on data curation. *Alexandria*, 2016, 26 (6), 73–76. DOI: <https://doi.org/10.1177/0955749016661067>.

46. Scaramozzino J. M., Ramírez M. L., McGaughey K. J. A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, 2012, 73 (4), 349–365. DOI: <https://doi.org/10.5860/crl-255>.

Статья поступила в редакцию 23.09.2019
Принята для публикации 25.10.2019

Received 23.09.2019
Accepted 25.10.2019

Сведения об авторе:

Колтай Тибор, доктор наук, профессор,
Институт обучающихся технологий, Универ-
ситет им. Кароя Эстерхази;
e-mail: koltay.tibor@uni-eszterhazy.hu

Information about the author

Koltay Tibor, Dr. habil., PhD, Professor, Institute
of Learning Technologies, Eszterházy Károly
University; <https://koltaytibor.uni-eszterhazy.hu/en>; https://www.researchgate.net/profile/Tibor_Koltay;
e-mail: koltay.tibor@uni-eszterhazy.hu