

Is data management a new “digitisation”? A change of the role of librarians in the context of changing academic libraries’ tasks

Anna Walek

Library Director, Gdańsk University of Technology Library, Poland

E-mail address: anna.walek@pg.edu.pl



Copyright © 2018 by Anna Walek. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:
<http://creativecommons.org/licenses/by/4.0>

Abstract:

Academic libraries’ tasks have been evolving over the years. The changes have been stimulated by appearing of electronic resources, automated library systems, digital libraries and Open Access (OA) repositories. Librarians’ tasks and responsibilities in the academic environment have been evolving in accordance with new tasks they were expected to assume. A few years ago there was a discussion during which an attempt was made to answer the question if a digital librarian is still a librarian. Digital librarians were expected to acquire a number of new skills and knowledge which were supposed to allow them to manage and organize the digital library and handle all the specialized tasks of massive digitization, storage, access and reference services.

The popularization of OA gave rise to new changes. The role of librarians evolved towards becoming specialists who not only supervised the process of uploading resources on open repositories, provided the training for academic staff, but also promoted the idea of OA.

The evolution of the idea of open access to the results of scientific research towards open access to research data, the advent of Data Science and Open Research Data (ORD) management were all lively discussed. The academic staff were faced with a new and difficult tasks of collecting, describing and disseminating the research data. Moreover, the academics were supposed to take advantage and care of the citation impact of their scholarly works. As a result, librarians started learning about ORD management and the strategies of supporting academic institutions’ further development. For the majority of librarians it was a natural way to develop new skills and knowledge. In the course of bringing about the changes a new job appeared – data librarians, soon becoming sought-after specialists at universities all around the world.

The article discusses the issues concerning the role of data librarians and their tasks resulting from European Union (EU) grant programmes and the guidelines of the publishers who are supposed to publish the research publications together with the research data. On the basis of this analysis the author will present the strategy for the creation of the Open Science Competence Centre at the Gdańsk University of Technology (GUT)..

Keywords: Open Research Data, data librarian, databrarian, academic libraries, Gdańsk University of Technology Library

Introduction

After almost 80 years we can see that Ranganathan's fifth law -- "The library is a growing organism" is still valid (Ranganathan, 2006).

The libraries are evolving and the tasks of librarians are changing. Only one thing is unquestionable – academic libraries serve the academic community as important partners in research process and information management.

In Poland the stage of implementing computer-based library systems was followed by the stage of libraries' involvement in the process of collections digitizing in the 1990s. Back in the 1990s the digitization of the cultural and research content was developing dynamically. The libraries responded to the emerging trends and needs by creating a new career specialization in the field of librarianship – a digital librarian. Multimedia nature of digital libraries required the digital librarians to become specialists who were supposed to manage digital libraries. The new tasks included handling the specialized tasks of massive digitization including graphic processing, creating websites, making metadata records based on Dublin Core Element Set and applying relevant licenses to make them accessible.

In the course of time the range of a digital librarian's tasks broadened including more and more duties associated with the creation of e-collections in digital libraries as well as management of the digital library resources (DelRosso, Lampert, 2013).

Along with the development of Open Access (OA), librarians became engaged in the process of depositing digital publications in OA repositories. Furthermore, librarians were developing their competences for creating OA policies, as well as developing institutional repositories services.

The next step was to disseminate research data management and curation needs.

Funding agencies and research institutions began to require data management policies. At the same time scientific journals recommended sharing data sets which allow researchers to verify the research results included in scientific articles. Nowadays modern digital scholarship requires librarians to face yet another challenge. They are obliged to adjust their services to suit the new needs and navigate a complex research and publication world.

At the present time librarians are supposed to specialize in new areas and a new specialization has been created – a data librarian. The responsibilities of the new type of a librarian embrace new skills and knowledge, and require librarians to become actively engaged in the research process.

Open Research Data

There are many definitions of research data. One of them was published at University College London (UCL) Blog in 2015.

„Research data are the original sources or material that you have created or collected to conduct your research project. They can be digital or non-digital. The response to your research question is based on the analysis of these research data” (UCL, 2015).

More and more research data is born in a digital form, although physical forms are still common within some fields of study. Researchers must keep track of both kinds of data usage simultaneously. Research projects are producing huge sets of data that would be impossible to manage without the aid of computers to process them.

Greater collaboration among researchers, engineers and computers scientists became a fact in all fields of study and librarians are being brought into these partnerships to contribute needed expertise in data management and preservation.

There are generally two types of data: qualitative (dealing with things numerically) and quantitative (descriptive in nature and dealing with the quality, categorization). Most researchers usually gather both types of data in their research (Krier, Strasser, 2014).

We can find also other categorizations of data. For example we can distinguish primary data (original data that arise from a particular experiment or observation), which are gathered and maintained by researchers and secondary data, often used by researchers but originally created by someone else (Krier, Strasser, 2014).

Both primary and secondary data take many forms, for example: observational data – data that have been gathered from observing a particular phenomenon, and experimental data – in contrast – derived from controlled, randomized experiments.

They both still could be produced by human researchers, taking notes and describing observations, but more and more often, data are gathered with the use of computers, sensors and other monitoring tools. These tools produce far larger data sets for researchers to collect and analyse (Krier, Strasser, 2014).

Data lifecycle

There are two ways to think about the lifecycle of data. We can consider it from a researcher's perspective and from a curator's perspective. The first one "Research data lifecycle" (UK Data Archive, 2018) created by the UK Data Archive could be useful if you think about the data from a researcher's perspective. It covers the lifespan of research data from the moment of its creation through the reuse of the data. The sequential steps of this lifecycle are: creating data, processing data, analyzing data, preserving data, giving access to data and reusing data.

On the other hand The Digital Curation Centre has created a "Curation lifecycle model" (Digital Curation Centre, 2018) that lays out all the processes and components involved in data curation from an archivist's or a curator's perspective.

In LITA Guide "Data Management for Libraries" we can find a short definition of data curation: "Data curation is the management of data once it has been selected for preservation and long-term storage" (Krier, Strasser, 2014).

The sequential steps of the curation lifecycle are: creating or receiving data, appraising and selecting data, ingesting, performing preservation actions, storing data, accessing data for use and reuse and transforming data.

Many researchers and librarians are involved at various stages of the data lifecycle, both during the research process and during the curation process.



Libraries have begun stepping in to assist researchers to data management. Some of them considered contributing their skills to support researchers an opportunity for professional growth, while others decided to learn new skills because of an external pressure. As Laura Krier said “in any case, librarians have a great opportunity to expand our services in ways that can benefit faculty, build stronger relationships between libraries and research communities, and continue to play a role in the preservation of scholarly communication”.

Data Management Plan (DMP)

Some funding bodies have defined requirements that plans for managing the data produced during a research project should include in grant proposals. These plans show how research data will be collected, organized, managed, and preserved during the project and after. What such a plan looks like depends on the specificity and conditions of the project. In general, these plans require a description of the data that will be produced or used, formats, metadata standards that will be used to store and organize data, the how and where of the data storage and accessibility.

Research data services

“Research data services are services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion” (Tenopir, Sandusky, Allard and Birch, 2012).

Research data are becoming more and more important for researchers who are beginning to see the advantage in sharing and reusing data sets, which has been confirmed by the results of the research conducted by Digital Science-Figshare. “The State of Open Data Report 2016” (Treadway, Hahnel, et. al. 2016) looked at global attitudes towards Open Research Data (ORD). It included survey results of more than 2,000 respondents whose key finding was that ORD had become more embedded in the research community. The 60% of the survey respondents were aware of the same importance of data sets quoting and research article quoting. More than 50% of the survey respondents agreed that they would need more assistance in data disseminating, choosing the relevant licenses and managing research data. The academics had a number of questions and express their doubts about the process of internationalization of research and funding agencies policies concerning research data (e.g. Horizon 2020, Wellcome Trust). Another aspect discussed in “The State of Open Data Report 2016” was the lack of appropriate infrastructure to store and organize data as well as inform about the data. The Report also shows how significant research data is, the medical data in particular in the context of rescuing human life, producing new drugs and creating innovative solutions and applications.

Another survey on ORD conducted by Digital Science-Figshare in 2017 – “The State of Open Data Report 2017” (Digital Science et. al., 2017) confirmed the results of the previous Report. “The State of Open Data Report 2017” showed the growth of the number of respondents (2, 300) from all over the world. 82% of the survey respondents recognized the necessity of research data dissemination (10% growth in the number of respondents supporting ORD dissemination in comparison with 2016). The number of survey respondents who would like to reuse the research data also increased by 10% - reaching 80%. All the respondents emphasized the lack of proper education and support for research data management and dissemination. the survey shows that researchers expect the support in this respect from

universities they are affiliated with. The main cause of motivation of the respondents to disseminate their scholar output is the awareness that visibility of their research and recognition in the scientific world may significantly increase. Last but not least, what encourages the faculty to share their research findings is the increased number of citations for their papers.

Another source – Council in Library and Information Resources report “The Problem of Data” note that researchers “understand that poor data management can be costly to their research and that access to greater technical expertise, through either a consultant or additional training, would be useful for their work” (Jahnke, Asher, Keralis, 2012). According to the report only few researchers are happy with their own data management practices. They usually complain that they do not have time for the administrative and organizational work and that they have never received explicit training in data management practices.

Data librarian

For a number of years the role of libraries and librarians in the data management and data curation processes has been discussed. The aim of the surveys conducted in various countries was to define data librarians’ professional identification, education background, experience, knowledge, skills and qualities that would allow them to perform the new duties in an effective manner and in line with new expectations.

The results of the surveys displayed the need for a set of particular qualities and skills that were essential for the librarians specialising in data management and data curation. The results of Kennan’s survey presented during IFLA Conference in Columbus (2016) concurred with the earliest findings of Kellam (Kellam, 2011) and Tenopir (Tenopir, Hughes, et. al., 2015) and later of Mizzy and Hayslett (Mizzy, Hayslett, 2016).

In 2015 Mary Anne Kennan conducted a survey among data managers, librarians, their supervisors and employers, to investigate key knowledge and skills for new professionals coming into data practice in universities and other scientific and research organizations. The results of these studies were announced during the IFLA conference in Columbus in 2016.

The following questions were asked:

- Where, and in what roles, are data librarians, managers and scientists likely to be employed?
- What knowledge and skills are required of people working in data librarianship and management roles?
- What are the educational and training requirements for data librarians and data managers?

The survey’s participants reported that the employment pattern differs across organizations. They worked in the libraries and other information departments such as IT, and in research offices. There was also a huge range of job titles (e.g. Data Librarian, Data Manager, Data Specialist) used by data specialists and curators.

As Kennan reports, almost all the employers indicated that major set of skills they required, or that they were looking for in librarians and other data specialists, were more generic than field specific and these could be labelled as interpersonal skills and behavioural characteristics. They said, for example, that they are looking for people who are flexible, that is why they are looking more at attributes than at particular knowledge or skills.

The most commonly mentioned interpersonal skills were: high level communication, advocacy, negotiation and capability building skills. In terms of personal characteristics they said that a data librarian or specialist should be: comfortable with change, have a service philosophy, willingness to learn, discretion, “boundless curiosity”, be adaptable, assertive and open to new experiences. Several participants also mentioned the ability to network and use network to learn.

Almost all participants mentioned the importance of contextual knowledge about the research environment of the university or scientific organisation, related funding agency policies, research measures and research evaluation activities. Moreover, some other factors were considered vital including understanding or being open to discipline specific research life cycles and cultures, processes, ethics, disciplinary research methods, intellectual property, licensing and policies, access norms and cultural sensitivities. It was also mentioned that subject specialty in specific areas would be encouraged.

The second most frequently mentioned requirement was knowledge in the data domain and related skills. Knowledge and skills related to facilitate data sharing, linked data, data management lifecycle, quality control, data processing, data management planning and ability to understand and support data storage requests were mentioned as crucial for the development of data librarianship. It is worth emphasising that data professionals should be aware of the legal and regulatory frameworks relevant to data, as well as ethics, copyright and creative commons issues.

Almost every participant mentioned the importance of knowledge about and skills in applying metadata, metadata standards, ontologies, structured information as well as possessing data mapping, cataloguing and harvesting competencies. Other important skills and knowledge indicated in the survey were as follows: enabling access, user experience, web interfaces, Graphical User Interfaces (GUI's), Digital Object Identifiers (DOI's), back up, version control and naming conventions.

A number of IT skills were also recognised as significant factors. Data specialists should have a sufficient understanding of IT to be able to communicate with both IT specialists and researches. Specific IT knowledge and skills mentioned in the survey were: XML, database design, APIs, user centered design, natural language processing tools and the Internet of things. Last but not least, the participants added to the above mentioned list of data librarians' skills awareness of project management, relationship management and change management principles.

Taking into account the education and previous experience of data professionals, it is apparent that most of them were librarians, some with prior qualification in IT. Another group of the librarians working as data professionals was made up of people who simply had some practical experience in IT without a college degree. These people had usually learned data specific knowledge and skills “on the job” working in the data management space in academic libraries or through additional targeted professional development.

What is also important, after surveying and interviewing a cross-section of librarians supporting scientific research authors by Mizzy and Hayslett (Mizzy, Hayslett, 2016) – the authors concluded that there is no one path to data librarianship. Some people start their professional careers in science and become librarians later, and some others start in librarianship and specialize in science over the course of their professional careers. The authors

prepared interview questions, based on the Kellam's survey about "experiences becoming and being data librarians" (Kellam, 2011).

European Open Science Policy

European Open Research Data Policy was developed in July 2012. The European Commission (EC) published the Communication "Towards better access to scientific information" (European Commission, 2012a) and recommendations (European Commission, 2012b) regarding the implementation of Open Access policy in relation to research data, including Horizon 2020. It was agreed that, as in the case of FP7 and the OA pilot program, Horizon 2020 would have a pilot program on the sharing of research data. This initiative is part of a broad project called "The European Research Area".

In October 2013 EC published a report on public consultation on Open Research Data (European Commission, 2013) and two years later adopted the "Digital Single Markets Strategy" (European Commission, 2015). Another action of EC was to set up a cloud for research data – The European Open Science Cloud, an infrastructure which is supposed to provide web hosting, make ORD available to European institutions and strengthen the position of the European Union. It is crucial to increase Europe's competitiveness through development and implementation of ORD policy and innovative solutions in research data management.

An important step towards strengthening the policy and monitoring its progress was the establishment of an expert committee „Open Science Policy Platform” in 2016. Since then the Committee has been monitoring the implementation of Open Science recommendations.

The result of later activities was the publication of the instructions called "FAIR Data Management in Horizon 2020" (European Commission Directorate General for Research & Innovation, 2016) in 2016 and "FAIR Open Research Data report" in January of 2017 (European Commission, 2017).

The European Commission recommended member states of the European Union to develop their own national Open Access policies.

Only by taking into consideration the conditions the academic libraries function in and the legal regulations existing in different Member States, can one understand the role of the libraries and their overall development in implementing the idea of OA and ORD throughout the European Union. In Central and Eastern European countries providing Open Access to results of scientific research takes place on a voluntary basis and is not a common practice. In other countries, e.g. United Kingdom national OA mandate has been implemented while in Germany there are different legal solutions concerning OA in each state of the country. Consequently, the degree of involvement in implanting OA of various universities across the European Community varies significantly (SPARC Europe, 2018).

Gdańsk University of Technology perspective

There is no research data policy in Poland. However, the need for action in this area is evident. Universities and libraries are beginning to take certain actions.

Gdansk University of Technology (GUT) Competence Centre will provide expertise and support including trainings among scholars about different aspects of Open Science and face to face consultations.

First of all, our skilled-based information service will be used as an on-site tailored training centre that will feature courses and events covering wide range of topics such as:

- Research Data Management Plan (how to create Plan that is for example required by Horizon 2020 Data Pilot Program)
- Open Research Data in different scientific disciplines (how to make research data opened?)
- Open Scholarly Communication Tools
- Open Evaluation and Peer-Review Systems
- Open Metrics
- Creating documentation and metadata
- Handling sensitive data
- Research Data and Open Access
- Backup of data and its security
- Different data formats
- Metadata portals
- Ethical issues of sharing data
- Legal aspects of sharing Research Data
- Open Research Data repositories and infrastructure
- Citing Research Data (how to cite datasets?)
- Open Research Data and Publishers' policies
- Digital preservation of data.

The specific tasks that the Competence Centre will undertake will respond to the specific needs of the academic staff. In order to set priorities and to plan appropriate staff training, a survey and a detailed needs analysis will be carried out.

GUT Competence Centre will be dedicated for all level researchers from all scientific disciplines. Similar centres have already operated in Europe e.g. at Cambridge University, Delft University or at Leiden University (The Centre for Digital Scholarship). GUT Competence Centre will be the first of this kind in Central and Eastern Europe and will have the important role in raising awareness on Open Research Data engaging stakeholders in Poland and other countries.

We would like to support different professionals from research and industry sectors. Research data education and trainings should be recognized as essential for open-oriented scientists. GUT Competence Centre will contribute to raising standards and skills regarding wide range of research data such as social networks data, environmental, business and economic data.

Conclusion

Libraries' services provided to researchers and research institutions in order to assist them with the research data management process are evolving. The new duties of the libraries differ from the previous tasks they have been responsible for so far. Participating in the data management librarians become a part of the research process from the very beginning of the data lifecycle. Their knowledge and engagement are essential from a very early moment of DMP creation

through collecting, describing, preserving and curating data. Additionally, librarians will also be responsible for ensuring that data sets prepared by other teams are properly used and cited. The services provided by libraries so far, e.g. digitization and digital resources management including Open Access implementation were about making already existing resources available to the users. The new tasks related to research data management pose different challenges for librarians that they will have to live up to. Librarians future roles and responsibilities will differ significantly from the tasks performed by library professionals today.

The top challenge for the librarian of the future will be how to balance the demand for research data management with all the other tasks librarians are responsible for. The number of electronic resources is increasing and paper books will not disappear from most library shelves, so librarians will be obliged to combine traditional tasks with duties involving quickly changing technology.

References

DelRosso, J., Lampert, C. (2013). So you want to be a digital librarian - what does that mean? [Electronicversion]. In: J. D. Monson (ed.), Jump-start your career as a digital librarian: A LITA guide(pp. 3-19). Chicago,IL: American Library Association. Retrieved from <https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=2013&context=articles>

Digital Curation Centre (2018), DCC Curation Lifecycle Model, Retrieved from www.dcc.ac.uk/resources/curation-lifecycle-model

Digital Science, Hahnel, M, Treadway, J., Fane, B., Kiley, R., Peters, D., et al. (2017), The State of Open Data Report 2017. figshare. Paper.
DOI:<https://doi.org/10.6084/m9.figshare.5481187.v1>

European Commission (2012a), Towards better access to scientific information: Boosting the benefits of public investments in research. Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, Brussels, 17.7.2012 COM(2012) 401 final. Retrieved from https://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

European Commission (2012b), Commission Recommendation of 17.7.2012 on access to and preservation of scientific information {SWD(2012) 221 final} {SWD(2012) 222 final} Brussels, 17.7.2012 C(2012) 4890 final. Retrieved from <https://ec.europa.eu/digital-single-market/node/66216>

European Commission (2013), Final Report of the European Commission Public Consultation on Open Research Data. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/final-report-public-consultation-open-research-data>

European Commission (2015), A Digital Single Market Strategy for Europe. Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, Brussels, 6.5.2015 COM(2015) 192 final {SWD(2015) 100 final}. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52015DC0192>



European Commission (2017), Open Science Monitoring Impact Case Study – Data FAIRport. Retrieved from

https://ec.europa.eu/research/openscience/pdf/monitor/fairport_case_study.pdf

European Commission Directorate General for Research & Innovation (2016), H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, Version 3.0 26 July 2016. Retrieved from

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Jahnke, L., Asher, A., Keralis, S.D.C. (2012), The Problem of Data (Washington, DC: Council on Library and Information Resources) Retrieved from <http://www.clir.org/wp-content/uploads/sites/6/pub154.pdf>

Kellam, L. M. (2011), Data librarianship: a day in the life, In: Numeric Data Services and Sources for the General Reference Librarian, ed. L. Kellam and K. Peter (Oxford: Chandos Publishing) pp. 151-160 DOI: <https://doi.org/10.1016/B978-1-84334-580-0.50005-5>

Krier L., Strasser C. A. (2014), Data Management for Libraries. A LITA Guide, Chicago

Mizzy, D., Hayslett, M. (2016), Data Librarianship: A Day in the Life-Science Edition In: Databrarinaship. The Academic Data Librarian in Theory and Practice, Chapter 21, ACRL, Chicago, Illinois, pp. 335-354.

Ohaji, I. K. (2016), Research Data Management: An Exploration of the Data Librarian Role in New Zealand Research Organisations, Victoria University of Wellington 2016 (PhD thesis). Retrieved from <http://researcharchive.vuw.ac.nz/handle/10063/5657>

Ranganathan, S. R. (2006), The Five Laws of Library Science, New Delhi [India]: Ess Ess Publications

SPARC Europe (2018), An Analysis of Open Data and Open Science v2.1 (January 2018). Retrieved from <https://sparceurope.org/download/2285/>

Policies#in#Europe,#May#2017 Tenopir, C. et al., (2017). Research Data Services in European Academic Research Libraries. LIBER Quarterly. 27(1), pp.23–44. DOI: <http://doi.org/10.18352/lq.10180>

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

http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., ... Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future? Journal of eScience Librarianship, 4(2). <http://dx.doi.org/10.7191/jeslib.2015.1085>



Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2012). Academic librarians and research data services: Preparation and attitudes (Session 116). Paper presented at IFLA World Library and Information Congress, Helsinki, Finland. Retrieved from <http://conference.ifla.org/past-wlic/2012/116-tenopir-en.pdf>

Treadway, J., Hahnel, M., Leonelli, S., Penny, D., Groenewegen, D., Miyairi, N., Hayashi, K., O'Donnell, D., Digital Science, and Hook, D. (2016). The State of Open Data Report. figshare. Paper. DOI:<https://doi.org/10.6084/m9.figshare.4036398.v1>

UCL (2015) What are research data? In Research Data Management Blog 15 September 2015 Retrieved from <https://blogs.ucl.ac.uk/rdm/2015/09/what-is-research-data/>

UK Data Archive (2018), Research Data Lifecycle. Retrieved from <http://data-archive.ac.uk/create-manage/life-cycle>.