Librarianship in the Age of Data Science: Data Librarianship Venn Diagram

Alexandre Ribas Semeler[1](✉) and Adilson Luiz Pinto[2]

[1] Geosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
alexandre.semeler@ufrgs.br
[2] PGCIN, Federal University of Santa Catarina, Florianópolis, Brazil
adilson.pinto@ufsc.br

Abstract. A Venn diagram is used in mathematics to graphically symbolise properties, axioms, and problems concerning sets and their theories. Thus, this study applied a Venn diagram to describe a theoretical background for data librarianship as a field relating to information science, e-science, and data science. Data librarianship is a new area of study that is located within the thematic core of the triad. The first set on the proposed Venn diagram is information science. Information technology concepts are fundamental to the comprehension of data librarianship in the context of information science. The second set is e-science, an innovative field that incorporates software and hardware that have been built by technology into science. The third set is data science, a way of representing data-driven research in the most diverse knowledge fields; it is a set of the skills, methods, techniques, and technologies of statistics and computer science used to extract knowledge and to create new products and services from data. To ensure greater comprehension of data librarianship, a relatively new field, we suggest some reading materials. The formal discipline of data librarianship is yet to be established in many countries across the globe. Thus, there is the lack of adequate information and certification on data librarianship.

Keywords: Data librarianship · Information science · E-science · Data science

1 Introduction

A Venn diagram is often used in mathematics to graphically symbolise properties, axioms, and problems concerning sets and their theory. The intention of this study is to apply a Venn diagram to describe a background for data librarianship as a practice field in relation to others fields as information science, e-science, and data science.

We propose that data librarianship in nowadays is a practice field located within the thematic core of the triad (information science, e-science, and data science). In this sense, the data landscape manifests as an area of growing research and a new practice that is being adopted by data librarians and other information scientists. Most of the investigations on e-science and data science are oriented towards the practice, usage, and consumption of data. In this sense, this paper proposes a Venn diagram

R. Mugnaini (Ed.): DIONE 2020, LNICST 319, pp. 118–130, 2020.
https://doi.org/10.1007/978-3-030-50072-6_10

to describe data librarianship, to provide a theoretical background for data librarians. Data librarianship is a field that will incorporate information science, e-science, and data science in libraries. Therefore, it is possible to say that the current tendencies in data librarianship research points to subjects related to […] intensive use of data leads to a technological data information environment that is ideal for studies aimed at models, techniques, and technologies used for data management and curation. This data environment must have competencies related to the creation, management, and preservation of data [1].

Information science is seen as a field of knowledge that involves a series of disciplines, such as computer science and librarianship. In this sense, the object of study of information science is information and the changes caused by the use and re-use of digital technology. Data in digital form are a new form of information generated by all human activities. We believe in a technological continuum that relates with the emergence of digital data in information science and modifies this discipline so that it can investigate the properties of information in all digital contexts in relation to all human activities.

Data librarianship in contexts of data-intensive sciences is related with e-science and data science process. E-science is the term adopted to refer to the use of computational technology during the practice of scientific investigation, which includes preparation, experimenting, data collection, results, storage, digital preservation, and accessibility for all the materials generated during scientific research. It can be said that e-science is preoccupied with the processes of representing information and scientific knowledge through technology that provides data-intensive usage [2].

The prioritization of data technology by scientists is a transformation variable that is common to almost all theories and scientific procedures. Currently, data science is regarded as the theory and practice of extracting knowledge from data; it is concerned with the creation of products and services based on data [3].

Thus, we propose these disciplines as the theoretical pillars for the investigation of data librarianship as a new area of study for librarians. Ultimately, our proposition is a Venn diagram for discussing the theoretical background of data librarianship.

2 Data Librarianship Venn Diagram

We believe that data librarianship is a central filed suitable for practical application to information science, e-science and data science in libraries. The field of data librarianship was made essential in the field of information by the influence of technologies and the direct impact of data-intensive science on the activities of data librarians and researchers, demanding that those professionals acquire skills for usage of all type of data. Thus, there is the question of how to delimit the relationship between information science, e-science, and data science and fashion a new practical and theoretical approach to data librarianship. To answer this question, which depicts the general objective of this paper, we propose a Venn diagram to present and summarise our proposal relating to the theoretical background of data librarianship (see Fig. 1).

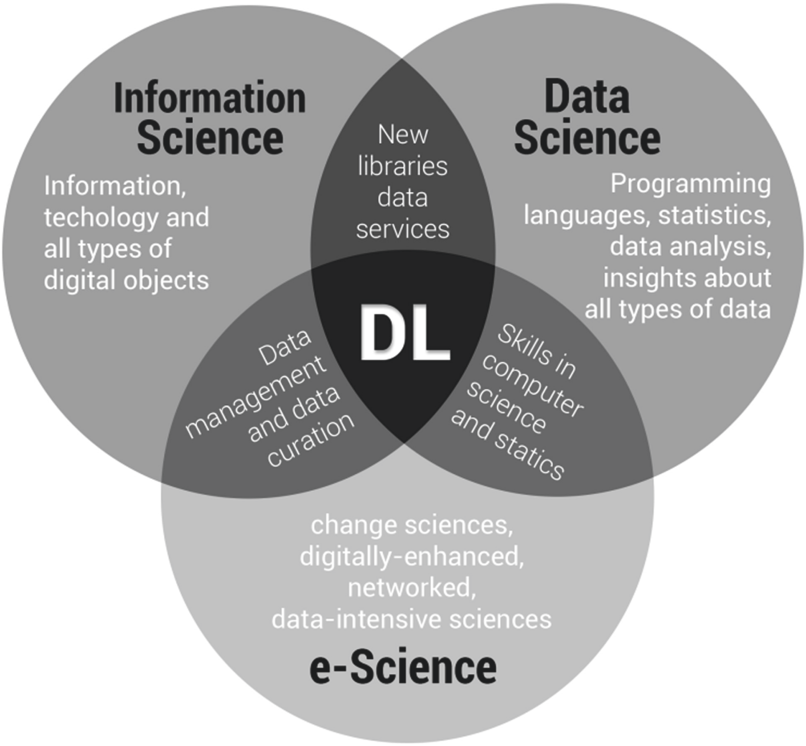

Fig. 1. Data librarianship Venn diagram [4].

The data librarianship Venn diagram (Fig. 1) describes the union and intersection between information science, e-science and data science. With this diagram, it is possible to conceptualise a background of data librarianship. Using three theoretical sets, it is possible to visualise the nucleus of data librarianship as a field: first set (information science) = {information, technology and all types of digital objects}; second set (e-science) = {change sciences, digitally-enhanced, networked, data-intensive sciences}; third set (data Science) = {programming languages, statistics, data analysis, insights into all types of data}.

Next, we present the core of the Venn diagram of data librarianship and the three theoretical sets related to it.

3 The Core of Data Librarianship Venn Diagram

Data librarianship is a area of study that is located within the thematic core of the triad fields, information science, e-science, and data science; its pillars are the new practical and theoretical activities related to data management and curation in digital repositories in the libraries context. The practical approach to data librarianship is geared toward activities such as the collection, manipulation, analysis, and visualisation of data, which are fundamental for data librarians toward offering new data services and products in libraries and other places of information.

Research into data librarianship has its origins in the mid 2000s. They are the result of the effort of North-American, British, and Canadian librarians to create services and products on advertising, consulting, management, preservation, and elaboration of metadata schemas for the effective incorporation of research data in document collections (books, papers, reports, and others). This effort is not something new; it has been discussed since the first applications of data librarianship in the 1960s, particularly in the creation of data services and data archives [1, 4–9].

Data-driven librarianship is not a new branch of librarianship; it is grounded in a diversity of skills already practised by librarians. Thus, traditional practices are reinforced with the objective of being applied to digital datasets. For instance, cataloging and organising all types of bibliographic materials, preservation and curation, user services and reference services, and consultancy and training, are redesigned to be combined with the new practices that encompass data management, data curation and data sharing. However, data librarianship is not only suited as support for the discovery of new skills in librarianship; it is involved with research data generation and preservation, and is concerned with practically all the traditional functions of the librarian trade such as acquisition and collection development, and collection and organisation of documents, cataloging, and user reference services implementation. Currently, the focus of data librarianship is a library services and new approaches for managing and curating digital data from scientific research [1, 4–9].

One of the main characteristics of data librarianship is a core preoccupation with upstream duties [10]; therefore, it cannot only be concerned about published information, but also with potential data sources, being that its objective is to comprehend how the different data types may be capable of generating useful research information [6, 9].

We propose that academic research on data librarianship has focused on fields such as information science, e-science, and data science. Thus, the next section presents the origins of information science and its unfolding in relation to the concepts of information and technology are presented the first set of Venn diagram.

4 Set 1: Information Science

The first set on the data librarianship Venn diagram is information science. Information science is a field of study where interdisciplinary issues circulate. It considers all forms of interaction between people and information [2]. According to [11], information science is a multidisciplinary field of study, involving several forms of knowledge, given coherence by its focus on the central concept of human-recorded information [2].

The concept of information and technology are fundamental to the comprehension of information science in digital context. Thus, [12] propose that information concept be applied in conjunction with documentation fields such as librarianship, among others aspects applied to technological information systems. The author further explained that author's information is a structured set of codified mental and emotional representations (signs and symbols), modeled with/by social interaction, and capable of being recorded on any material medium, and therefore, communicated in an asynchronous and multidirectional way [12].

According to [13–15] investigations about information in information science suggests that the implementation of critical investigations about the conceptual nature of the basic principles of registered information, including the dynamic utilisation of information in the sciences and the application of methodologies and theories to solve problems generated by computer usage.

Information science looks into theoretical and practical issues related to the behavioural characteristics and flows of information when we incorporate technology-mediated hybrid media such as e-science. The ubiquitous technology driving the pervasive behaviour of digital information has changed our existence that has been increasingly mediated by technology. This is not limited to the sciences but encompasses all domains of knowledge. Under this premise, the concept of technology is introduced [2].

When we say technology is a significant part of the contemporary world, we may be referring to our satisfaction by devices that make our lives more comfortable, our enthusiasm in the face of the possibilities guaranteed by computers and internet, or our fear of the increasingly more potent and sophisticated weapons it heralds. Thus, technology presents itself as a multifaceted reality, not only as objects and sets of objects but also as systems, processes, modus operandi, and mentality [16–18].

The most common way to think technology is to regard it as a technical device, a computer or smartphone, hardware, which is a manner of seeing it as something concrete. In the words of [16], technology manifests itself primarily as objects, as material artefacts made by man, whose function depends on a specific materiality. Under this bias, it is necessary to reiterate that technical devices are, in part, user dependent and using and/or handling them require special skills and training [16–18].

Nowadays, technology can be defined as hardware, as rules, and as systems. First, technology is usually thought of as being comprised of machines or tools—hardware—this is a way of looking at it as something concrete. However, a distinction between these concepts should be made. While users directly manipulate tools, machines are more independent of the user and require skills and training to be used or handled. A second standpoint views technology as rules. Software can be seen as a metaphor for this approach; it involves patterns of meaning and relationships, such as political rules, laws, and scientific principles that are systematically developed. The third way of viewing technology is as systems. This suggests that hardware artefacts and software are technologies, but that they need to be considered in the context of their user [1, 2].

Technology can be a specific form of knowledge, comprising of specific ways of knowing the material world, ways that incorporate scientific knowledge, but also equally possess their own characteristics. In this sense, technology defines itself as a multifaceted, artifact manufacture and usage; a form of human knowledge addressed to create a reality according to our own purposes; know-how; practical implementations of intelligence; humanity at work; nature available to man as a resource; the field of knowledge related to artifact design and the planning of its achievement, operation, adjustment, maintenance, monitoring, under the light of scientific knowledge; Modernity own lifestyle; the totality of methods that can be reached rationally; in all fields of human activity; the material structure of Modernity, in four dimensions: as objects; way of knowledge; specific form of activity or will (determinate human action when facing reality) [16–18].

This examination aims to discuss information and technology to examine the impact of technology on scientific knowledge and information concepts, i.e. research activities, the discourse about scientific culture, and its influence on the processes of interaction between information and technology. When information is mediated by technology, it can reveal the complexity of new ways of scientific investigation [2]. Thus, the next section is a presentation on e-science, the second set of data librarianship Venn diagram.

5 Set 2: E-science as a Contemporary Data and Information Phenomenon

The investigation methods based on the use of hardware infrastructure and scientific software, have allowed the emergence of e-science. In the 1990s, John Taylor, the Director General of the United Kingdom's Office of Science and Technology, created this term to refer to the use of technology to conduct scientific investigations. In United States a 2007 lecture, Jim Gray described the e-science as a new way of doing science. He believes that e-science is the fourth paradigm of science [2, 19–21].

E-science is one of many terms used to describe recent transformations in the scientific enterprise. It is a way to translate methods and scientific disciplines to computers, allowing scientific collaboration on a global scale, in addition to intensive data usage [21].

Thus, it is perceived that, beyond the concern with immense data volume, generally, e-science basic characteristics are being digitally-enhanced, networked, and data-driven. The digitally-enhanced aspect of e-science embodies how e-scientists analyse, manage, gain access to, and share data digitally. Without the technological advances of our time, the amount of data we have access to and are able to create would significantly decrease. Almost exclusively through online databases, networks, shared digital repositories, and digital files, datasets are shared, and therefore, are digitally accessible. Networking is a very important feature of e-science because this is what makes e-science particularly fascinating, the networked and multidisciplinary nature of the field that creates initially unintended links among datasets. E-scientists utilise networked data and materials to formulate new information through cross-comparison and manipulation. This creates another form of scholarly communication that is networked and constructed extemporaneously, depending on the current usage; such as how articles are connected through citations and websites. Therefore, e-science thrives when datasets are shared and accessible. The final characteristic of e-science is that it is data-driven. E-science exploration is achieved through data manipulation; thus, datasets are used as the primary form of experimentation. By manipulating and cross-comparing datasets, researchers are able to find patterns and develop new inquiries across disciplines [22].

Current studies about data-driven sciences amplify themselves when mediated by computers, mainly if networks like the web are included. This gives rise to different methods applicable to data collection and/or usage; and thus, a new way to practise science, based on technological instruments and technical devices created by contemporary computing that has as its basis the data-intensive usage of the world of science. Data-driven sciences opened a new dimension to scientists, the data universe, causing a revolution in the process of scientific thinking.

The e-science phenomenon is responsible for the generation and use of data in large quantities. In the Big Data era, data, information and technology are pervasive. E-science is a contemporary data/information phenomenon. E-scientists believe that data-driven sciences will yield optimum results with the intensive use of technological instruments and devices. The theoretical understanding of these studies is based on appreciating the fact that the nature of scientific things depends on the materiality of technical devices. Therefore, scientific instruments are the key to contemporary scientific practice. This conclusion makes science less about knowledge than about a practice that involves using machines and technical devices [2, 23].

The author [20] introduced the idea of a fourth paradigm in a lecture delivered in California in 2007 for the National Research Council of the United States. He considers computational science as the third type of Science, aiding the various domains of knowledge, for example, the hard sciences such as Astronomy, Physics, and Medicine, amongst other categories of science. He amplifies the idea of e-science, thus, making the term synonymous to the fourth paradigm of science. It is believed that digital technology revolutionised the scientific methods, as posited by computer sciences made new scientific paradigms emerge which modified the way scientific research is.

Investigations in the fields of e-science were simulated against the backdrop of hard science disciplines such as astrophysics, genetics, oceanography, climatology, nuclear physics because the main concern of these fields is the immense volume of data generated during scientific research. E-science channels science transformation towards the intensive usage of huge data volumes.

E-science methods were first used by Biology, Physics and Astronomy researchers in projects such as the GenBank [24], the Large Hadron Collider (LHC), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) [25] and other projects in the world of hard sciences. For example, the World Wide Web, invented by Tim Berners-Lee at the European Organization for Nuclear Research (CERN), was createdfor sharing document and data amongst researchers. For example, the Pan-STARRS project will retain 2.5 petabytes of data each year, when it starts operation [31]. CERN's Great Hadrons Collider (GHC) will generate 50 to 100 PB of data each year, with circa 20 PB stored and processed in a global federation of national arrays linking 100 thousand CPUs. Therefore, scientists and institutions need a model and better set of practices to provide balanced hardware architecture to the corresponding software capable of dealing with those data volumes [21, 24, 25].

The large data volume generated by e-science gives rise to data science. Data science is regarded as a field of study connected to data analysis and management. In the next section, we present data science as a scientific theory and practice used for the comprehension of the diverse dimensions of data-intensive sciences.

6 Set 3: Data Science

Data science is a field of knowledge that merges techniques of computer science and statistics. It can be considered as a new area of study for librarians seeking to engage with issues related to data management and analysis. It is a field of knowledge that demands specific expertise, usually related to mathematics and programming languages [1].

The term data science has its origin in disciplines such as Econometrics, Physics, Bio-statistics, computer science, and statistics and engineering. It can be a synonym or related term for concepts such as business analytics, operational research, competitive intelligence, data modelling and analysis, and the discovery of knowledge at databases. The time-line (see Fig. 2) presents data science evolution.

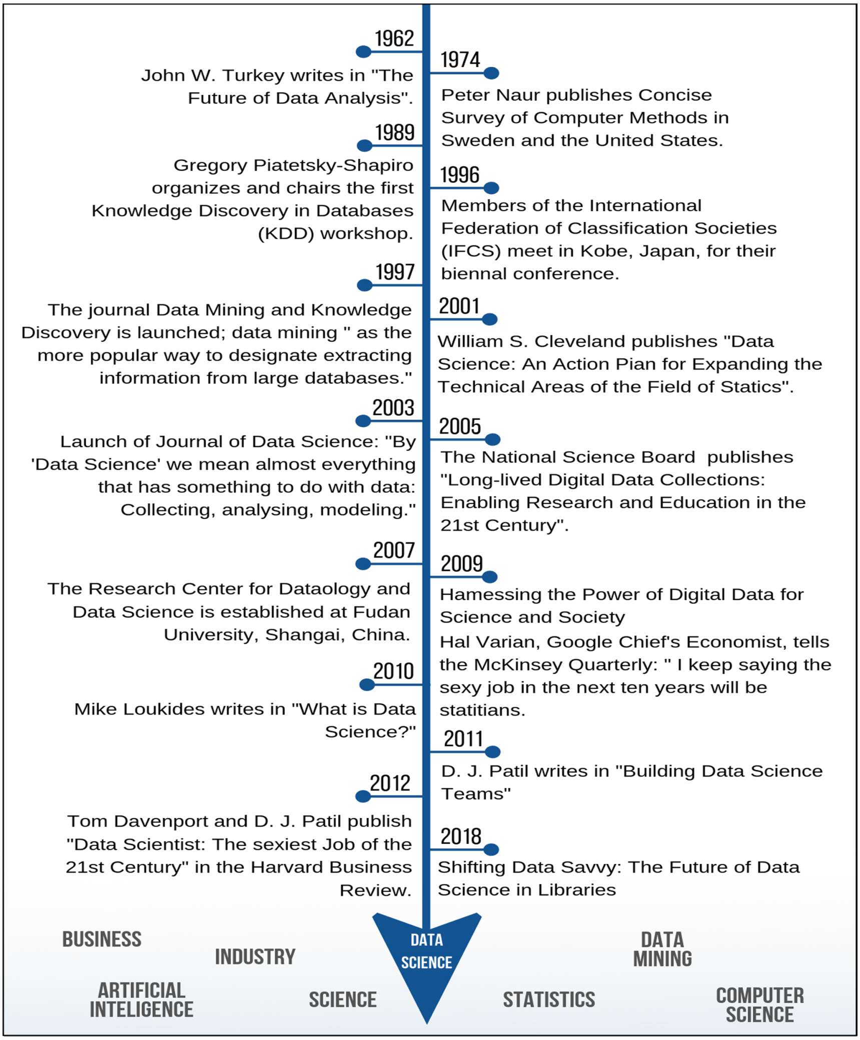

Fig. 2. Data Science time-line [4].

Although data science is often viewed as a new term, it has been used in the field of computer science since the 1960s. [27, 28] reveal that [29] proposed that computer science be called “Datalogy”. In the 1990s, the Committee on Data of the International Council for Science (CODATA) established as priority the realisation of studies about advances in data science, CODATA using the term data science as a way of representing data-driven research in the most diverse knowledge fields. In the 2000s, the first journals appeared, such as the Journal of Data Science from the Columbia University in the USA. Finally, data science appears, in current usage, as a combination of data analysis development and the elaboration of new intelligent algorithms in artificial intelligence, on a par with statistics and computer science [26–29].

Data science possesses two basic functions: the analysis and invention of new techniques that may generate insights during the usage of data, which was not possible before the 2000s. In that sense, data science can be seen as the theory and practice of extracting knowledge from data. Being a scientific discipline, which is preoccupied with creating products and services from data, its aims are to use data-driven sciences for the benefit of the environment, such as the use of Internet products as data repositories, climatic sensors, GPS, and the most recent tendency to connect objects to network [27, 28].

However, one of the main investigation problems in data science is geared towards establishing it as a science of data and not a mere extension of statistics and computer science techniques. Data science aims to analyse and understand contemporary phenomena using data. [30]. The data science is as a new area of study for information professionals seeking to engage in data management and data analysis issues based on statistics and computer science [27, 28].

In a recent paper, [31] presented a definition of data science and the crucial role of librarians in its development. Advances in statistics and computer science, combined with an abundance of data, have given rise to a new professional ecosystem. This ecosystem is concerned with generating insight from data to inform decision-making. Data science methods and products have transformed commerce, health care, and government, and they will continue to transform other sectors [1].

Data science requires knowledge and techniques to data analysis, including statistics and computer programming, requires systematic thinking combined with a creative approach to solve general pragmatic problems. Data science combines data mining, machine learning, system design, distributed computing, statistics, industrial engineering, domain knowledge, visualisation, and Big Data [32, 33].

The scientific status of data science is connected to different objectives such as: (a) the study of scientific data, which sees data science as a method and technology used to conduct scientific researches through scientific data management; it refers to the data used during scientific researches; (b) the study of business data, which sees data science as a technique for creating data products or services, and as a methodology to generate solutions for problems in a business environment (business intelligence); (c) the integration of statistics, computational technology, and artificial intelligence; this objective is preoccupied with the requisite skills for the data scientist; and (d) a method for solution of scientific and/or business issues related to the extraction of knowledge from data; that is, an objective which integrates all other objectives [1, 32, 33].

Therefore, following the identification of the combination of disciplines and the practical development process that compose data science, it becomes imperative to identify the skills needed by data scientists. Data science can be regarded as a new area of study for data librarians seeking to become involved in issues related to data management and analysis. Thus, data librarians should understand the complexity and variation associated with data science praxis because it provides new methods and practices for data librarianship. A data librarian needs not become a programmer, statistician, or database manager, but should possess foundational competencies in the languages and programming logic of computers, databases, and information retrieval tools [1].

Finally, we will suggest some readings, courses and suggested themes for investigation, in relation to the triad, information science, e-science, and data science.

7 Final Remarks: Literature, Courses, and Future Work in Data Librarianship

The final question is related to the difficult localisation of literature about data librarianship. It reveals that data librarianship is still developing and needs attention from the professionals of traditional librarianship. In that sense, in order to acquire minimum theoretical knowledge in data librarianship, few essential works are recommended. The work, organised by Lynda Kellam and Kristi Thompson in 2016, is the anthology of articles, *Databrarianship: the academic data librarian in theory and practice*, which presents diverse questions about data librarianship and research data usage based on the most distinct scientific disciplines [8]. Furthermore, another significant work The book *Data Librarian's Handbook*, written in 2016 by Robin Rice and John Southall; it focuses on the data librarian's field of action [5, 34, 35].

Examining the questions of e-science, the book writings by Tony Hey; Stewart *T*ansley and Kristin Tolle, *Fourth Paradigm: Data-intensive Scientific Discovery* (2009) is another significant reading material. The importance of this text lies in the fact that it sees e-science in all its aspects, in relation to scientific and technological advances, scientific communication or the intensive use of data extracted from technological sensors [35].

The questions related to data science and the requisite skills for the data scientist appear in first-time in the text, *Data Scientist: The Sexiest Job of the 21st Century* published by the Harvard Business Review, in 2012, written by Thomas Davenport and D. J. Patil [36]. Another essential text on data science is *The Data Science Venn diagram*, elaborated by Drew Conway in 2010 available free in the web. The report about data science in libraries was published in 2018, *Shifting to data savvy: The future of data science in libraries*; it presents the synthesis of the discussions on data science as applied in North American libraries. The report represents data science projects in libraries in the USA. The project explores the challenges associated with the implementation of data science in different libraries environments. The report is oriented around four distinct facets of a multi-faceted framework: structures, skills, services, and stakeholders. The report discusses specific perspectives in relation to the lack of skills militating against the practice of data librarianship, in other words, what is missing in librarians' information management skills; and also in relation to data management, in other words, the capability of library managers to understand and appreciate the benefits of the new skills brought by data science to offer organizational and managerial support to libraries [37].

In this sense, to acquire skills, there are some courses and formation courses offered by international universities, whether in the form of proximate or distance learning. For example [38], we have the Research Data Management and Sharing course, offered online for free by the Curating Research Assets and Data Using Lifecycle Education (CRADLE), of the North Carolina University, at Chapel Hill, in partnership with project EDINA, of the Edinburgh University. An online course is the Research Data Management Training MANTRA [39], offered by project EDINA, of the Edinburgh University, focused on research data management. The course is useful for both data librarians and other researchers. These courses offer an introduction to research data management and sharing, life cycle, data management, and best practices in relation to working with data (including organisation, documenting, storage, and security), and research data repositories.

The courses offered by the Library Carpentry initiative website is made up of software and data skills training originally based on lessons developed by the Carpentries. The aim of this website is help library professionals work more efficiently, and potentially teach the skills they have learned to colleagues, students, and researchers. Another course is offered by the Harvard-Smithsonian University, at the Center for Astrophysics, John G. Wolbach, and by the Harvard Library focused on preparing data librarians for acquiring knowledge about data science [40].

To learn statistic techniques and programming languages used in data science, data librarians might opt for courses offered by universities in the field of statistics or computer science. They might opt for completely different formation courses offered by online training platforms such as Coursera, Udemy, Data Science Academy, and EDX.

For further studies, it is suggested that data librarians investigate themes related to research data usage in diverse contexts such as Big Data, Internet of Things, artificial intelligence, informational patterns, and scientific communication. It would be interesting to accomplish a study about the possible analysis of the metadata of these fields and the content stored at research data repositories. Thus, in conclusion, this paper recommends that data librarians should work with research data repositories, supporting researchers, from the first steps of scientific investigations, assisting with the documentation process and providing the assurance that the research data will be preserved, usable, and reusable in the long term. A data librarian must use the fundamental values, ethical principles, skills, and professional knowledge of data librarianship to work with research data, especially at digital repositories.

Finally, data librarians should be able to combine the skills and knowledge related to information science, e-science, and data science, to consolidate the Venn diagram of data librarianship. This diagram was proposed to enable for data librarians to consolidate their practical activities supported based on consolidated theoretical fields such as information science, in addition to new knowledge in fields such as e-science and data science.

Funding. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Semeler, A., Pinto, A., Rozados, H.: Data science in data librarianship: core competencies of a data librarian. J. Librariansh. Inf. Sci. **51**(3), 771–780 (2019)
2. Semeler, A., Pinto, A.: E-science: an epistemological analysis based on the philosophy of technology. IFLA J. **43**(2), 198–209 (2017)

3. Nielsen, L., Bulingame, N.: A Simple Introduction to Data Science. New Street Communications, Wickford (2012)
4. Semeler, A., Pinto, A.: Data Librarianship Venn Diagram Handbook (Beta Version). IGEO/UFRGS, Porto Alegre (2019)
5. Rice, R., Southall, S.: The Data Librarian's Handbook. Facet Publishing, London (2016)
6. Koltay, T.: Data literacy for researchers and data librarians. J. Librariansh. Inf. Sci. **49**(1), 3–14 (2017)
7. Tenopir, C., Talja, S., Horstmann, W.: Research data services in European academic research libraries. LIBER Q. 2017 **27**(1), 23–44 (2015)
8. Kellam, L., Thompson, K.: Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice. Association of College and Research Library, Chicago (2016)
9. Xia, J., Wang, M.: Competencies and responsibilities of social science data librarians: an analysis of job descriptions. Coll. Res. Libr. 2014 **75**(3), 362–388 (2015)
10. Gold, A.: Cyberinfrastructure, data, and libraries, part 1: a cyberinfrastructure primer for librarians. D-Lib Magazine, 13(9/10) (2007)
11. Bates, M.: The invisible substrate of information science. J. Am. Soc. Inf. Sci. **50**(12), 1043–1051 (1999)
12. Silva, A.M., Ribeiro, F.: Information science and philosophy of information: approaches and differences. In: Demir, H. (ed.) Luciano Floridi's Philosophy of Technology: Critical Reflections. Springer Science+Business Media, Dordrecht (2012)
13. Floridi, L.: Open problems in the philosophy of information. Metaphilosophy **35**(4), 554–582 (2004)
14. Floridi, L.: The Philosophy of Information. Oxford University Press, Oxford (2010)
15. Floridi, L.: Steps forward in the philosophy of information. Etica Politica/Ethics Politics **14**(1), 304–310 (2012)
16. Cupani A.: Filosofia da Tecnologia: um convite. Ed. da UFSC, Florianópolis (2013)
17. Scharff, R., Dusek, V.: Philosophy of Technology: The Technological Condition: Ananthology. Blackwell Publishing, Malden (2006)
18. Kline, J.: What is technology? In: Scharff, R., Dusek, V. (eds.) Philosophy of Technology: The Technological Condition an Anthology. Blackwell Publishing, Malden (2006)
19. Yang, X., Wang, L., Laszewski, G.: Recent research advances in e-science. Cluster Comput. **12**(4), 353–356 (2009)
20. Gray, J.: Jim Gray on eScience: a transformed scientific method. In: Hey, T., Tansley, J., Tolle, E. (eds.) Fourth Paradigm: Data-intensive scientific discovery. Microsoft, Redmond (2009)
21. Jankowski, N.: Exploring e-science: an introduction. Special theme: e-science. J. Comput.-Med. Commun. **12**(2), 549–562 (2007)
22. Shcmillen, H.: Library and Information Science Education and eScience: Current State of ALA Accredited MLS/MLIS Programs in preparing Librarians and Information Professionals for eScienceNeeds. Capstone Projects, Denver (2015)
23. Davenport, T.: Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business Review Press, Boston (2014)
24. Genbank Homepage (2018). https://www.ncbi.nlm.nih.gov/genbank/. Accessed 15 Aug 2019
25. PAN-STARRS Homepage. https://panstarrs.stsci.edu/. Accessed 15 Aug 2019
26. Fay, D.: Earth and environment introduction. In: Hey, T., Tansley, J., Tolle, E. (eds.) Fourth Paradigm Data-intensive scientific discovery. Microsoft, Redmond (2009)
27. Zhu, Y., Xiong, Y.: Towards data science. Data Sci. J. **14** (2015)
28. Chen, L.M.: Introduction: data science and bigdata computing. Mathematical Problems in Data Science, pp. 3–15. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25127-1_1
29. Nauer, P.: https://en.wikipedia.org/wiki/Peter_Naur. Accessed 15 Aug 2019

30. Hayashi, C.: What is data science? Fundamental concepts and a heuristic example. In: Hayashi, C., Yajima, K., Bock, H.H., Ohsumi, N., Tanaka, Y., Baba, Y. (eds.) Data Science Classification, and Related Methods. STUDIES CLASS. Springer, Tokyo (1998). https://doi.org/10.1007/978-4-431-65950-1_3
31. Burton, M.: Data science in libraries. Bull. Assoc. Inf. Sci. Technol. **43**(1), 33–35 (2017)
32. Aalst, W.M.P.: Data scientist: the engineer of the future. In: Mertins, K., Bénaben, F., Poler, R., Bourrières, J.-P. (eds.) Enterprise Interoperability VI. PIC, vol. 7, pp. 13–26. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04948-9_2
33. Voulgaris, Z.: Data Scientist: The Definitive Guide to Becoming a Data Scientist. Technics Publications, Basking Ridge (2014)
34. Koos, J.: The data librarian's handbook, by Robin Rice and John Southall. Med. Ref. Serv. Q. **37**(2), 217–218 (2018). https://doi.org/10.1080/02763869.2018.1439231
35. Hey, T., Hey, J.: E-science and its implications for the library community. Libr. Hi-Tech **24**(4), 515–528 (2006)
36. Davenport, T., Patil, D.: Data scientist: The Sexiest Job of the 21st Century. Harvard Business Review, Cambridge (2012)
37. Burton, M., Lyon, L., Erdmann, C., Tijerina, B.: Shifting to data savvy: the future of data science in libraries. Project Report. University of Pittsburgh, Pittsburgh, PA (2018)
38. Cradle Homepage. https://pt.coursera.org/learn/data-management. Accessed 15 Aug 2019
39. Mantra Homepage. https://mantra.edina.ac.uk/. Accessed 15 Aug 2019
40. Library Carpentry Homepage. https://librarycarpentry.org/. Accessed 15 Aug 2019