

Data librarianship as a field study

Biblioteconomia de dados como campo de estudos

Alexandre Ribas SEMELER¹  0000-0002-8036-4271

Adilson Luiz PINTO²  0000-0002-4142-2061

Abstract

Data are generated during all human activities related to digital technology. In recent times, scientific research has increasingly opted for digital data as its primary source of data; data definition changes for different disciplines and researchers. In this context, we study the main characteristics of data librarianship as a specialized field of traditional librarianship concerned with data use in libraries. Our work is organized as follows: First, we present a proposed Venn diagram on the theoretical foundations of data librarianship; then, we point out the core skills needed by data librarians. Based on a non-exhaustive literature review, we point out the main topics of research in data librarianship. We describe the significance of research data, data management, data curatorship, and data repositories. Finally, we list a few certification courses in data librarianship. We conclude that data librarianship plays a dynamic role in the practical application of data technologies in libraries, and that professional development, certification, and training in data librarianship are interdisciplinary tasks linked to digital technologies.

Keywords: Information management. Professional competences. Research data. Scientific data management. Specialist librarianship.

Resumo

Dados são gerados durante todas as atividades humanas ligadas à tecnologia digital. Atualmente, os dados digitais são uma fonte primária para a geração de pesquisa científica; no entanto, a definição de dados varia entre as diferentes disciplinas e entre os pesquisadores. Nesse contexto, este estudo analisa as principais características da biblioteconomia de dados como um campo especializado da biblioteconomia tradicional, que se preocupa com o uso de dados em bibliotecas. O trabalho está organizado da seguinte forma: primeiro, apresenta uma proposta de diagrama Venn sobre os fundamentos teóricos da biblioteconomia de dados e, depois, aponta o núcleo de habilidades necessárias aos bibliotecários de dados. Com base em uma revisão não exaustiva de literatura, são assinalados os principais tópicos de pesquisa em biblioteconomia de dados. O estudo descreve o significado de dados de pesquisa, seu gerenciamento e curadoria, bem como seus repositórios e, por fim, lista alguns cursos de certificação em biblioteconomia de dados. Conclui-se que a biblioteconomia de dados desempenha um papel dinâmico na aplicação prática das tecnologias de dados em bibliotecas, e que o desenvolvimento profissional, a certificação e o treinamento na área são tarefas interdisciplinares ligadas às tecnologias digitais.

Palavras-chave: *Gestão da Informação. Competências profissionais. Dados de Pesquisa. Gestão de dados científicos. Biblioteconomia especializada.*

Introduction

Data is generated during almost all human activities linked to digital technology. The data concept depends on context, researcher, and area of knowledge in which the research is conducted and depends on how the data

¹ Universidade Federal do Rio Grande do Sul, Instituto de Geociências. Av. Agronomia, 9500, Campus do Vale, 91501-970, Porto Alegre, RS, Brasil. *Correspondência para/Correspondence to:* A. R. SEMELER. E-mail: <alexandre.semeler@ufrgs.br>.

² Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Departamento de Ciência da Informação. Florianópolis, SC, Brasil.

Received on May 19, 2020, final version resubmitted on July 28, 2020 and approved on September 11, 2020.

Como citar este artigo/How to cite this article

Semeler, A. R.; Pinto, A. L. Data librarianship as a field. *Transformação*, v. 32, e200034, 2020. <https://doi.org/10.1590/2318-0889202032e200034>



is generated and at what stage of the data life cycle the information will be present. A computer scientist may use the term to refer to algorithms and scripts in a computer network. A statistician may consider data a set of natural numbers. A social scientist may consider data the interactions and behavior of people in a specific community. So, how can librarians use the concept of data in libraries?

In this context, this study aims to present the main characteristics of data librarianship as a field. Data librarianship is not something new; it has been discussed since the 1960s in North America and Europe. Today, data librarianship emphasizes is the creation of library services related to the use of digital data, data repositories, data curation, and data management in libraries. Data librarianship is influenced by directly influenced by the impact of data technologies and the new data-intensive research.

Data librarianship requires academic and professional skills, including traditional areas in librarianship, as reference service, and support to academic research. Nowadays, data librarians need data skills as data transformation, data manipulation, and data analysis. A data librarian needs to combine relevant skills and knowledges for data management and data curation.

Data librarianship can be discussed as a theoretical background to enable data librarians to participate in these activities. We propose a methodological pillar to investigate data librarianship as a field, through a Venn diagram, we will describe the main fields that influence data librarianship. In other words, with the union and intersection between information science, e-science, and data science. These knowledge fields represent the theoretical and practical links between areas that make up data librarianship as an interdisciplinary field.

Our work is organized as follows: first, we present a proposed Venn diagram for data librarianship, then we point out the suggested skills to data librarians. We briefly describe the essential meaning of research data, data management, data curation, data repositories, and some courses related to data librarianship certification. To delimit our proposal, we report theoretical issues and presents relevant research findings on these topics, based on a non-exhaustive review of recent literature extracted from databases such as Web of Science, Scopus, ResearchGate, and Google scholar.

Data librarianship Venn Diagram

The Venn diagram is used in mathematics to represent the union and intersection between knowledge fields, which interrelate with each other through theoretical sets. The sets of the diagram represent problems related to these sets and their theoretical connections. Thus, this section of the study applied a Venn diagram to describe the theoretical backgrounds of data librarianship as a field of study related to information science, e-science, and data science. Data librarianship is suitable for practical applications in information science and data science as a result of data-intensive research. Thus, this paper proposes a Venn diagram to present and summarize the methodological pillar to investigate of data librarianship (Figure 1).

The Venn diagram describes the union and intersection between information science, e-science, and data science, but we will not talk about every facet explored in the fields under review. It helps to conceptualize a new way of practicing data librarianship. Figure 1 gives an idea of what data librarianship involves. From the diagram, data librarianship is the intersection of the three theoretical sets. The set elements represent the theoretical and practical connections between the sets, and they indicate the fields that constitute data librarianship as an interdisciplinary field.

The first set is information science. Information science does not only focus on the use of technology and digital objects but much more on the nature of information and how people interact with it. However, in this paper, information science refers to the specific application of data technologies. Information and technology concepts are fundamental to the comprehension of data librarianship in the context of information science.

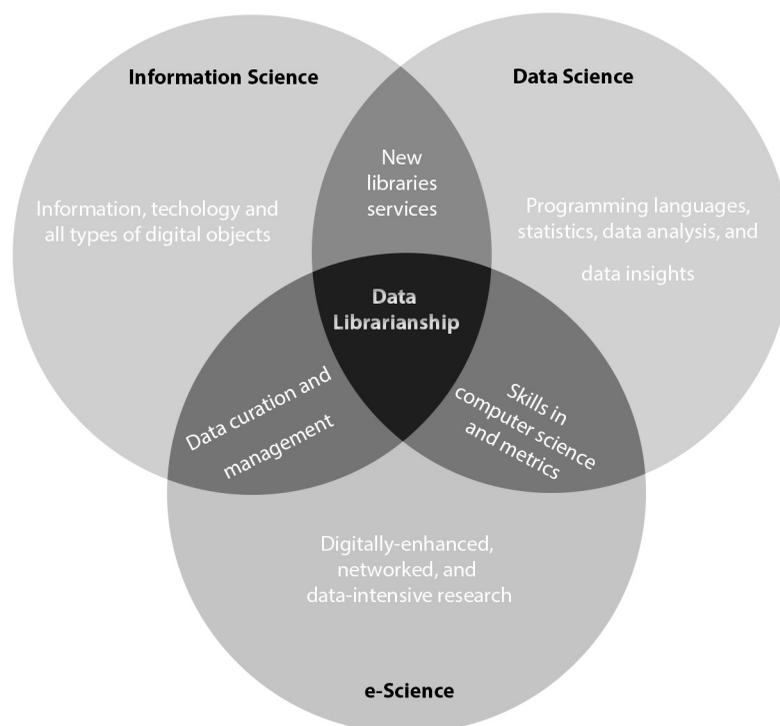


Figure 1. Data librarianship Venn Diagram.

Source: Adapted from Semeler and Pinto (2019b), Semeler and Pinto (2020).

Information Science is a field of study where interdisciplinary issues circulate. The theoretical argument takes place mostly through dialogue between classic information science texts such as Borko (1968); Buckland (1991); Saracevic (2010); Bawden and Robinson (2012); Morville (2014); These authors take the impact of technology on scientific knowledge generated in information science as a point of view, that is, the influence of technology on society and the interaction between the concepts of information and technology applied to information science issues.

Information science is an interdisciplinary field that comprises several disciplines. Thus, the one focus of study in information science is information and the revolutions caused by the use and reuse of digital data technology. Digital data are input for new technological systems; digital data is a new form of information composed by the human behavior of mediated data and generated by the data-intensive use of technologies necessary for the current human activities of work, health, education, security, sports, etc.; a technological continuum lives between information science and digital data objects. This technological continuum highlights the concept of data as an indicator of a new paradigm of studies in information science. The properties of information now transform when inserting digital contexts of artificial intelligence, big data, internet of things applied to understand the behavior of data generated during human activities mediated by data technologies (Semeler; Pinto; Vianna, 2017; Semeler; Pinto, 2020).

The second set is e-science. E-science is an innovative field that incorporates technical tools and devices built with contemporary technology into science. It is the science that uses large datasets to perform data-intensive research. These research trends have influenced some scholars to characterize the new research landscape as Research 2.0, Science 2.0, or e-science (Koltay, 2019).

Data librarianship in the context of data-intensive research is related to e-science and data science:

With the research landscape tilting toward a more intensive research, propelled by the power of information and communication technologies, the research community is overwhelmed with tons of research data generated on a daily basis. Research data is hailed as the cornerstone for current and future science discoveries (Chawinga; Zinn, 2020, p.1).

E-science is a new way of practicing science based on technical tools and devices built with contemporary data-intensive technology. It refers to the use of computational technology in scientific investigation processes, including preparation, experimentation, data collection, result discovery, storage, maintenance, and accessibility of all scientific research results. Thus, e-science provides information and scientific knowledge for data-intensive use via technology (Semeler; Pinto; Vianna, 2017, Semeler; Pinto, 2020).

Data science is a means of representing data-intensive research in diverse fields. The third set is based on skills, methods, techniques, and technologies of statistics and computer science used to extract knowledge and create new products and services from data. The prioritization of data technology by scientists is a common variable in almost all areas of scientific knowledge.

Data science and information science are twin disciplines by nature. According Wang (2018) that analyzed existing work in information science and data science the basic idea is that the nature of data science is consistent with those of information science and that they are closely interrelated and together form the components of information research.

Currently, data science combines the techniques of computer science and statistics. It encompasses specific knowledge domains, including programming languages and scientific methodologies for data collection, manipulation, analysis, and visualization. The disciplinary bases of data science combine a series of knowledge, such as data mining, machine learning, system design, distributed computing, statistics, industrial engineering, domain knowledge, visualization, and big data. Under this view, it is noted that data science has emerged as a multidisciplinary effort that combines numerous skills and various fields of knowledge, which allows for the thinking or prediction of prototypes and demonstrations that help to illustrate data-intensive research (Semeler; Pinto, 2019b; Semeler; Pinto, 2020).

Finally, contained in the core of the diagram is data librarianship, represented as the central intersection between the three sets (information science, data science, and e-science) representing the impact of data technologies and new data-intensive research in libraries. It relates to the interest of librarians in understanding the use and reuse of data technologies, which can be used to create library services (Xia; Wang, 2014; Kellam; Thompson, 2016; Rice; Southall, 2016; Koltay, 2017).

We believed that a data librarian needs to possess relevant competences for data management, data curation, and to support of research data services. Tenopir *et al.* (2014) argued that data librarians could be involved in the research process and be actively committed to data curation and data management. Also, it is a research tendency connected to academic libraries in North America and Europe. Data librarians require expertise to extract, analyze, and present data and some data librarians can work as reference librarians, data curators, library instructors, data support specialists and researchers. The suggested additional roles for libraries include training researchers, performing data archiving and preservation tasks, and training data librarians on data analysis.

Data librarianship skills

Data librarians, works with researchers from the first stages of research, to assist them with research documentation processes and ensure that digital data will be preserved, usable, and reusable in the long run.

Data librarians are professionals, trained and specialized in the curation, preservation, and archiving of data. More recently they have been involved in advocacy, research data management, and delivering training at universities, research institutes, and their libraries (Koltay, 2020, p.4).

Professionals in the data librarianship field have aggregated to themselves the core values of traditional librarianship, which includes knowledge of the processes of maintenance (selection, acquisition, preservation, and management of collections), and provision of reference services, providing technical assistance to ensure the preservation and continuous access to digital collections for an entire community of users (International..., 2016).

Data librarians need to know the skills required in data science. Data librarians also needs to possess the same standard development skills as a data scientist. The skills required of a data scientist include those of a hacker, scientist, quantitative analyst, trusted adviser, and expert businessperson. These skills do not require the data librarian to undertake a computer science, statistics, or business course; however, he/she needs to have at least an intermediate level of knowledge in these areas. Data librarians need to understand how to integrate multiple data technologies into library services. He/she needs the ability to demonstrate hacking skills to program, extract, and structure data. He/she also needs to know the basic methods and tools adopted in metrics studies (bibliometric and informetric). Finally, data librarians need to acquire some of these skills gradually (Xia; Wang, 2014, Semeler; Pinto; Rozados, 2019a).

The work of a data librarian involves not only data manipulation, but also human interpretation, decision making, and communication of that data. McCaffrey and Giesbrecht (2016) suggest three areas of activity: (a) data management and curation, (b) data visualization and geospatial representation; and (c) advanced data reference services. In this regard, Kennan (2016) presents the necessary skills for data librarians at research organizations and universities in North America and Europe (Chart 1).

Chart 1. Required skills for data librarians.

Skills	Description
Interpersonal and behavioral characteristics	Capability related to formal communication, such as technical document writing and case studies. Another point associated with this profile is the capability to adapt to new experiences and constantly seek updates.
Contextual knowledge about institutional environment	This ability involves knowledge about funding policies offered by agencies that support scientific research. This ability demands the understanding of ethical procedures in the scientific process, disciplinary research methods, scientific communication methods, intellectual property, access methods, copyrights, creative commons and royalties.
Data usage	This includes the understanding of data types (quantitative, qualitative), patterns and metadata schemas (Dublin Core, Resource Description Framework), as well as questions related to unique identifiers (Digital Object Identifiers) and the preservation of digital data.
Knowledge about data manipulation technologies	The most relevant technology knowledge for data librarians includes programming languages (Python, Structured Query Language, Java), design and structuring of databases, Application Programming Interface for data retrieval, user-centered design, natural language processing tools, the Internet of Things, and Big Data.

Source: Adapted from Kennan (2016).

The author includes the need for a data librarian to adapt to new experiences and continually pursue professional development. She also emphasizes the need for these librarians to know the research environment of the university or scientific organization in which they work.

A data librarian must transform data into something reusable and shareable that can be preserved for a long time. He/she must have the knowledge of metadata schemas, data repositories, datasets that are used by researchers, as well as a good notion about diverse methodologies and scientific disciplines. However, most importantly, he/she must possess relevant skills for data collection, data description, data literacy, system construction, elaboration of

reference services, institutional data management, and policy creation services (Dekker; Lackie, 2016; Kennan, 2016; McCaffrey; Giesbrecht, 2016).

Ideally, a data librarian participates in decision-making processes at all stages of scientific research. A data librarian is an extension of academic librarian. To facilitate thorough data planning activities, a data librarian must organize, treat, and preserve research data. This requires the skills necessary to work with many kinds of data technologies (Wang, 2018).

Furthermore, a data librarian must possess the necessary skills to work with any kind of data, including observational, computational, or experimental data. He/she must understand how investigations about data modify the theories and practices theories of librarianship. Besides helping researchers to find and acquire data as well as recommending proper data resources, such as research data services, data librarians may also offer consulting services related to data preparation and acquisition in specific disciplines (Dekker; Lackie, 2016). In conclusion, data librarians need to combine several skills and knowledge from various disciplines to work with various types of data. Meanwhile, their focus is on research data.

Data librarians and research data

Research data can be a primary source in the scientific discovery process. The research data are events and evidence produced, obtained, or used during scientific research, can be used and re-used for scientific analysis, and can produce research results for required to validate research results.

Research data can be characterized in various ways, according to its nature, origin, or status in the scientific research workflow. According to the *Guia de Gestão de Dados de Pesquisa* prepared by the Brazilian government's Nuclear Energy Management Commission, these data are categorized to their nature: observational: generated through direct observations that can be associated to specific places and time; computational, are the result of the execution of computational models or simulations; experimental, come from controlled situations in laboratories; can be numbers, images, videos, audio, software, algorithms, equations, animations, models and simulations; finally, according to the research flow, (a) raw data, collected directly from scientific research instruments; (b) derivatives, results of processing or combination of raw data or other data; and (c) canonical or referential, collections of consolidated data and generally archived in large data centers (Sayão; Sales, 2015).

Data is a product of research, as well as an input for research. The definition of research data is intricate and fluid and often refers to the recorded information necessary to support or validate research project observations, findings, or outputs. It comprises records, files, or other evidence, irrespective of their content or form (e.g., print, digital, or other forms) that facilitate research observations, findings, or outcomes, including primary materials and analyzed data (Rice; Southall, 2016; Henderson, 2017; Koltay, 2017).

The complexity of the research data concept reveals that data librarianship is a theory and practice needed for all professionals that aim to understand the different dimensions of research data. Data librarians need to works with researchers from the first stages of research. Data librarians should offer services to help researchers understand and document their research. These services include disclosing and evaluating research data, consulting for data management plan, metadata schemas, incorporating data in conventional collections, data literacy, reference services to help researchers find data, and generating solutions for publishing and preserving digital data (Witt; Horstmann, 2016).

Research data management is a task that must be undertaken by data librarians. Research data is an essential part of scientific academic records, and data management is an elementary task of academic libraries. Research data management is central for scientific research transparency, data preservation, and reuse and analysis of data, thereby ensuring the advancement of scientific knowledge (Tenopir *et al.*, 2017).

A data librarian must acquire these skills to support research data management. In addition, existing roles in open access promotion are also related to research data management. Librarians need to understand the theories, methodologies, and technical jargon of each scientific field to be prepared for a new demand of professional librarianship skills present in a data librarian. Furthermore, librarians need to increase their understanding of the typology and the significance of research data to researchers. Hence, the development of practical skills is highly necessary (Cox *et al.*, 2013).

There are two kinds of data services in libraries: data management services, which focus on helping researchers to work with the data (data management development plans, data organization and documentation training, safe research data storage); and data curation services, based on safe and long-term data storage and preservation. After the research is concluded, it is necessary to ensure that the data is shared and reutilized. Research data services are required by the research paradigm based on data-intensive usage and must be integrated with library services.

As research becomes more data-intensive and researchers face new challenges in managing and sharing research data, libraries have begun to offer a variety of data support services, including instruction and training, data management planning guidance, data stewardship and curation, and data visualization, research over the last few years indicates a growth in library data services (Federer, 2018, p. 294).

These services consist of the management and curation of research data. Although these activities are not clearly distinct from each other, they are independent. A data librarian needs skills in data management and data curation (Tenopir *et al.*, 2014; Koltay, 2017).

Data librarians and data management

Research data services involve skills in data management and data curation, which are independent activities, although they are not clearly distinct from each other (Koltay, 2017). Data management comprises data organization, storage, access, and preservation. This activity includes file naming, data assessment, data documentation, metadata creation and controlled vocabulary, data storage, data archiving and preservation, data sharing and re-utilization, data privacy, data copyrighting, and data publishing (Henderson, 2017).

Data management is a process that includes obtaining, validating, storing, curating and processing data to ensure continuous access and preservation over time for its users. Data management plans are important because they illustrate how data may be visible, accessed, and reused by the public. Research data services are demanded by a research paradigm that involves intensive data usage and must be incorporated into library services (Tenopir *et al.*, 2014).

Relating to the usage of research data by researchers, data librarians must possess competencies, including organizing, storing, accessing, and preserving research data. Furthermore, data librarians must have the knowledge and skills necessary to advise researchers on (a) preservation of research results; (b) data management and data curation, including discovery, access, dissemination, preservation, and portability; (c) meeting various deadlines required by funding institutions, including open access requirements; (d) potential data manipulation tools that can be used for data mining; and (e) the use of metadata (Rice; Southall, 2016; Henderson, 2017; Rice, 2019).

Additional functions suggested for libraries include training researchers to accept data archiving and preservation rules and training data librarians on how to work with any kind of data. Libraries must also be involved throughout the research route and be actively engaged in curating, advising, and preserving research outputs. The roles suggested by members of the Association of Research Libraries in the United States and Canada include training on e-science, digital curation, and data management (Tenopir *et al.*, 2014).

Data librarians and digital curation

Data librarians require technical abilities to extract, analyze, and present digital research data. Data librarianship is a field of specialization for digital curators. Thus, a data librarian must combine relevant skills for digital data curation. These abilities comprise the organization, storage, access, and preservation of digital research data (Cassella, 2016; Henderson, 2017).

The objective of digital curation is to reduce the risks of digital data obsolescence in computers or other devices. Its main function is to ensure access and continuous preservation of digital data over time. Digital data curation involves concepts such as digital preservation, data curation, electronic resource management, and digital asset management. The exponential growth of digital information has contextualized digital curation under three essential approaches: a process for selecting and filtering content on the web, an active management action focused on digital preservation, and an active process linked to e-science (Siebra; Borba; Miranda, 2016).

Some institutions in United States, Canada, the United Kingdom, the Netherlands, Italy, and New Zealand adopt an international curriculum, a Digital Curation describe digital curation as a task related to large data generation from scientific projects and projects related to cyber-infrastructure in scientific research. According to the Digital Curation Center, digital curation is concerned with the active management of the entire lifecycle of digital objects. This definition of digital curation is closely linked to the academic sphere. In this context, the term data can be used in a broad sense, which includes any digital resource (Higgins, 2008; Dutra; Macedo, 2016; Cassella, 2016).

According to About (2008), digital curation is a set of activities that encompass the management, planning, creation, and guarantee of access for the use and reuse of digital objects. Oliver and Harvey (2016) regarded digital curation as a necessary process for several reasons. For example, it is important in scientific research due to the digital focus of sciences that are largely mediated by technology. According to the same authors, the Sciences roams *in vitro* for the *in silico*. Science follows the path of intensive usage of data based on supercomputers, and an example is the Large Hadron Collider at European Organization for Nuclear Research.

Another curation trend is related to government data. Governments generate large amounts of data; hence, they need to ensure the integrity of the data and access to that data. Therefore, it is essential to develop techniques to protect data in digital media, thereby reducing the risk of loss and enabling the future use and reuse of data.

According to Dutra and Macedo (2016), the third trend is about maintaining and creating value in datasets that envision the current and future use of digital data. Data curation services provide an infrastructure that supports data management by providing permanent storage, as well as assigning unique identifiers and metadata. However, Koltay (2019) argued that research data curation services not only aim at digital preservation of data but also at the use and reuse of data. The data is prepared to be accessed and used by users, similar to a museum collection.

Hence, a data librarian must investigate the systems and rules that relate the research with research data on digital repositories; further, a data librarian must be a professional who is knowledgeable about research data collection, manipulation, analysis, and visualization to offer products and research data services related to investigations on data management and curation. A data librarian must have the necessary skills to work with any kind of data, including observational, computational, or experimental data. He/she must understand how investigations about technology can modify the theories and practices of data librarianship (Semeler; Pinto, 2019c).

Data librarians and research data repositories

The different kinds of technological systems created to support research data comprise a rich source of digital information based on scientific investigation. The diverse ways in which research data flows in these systems,

especially on the internet, accentuates the need to organize, comprehend, preserve, and analyze the information that can be extracted from the digital medium.

Research data repository is a technical information and digital organizational system that helps researchers in data management and data curation. It also eases the search for and access to research data in one or multiple sources, both internal and external to the repository. Research data repositories are an essential part of the research cyberinfrastructure that ensure the preservation, long-term access, and reutilization of data (Kindling; Pampel, 2017).

Research data repositories are an evolution of document repositories. According to Monteiro (2017), research data repositories are environments that are implemented with the appropriate infrastructure in universities to help researchers to manage and access all types of scientific data, thereby contributing to the reutilization and replication of scientific procedures. Furthermore, research data repositories help researchers in data management to plan and structure their research from the beginning of the research process. Research data repositories can be institutional, disciplinary, and multidisciplinary, and may result from specific research projects. Generally, all kinds of research data repositories emphasize the access and preservation of research data (Re3data, 2012; Semeler; Pinto, 2019c).

The global distribution of research data repositories is cataloged by the Registry of Research Data Repositories (re3data). The re3data repository was launched in 2012 and currently covers different academic subjects. It permanently stores data repositories and provides access to research datasets for funding bodies, editors, and academic institutions, thereby promoting a culture of sharing, access, and visibility for research data (Re3data, 2012).

Re3data is funded by *Deutsche Forschungsgemeinschaft* and congregates initiatives from the Library and Information Services of the Humboldt-Universität. German Research Centre for Geosciences, Karlsruhe Institute of Technology, and the Berlin School of Library and Information Science of Humboldt-Universität zu Berlin, Germany. Re3data is internationally indexed, with about 2,500 research data repositories since the fall of 2012. It offers researchers, funding organizations, libraries, and editors a systematic overview of the heterogeneous landscape of research data repositories (Costa; Braga, 2016; Kindling; Pampel, 2017).

Research data repositories allow access to a wide array of research data and ensure its preservation. Based on re3data, Costa and Braga (2016) stated that thematic research data repositories are usually hosted in North American and European countries. These repositories are stored as texts and images, with little representation of the metadata patterns and certification. Generally, most data repositories distribute their content under a copyright license, although their data are characterized as open access. Based on the context described above, there is a consolidated environment of research data repositories that focus on diverse knowledge fields and are globally distributed.

Certification on data librarianship

The development of data librarianship is related to graduate studies in librarianship and information sciences. Certification in data librarianship is interdisciplinary. In this sense, to acquire knowledge about data-oriented librarianship, there are some courses offered by international institutions. The data librarianship is ad hoc. Formal academic education of data librarians involves studies at the graduate level; we assume that a data librarian must have extensive research experience. Qualification and certification in related areas, on the other hand, can be acquired through courses that teach techniques of data technology application, data management, and digital data curation.

On data management and curation, we suggest the Research Data Management and Sharing course, offered online for free by the North Carolina University, at Chapel Hill, in partnership with division of Information services of the Edinburgh University. The focus of Research Data Management and Sharing course courses is on creating educational resources to empower librarians, archivists, library students in data management and curation.

Another course is Research Data Management Training, offered by the division of Information services of the Edinburgh University, focused on data management and research data repositories. The course presents theoretical and practical definitions of research data (organization, documentation, storage, preservation, and security), data management, data sharing, data life cycle, and best practices related to research data (Molloy, 2012).

The courses offered by the Library Carpentry, have the objective of offering data science techniques and skills that can be applied for the creation of services in Libraries. The Library Carpentry website is made up of software and data skills training; website course aims to help librarians work and potentially teach the abilities they have learned to colleagues, students, and researchers.

Additional course is Research Data Management Librarian Academy, a program to enable data librarians to work in a global research-intensive. Finally, LIS 628 – Data Librarianship and Management by Steeves (2020) at the Pratt Institute School of Information in the United States is an open academic class focused on the introduction of concepts and terminology related to data and data services in libraries. Lastly, we will conclude this work by showing the main characteristics of the data librarianship as a field.

Conclusion

The usage of data technology in libraries defines the main characteristic of data librarianship as a field. Applications linked to data management technologies and data curation are leading trends in data librarianship. However, data librarians must use the ethical principles and professional skills of traditional librarianship to work with research data. Data librarianship is related to harvesting, manipulation, analysis, and visualization of data.

Data librarianship focused on studies of data provenance and best practices for creating reference services using research data, among other scientific activities related to data theory and praxis, primarily when these data are related to scientific research processes, and in digital media such as data repositories.

This paper recommends that data librarians work with research data repositories to support researchers from the first stages of scientific investigation, assist with the documentation process, and ensure that the research data is preserved, usable, and reusable for a long time. We concluded that data librarianship is a practical field that incorporates the concept of digital data in libraries. Therefore, the field of data librarianship sees itself as a field strongly related to the practice, use, and consumption of data technologies. Data librarians should participate in discussions on data sharing, access tools, preservation best practices, and global resources. They should also answer questions related to statistical literacy, data management, data curation, data sharing challenges, metadata formats, digital repositories, data preservation, and citation. It is a new approach to the practice of librarianship based on contemporary data technology. Data librarianship as a field plays a fundamental role in the data management and curation of datasets generated during research processes.

Finally, we point the limitation of the research regionalities of data librarianship is that our next work aims to delimit the regionalities of data librarianship. We will conduct a bibliometric study to compare the trends and differences in data librarianship in countries like the United States, Canada, the United Kingdom, and Brazil. For further studies, it is suggested that data librarians investigate themes related to research data usage in diverse contexts such as big data, the Internet of Things, artificial intelligence, information patterns, and scientific communication.

Contributors

A. R. SEMELER contributed to the study conception, design and interpretation. A. L. PINTO contributed to the review and approval of the final version of the manuscript.

References

- About, D. *What is digital curation?* Edinburgh: Digital Curation Centre, 2008. Available from: <https://era.ed.ac.uk/handle/1842/3362>. Access on: Mar. 1, 2020.
- Bawden, D.; Robinson, L. *Introduction to information science*. London: Facet Publishing, 2012.
- Borko, H. Information science: what is it? *American Documentation*, v. 19, n. 1, p. 3-5, 1968. Doi: <https://doi.org/10.1002/asi.5090190103>.
- Buckland, M. Information as thing. *Journal of the American Society for Information Science*, v. 42, n. 5, p. 351-360, 1991. Available from: <http://people.ischool.berkeley.edu/~buckland/thing.html>. Access on: Mar. 1, 2020.
- Cassella, M. Dal digital curator al data librarian. *Biblioteche Oggi*, v. 34, n. 3, p. 13-21, 2016. Doi: <http://dx.doi.org/10.3302/0392-8586-201603-013-1>.
- Chawinga, W.; Zinn, S. Research data management in universities: a comparative study from the perspectives of librarians and management. *International Information and Library Review*, p. 1-15, 2020. Doi: 10.1080/10572317.2020.1793448.
- Costa, M.; Braga, T. Repositórios de dados de pesquisa no mundo. *Cadernos Bad*, n. 2, p. 3-5, 2016. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1585>. Acesso em: 1 mar. 2020.
- Cox, A.; Verbaan, E.; Sen, B. Upskilling liaison librarians for research data management. *Ariadne*, n. 70, 2013. Available from: <http://www.ariadne.ac.uk/issue70/cox-et-al>. Access on: Mar. 1, 2020.
- Dekker, H.; Lackie, P. Technical data skills for reproducible research. In: Kellam, L.; Thompson, K. *Introduction to databrarianship: the academic data librarian in theory and practice*. Chicago: Association of College and Research Library, 2016.
- Dutra, M.; Macedo, D. Curadoria digital: proposta de um modelo para curadoria digital em ambientes big data baseado numa abordagem semi-automática para a seleção de objetos digitais. *Informação e Informação*, v. 21, n. 2, p. 143-169, 2016. Doi: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p143>.
- Federer, L. Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association*, v. 106, n. 3, p. 294-303, 2018. Doi: <https://doi.org/10.5195/jmla.2018.306>.
- Henderson, M. *Data management: a practical guide for librarians*. Lanham: Rowman and Littlefield Publishers, 2017.
- Higgins, S. The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, v. 3, n. 1, p. 134-139, 2008. Doi: <https://doi.org/10.2218/ijdc.v3i1.48>.
- International Association for Social Science Information Services and Technology. *Defining data librarian: call for comments*. Edmonton: International Association for Social Science Information Services and Technology, 2016. Available from: <http://www.iassistdata.org/blog/defining-data-librarian-call-comments>. Access on: Mar. 1, 2020.
- Kellam, L.; Thompson, K. *Introduction to databrarianship: the academic data librarian in theory and practice*. Chicago: Association of College and Research Library, 2016.
- Kennan, M. *Data management: knowledge and skills required in research, scientific and technical organizations*. Haia: Ifla, 2016. Available from: <http://library.ifla.org/1466/1/221-kennan-en.pdf>. Access on: Mar. 1, 2020.
- Kindling, M.; Pampel, H. Informationsinfrastrukturangebote für digitale Forschungsdaten. *E(hren) Journal*, n. 2, p. 15-33, 2017. Doi: <http://doi.org/10.18452/2341>.
- Koltay, T. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, v. 49, n. 1, 2017. Doi: <https://doi.org/10.1177/0961000615616450>.
- Koltay, T. Accepted and emerging roles of academic libraries in supporting research 2.0. *The Journal of Academic Librarianship*, v. 45, n. 2, p. 75-80, 2019. Doi: <https://doi.org/10.1016/j.acalib.2019.01.001>.
- Koltay, T. Quality of open research data: values, convergences and governance. *Information*, v. 11 n. 4, p. 1-15, 2020. Doi: <https://doi.org/10.3390/info11040175>.
- Mccaffrey, M.; Giesbrecht, W. Teaching data librarianship to LIS students. In: Kellam, L.; Thompson, K. *Introduction to databrarianship: the academic data librarian in theory and practice*. Chicago: Association of College and Research Library, 2016.
- Monteiro, E. *Direitos autorais nos repositórios de dados científicos: análise sobre os planos de gerenciamento dos dados*. 2017. 118 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2017.
- Morville, P. *Ambient findability: what we find changes who we become*. Farnham: O'Reilly, 2014.
- Molloy, L. *JISC Research data MANTRA project at EDINA, information services*. Edinburgh: University of Glasgow, 2012. Available from: eprints.gla.ac.uk/71435/1/71435.pdf. Access on: May 1, 2020.
- Re3data. *About*. Berlin: Humboldt-Universität, 2012. Available from: <http://www.re3data.org/about>. Access on: Mar. 1, 2020.
- Rice, R.; Southall, S. *The data librarian's handbook*. London: Facet Publishing, 2016.
- Rice, R. Supporting research data management and open science in academic libraries: a data librarian's view. *Communications of the Association of Austrian Librarians*, v. 72, n. 2, p. 263-273, 2019. Doi: <https://doi.org/10.31263/voebm.v72i2.3303>.
- Saracevic, T. Information science. In: Bates, M. J.; Maack, M. N. (ed.). *Encyclopedia of library and information science*. New York: Taylor and Francis, 2010. p. 2570-2586. Available from: <https://comminfo.rutgers.edu/~tefko/SaracevicInformationScienceELIS2009.pdf>. Access on: May 1, 2020.
- Oliver, G.; Harvey, R. *Digital curation*. Chicago: American Library Association, 2016.
- Sayão, I. F.; Sales, L. F. *Guia de gestão de dados de pesquisa para bibliotecários de pesquisadores*. Rio de Janeiro: Cnen, 2015.

Disponível em: http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 1 maio 2020.

Semeler, A.; Pinto, A.; Vianna, W. E-science: an epistemological analysis based on the philosophy of technology. *Ifla Journal*, v. 43, n. 2, p. 198-209, 2017. Doi: <https://doi.org/10.1177/0340035216678235>.

Semeler, A.; Pinto, A.; Rozados, H. Data science in data librarianship: core competencies of a data librarian. *Journal of Librarianship and Information Science*, v. 51, n. 3, p. 771-780, 2019a. Doi: <https://doi.org/10.1177/0961000617742465>.

Semeler A.; Pinto, A. *Data librarianship Venn diagram handbook (beta version)*. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2019b.

Semeler A.; Pinto, A. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, v. 48, n. 1, 2019c. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 1 maio 2020.

Semeler A. R.; Pinto A. L. Librarianship in the age of data science: data librarianship Venn Diagram. In: Mugnaini, R. (ed.). *Data and information in online environments*: DIONE 2020. Cham: Springer, 2020. Doi: https://doi.org/10.1007/978-3-030-50072-6_10.

Siebra, S.; Borba, V.; Miranda, M. Curadoria digital: um termo interdisciplinar. In: Encontro Nacional de Pesquisa em Ciência

da Informação, 27., 2016, Salvador. *Anais* [...]. Salvador: Universidade Federal da Bahia, 2016.

Steeves, V. *LIS 628: data librarianship and management*. New York: Pratt Institute School of Information, 2020. Available from: <https://vickysteeves.gitlab.io/lis-628-datalibrarianship/syllabus.html#instructor-information>. Access on: Mar. 1, 2020.

Tenopir, C. et al. Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research*, v. 36, p. 84-90, 2014. Doi: <https://doi.org/10.1016/j.lisr.2013.11.003>.

Tenopir, C. et al. S. Research data services in european academic research libraries. *Liber Quarterly*, v. 27, n. 1, p. 23-44, 2017. Doi: <https://doi.org/10.18352/lq.10180>.

Xia, J.; Wang, M. Competencies and responsibilities of social science data librarians: an analysis of job descriptions. *College and Research Libraries*, v. 75, n. 3. p. 362-388, 2014. Doi: <https://doi.org/10.5860/crl13-435>.

Wang, L. Twinning data science with information science in schools of library and information science. *Journal of Documentation*, v. 74, n. 6. p. 1246-1257, 2018. Doi: <https://doi.org/10.1108/JD-02-2018-0036>.

Witt, M.; Horstmann, W. International approaches to research data services in libraries. *Ifla Journal*, v. 42, n. 4, p. 1-2, 2016. Available from: https://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-42-4_2016.pdf. Access on: Mar. 1, 2020.