

GOOD SCIENTIFIC PRACTICE IN THE DRAWING UP OF DATA MANAGEMENT PLANS

BOAS PRÁTICAS CIENTÍFICAS NA ELABORAÇÃO DE PLANOS DE GESTÃO DE DADOS

¹Clara Alcina Guandalini

²Ariadne Chloë Mary Furnival

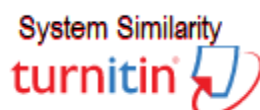
³Ana Carolina Simionato Arakaki

Universidade Federal de São Carlos^{1,2,3}

Correspondence to Author

Clara Alcina Guandalini
Universidade Federal de São Carlos
São Carlos, SP - Brazil
E-mail: clara.alcina@gmail.com

 ORCID: <https://orcid.org/0000-0002-4382-1779>



JITA: FJ. Knowledge management

e-Location: 019034

ABSTRACT

Institutions and research agencies now require scientific conduct that collaborates with the agenda for Open Science. The drawing up of a Data Management Plan can be characterized as an initial procedure for the development of scientific research, a document that proposes that researchers manage the raw data of their research, promoting sharing, the opening up of data and its reuse for the benefit of researchers and society. The aim of this work was to assess requirements for Data Management Plans as part of new conduct for good practices in scientific research. The research was characterized as exploratory; document and bibliographic methods were used for the identification, collection and systematization of information. The results point to the need for the availability of these Plans in open data repositories and their relationship to higher education institutions, funding agencies and researchers. Actions that are carried out in the creation and development of Data Management Plans were also identified, mainly for the development of institutional policies for the management of research data, but also respecting different areas of knowledge, since not all data is generated in the same way. In addition, it is evident that there is a new paradigm in scientific publication and dissemination, which also influences its communication, be it within the academic environment or with society in general.

KEYWORDS

Data management plan. Research data. Data management. Open Science.

RESUMO

As instituições e as agências de pesquisa vêm exigindo uma conduta no contexto da agenda em prol da Ciência Aberta. Sendo que, a elaboração de um Plano de Gestão de Dados pode ser caracterizada como um procedimento inicial para o desenvolvimento da pesquisa científica, documento que propõe ao pesquisador gerenciar os dados brutos de sua pesquisa, valorizando o compartilhamento em conjunto com a abertura dos dados e seu reuso em benefício aos pesquisadores e à sociedade. Diante disso, o objetivo desse trabalho foi verificar a elaboração de um Plano de Gestão de Dados como parte das novas condutas para as boas práticas na pesquisa científica. A pesquisa caracterizou-se como exploratória, empregou-se métodos documentais e bibliográficos para a identificação, coleta e sistematização das informações. Os resultados visaram a disponibilidade desse Plano em repositórios de dados abertos e a relação das instituições de ensino superior, agências financiadoras e pesquisadores com o levantamento de ações que são realizadas na criação e desenvolvimento de Plano de Gestão de Dados, principalmente para o desenvolvimento de políticas institucionais para o gerenciamento dos dados de pesquisa, porém, respeitando também as áreas do conhecimento, visto que nem todos os dados são gerados do mesmo modo. Por fim, considera-se a presença de um novo paradigma na publicação e divulgação científica, o que influencia também sua comunicação, seja com o meio acadêmico ou com a sociedade.

PALAVRAS-CHAVE:

Plano de gestão de dados. Dados de pesquisa. Gerenciamento de dados. Ciência aberta.

1 Introduction

The production of scientific knowledge involves principles of integrity, respecting an ethical conduct by researchers. With this in mind, the good practices that are presented throughout this text reflect this scientific integrity. The term “good practices” can be understood as the conduct adopted for greater development and dissemination of science, aiming for its openness to society.

The steady growth of scientific production is making educational institutions and funding agencies increasingly aware of this conduct in the context of the Open Science agenda. Thus, good practices in scientific research may include the drawing up of a Data Management Plan (hereafter, DMP), which proposes that researchers manage the raw data of their research. It can be said that the DMP is a new way of doing science, which values the sharing, openness and reuse of data for the benefit of researchers and society: so the need to develop a DMP responds to new demands and recommendations concerning the opening up of scientific data.

The Open Data movement is a term that proposes transparency in data dissemination, aiming for its possible reuse. This movement arises in the context of a broader one, Open Science, which includes the promotion of open data in science and consequently portrays a more democratic and accessible way of doing science, as well as emphasizing the need and importance of replicating and reproducing research, especially experimental research. These movements in science lead to new perspectives in the development of scientific research and lead to the emergence of questions in scientific publications related to: data archiving, support in opening data, advantages and disadvantages of open data, how to open up data, and how to describe open data in a DMP.

This study took as its starting point that the elaboration of a DMP is already a requirement in the scientific community, given that institutions and especially the research funding agencies – national and international – stipulate such a requirement. Given this, the aim of this work was to verify the elaboration of a Data Management Plan as part of the new conduct for good practices in scientific research.

The research was exploratory, using bibliographic and document sources which enabled an analysis and discussion of the topic, after careful identification, collection and systematization of information. Document collection was conducted mainly from internet pages of institutions, projects, repositories and databases. This information search and collection prioritized articles by professionals from the area, institutions that have open initiatives, and governmental and academic institutions, based on the assumption that they would be aware of these actions in scientific communication. The search for bibliographic publications also used the CAPES Journal Portal, Google Scholar search tool and other databases such as SciELO and BRAPCI.

Before addressing DMPs specifically, we need to contextualize them with a brief outline of the Open Science movement, within which a concern for open scientific data, as well as ways of managing them arises, which we will do in the following sections.

2 Open Science

Since 2001, open access to scientific communication has emerged in order to disseminate, openly and cost-free, the fruits of scientific research, both in the form of traditional publications, such as articles, reports, theses and dissertations, and later, to scientific data. The Budapest Open Access Initiative (BOAI), a meeting organized in 2001 by the Open Society Institute (OSI), aimed to discuss open access to scientific literature and which initiatives could drive this movement. The BOAI declaration states that open access to scientific literature means:

[...] its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution and the only role for copyright in this domain should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. (BOAI15, 2017)

Ten years later, in 2012, the BOAI organizers produced a publication stressing that open access to scientific publications was still far from being achieved, and made further recommendations for the next ten years on how the initiative could continue to grow and consolidate. As noted by Furnival and Silva-Jerez (2017, p.153), open access to scientific publications in fact:

[...] is part of a broader scenario for openness to knowledge in general (open access, open data, open educational resources, free software, open licenses) and is essentially a movement towards information and knowledge as a public good..

This goal to attain openness to knowledge is known as Open Science, and the expansion of the internet and resulting Information and Communication Technologies (ICT) has propelled the context of Open Science to a new proportion, enabling other ways of sharing and collaboration primarily between researchers and society. Open Science can be considered as a social movement, as it demonstrates new concepts in scientific practices, as stated by Albagli (2015, p.14):

The open science movement, in its present form, actually reflects new ways of thinking and practicing science, with direct repercussions on institutional commitments, norms and frameworks that directly interfere with scientific practice and its relations with society (our translation).

The concept of Open Science today goes beyond providing open access content: the practice of science itself becomes accessible by offering access not only to the final research product but to all its development, allowing the data to be used, reused and distributed. This change in research implies greater collaboration, facilitating scientific communication and consequently further progress, as Chan, Okune and Sambuli point out that:

[...]open and collaborative science also promises to increase the visibility and impact of research at the local level, facilitate the participation of researchers in local and international collaborations, stimulate public engagement with science through activities such as citizen science, and promote research. knowledge sharing culture, as well as a new reflection on social innovation. (CHAN; OKUNE; SAMBULI, 2015, p. 103, our translation).

In order for this new format of science to take place in a viable way, and since science is data-based, initiatives have emerged to promote and enable the open sharing of scientific data generated in research on which the writing of scientific articles and reports is based. Open scientific data imply new challenges for science, both technological and political, as Albagli points out:

[...] in the development of open science, beyond the technical and technological aspects (such as the development of free tools, availability of open computing platforms, and technological infrastructure for data sharing), there are the cultural, political and institutional issues (formal and informal) that most interfere with the open or proprietary character of these practices. (ALBAGLI, 2015, p.17, our translation).

In 2011, the Research Information Network (RIN) and the National Endowment for Science, Technology and the Arts (NESTA), both from the United Kingdom, published the report *Open to All? Case studies of openness in research*. This report aims to provide researchers, institutions and funders with better insight into Open Science methods. To that end, the report analyzed different case studies and, according to the participating researchers, what were the consequences, benefits and barriers in opening up research data. The benefits that stood out were: research efficiency; research quality and academic rigor; visibility and scope of engagement; new research questions; collaboration and community building; and social and economic impact. Barriers and restrictions were identified as: lack of evidence of benefits; lack of incentives, rewards and support; lack of time, skills and other resources; cultures of independence and competition; quality and usability concerns; and ethical, legal and other restrictions to openness.

While scientific communication was undergoing the impact of information and communication technologies, the Royal Society's report *Science as an Open Enterprise: Open Data for Open Science* (2012) made some recommendations on fundamental principles:

This report analyses the impact of new and emerging technologies that are transforming the conduct and communication of research. The recommendations are designed to improve the conduct of science, respond to changing public expectations and political culture and enable researchers to maximise the impact of their research. (ROYAL SOCIETY, 2012, p.10).

Another term that emerges in this scenario is “Citizen Science”, as a more participatory science, with the factors that give rise to this greater collaboration being the expansion of digital and sharing technologies. More active citizen participation through the internet and mobile apps points to greater transparency of data and information in accordance with public interests on certain societal issues, including academia and science. According to Parra (2015, p. 126): “Citizen science refers to the engagement of the general public in scientific research activities when citizens actively contribute to science, whether through their intellectual endeavour, their local knowledge or their tools. and resources.” Examples of far-reaching citizen science projects include Zooniverse, Asteroid Zoo, and Cochrane Crowd, among many others around the world¹.

Through this greater citizen participation and collaboration, the production of data is currently exponential, and as citizens, we make our data available to different companies every day through social networks, applications and websites, usually automatically, without worrying about the destination and long-term archiving of this data. Here, we can introduce the concept of Big Data, which comprises:

[...] the generation, processing and analysis of large volumes of data that exceed conventional processing capabilities, and which is also being exploited by companies, governments, and other sectors interested in extracting information from large amounts of unstructured data. E-Science, on the other hand, incorporates, besides the intensive use of data, collaborative scientific research and the use of shared resources for its exploration. (ALBAGLI; APPEL; MACIEL, 2014).

The concept of e-Science differs from Open Science, as the former refers primarily to the collaboration that can be made at a distance between researchers and research groups using ICT. Undeniably, the term e-Science contributed to the principle of Open Science; but it is also important to stress that research carried out in collaboration and mediated by computational informatics infrastructure can be identified as e-Science, but not always as Open Science (ALBAGLI; APPEL; MACIEL, 2014).

Open access concepts have expanded, driving a movement of change in both governmental and scientific data sharing. From Open Science, one is aware that scientific data also needs to be more widely disseminated in collaboration with the scientific community, according to new technologies that jointly enhance a more participatory Web. It is clear that the Open Science movement encompasses open data and it is necessary to reflect on how this data can be ethically disclosed and how it can be stated in a Management Plan.

¹ See the list of more than 200 citizen science projects at: https://en.wikipedia.org/wiki/List_of_citizen_science_projects

3 Open Scientific Data

Open data, defined by the Royal Society (2012, p.12) as: “Data that meets the criteria of intelligent openness, and must be accessible, usable, assessable and intelligible”, represent a new format of open access, and this openness is of paramount importance for Open Science to develop, as Molloy points out:

The more data is made openly available in a useful manner, the greater the level of transparency and reproducibility and hence the more efficient the scientific process becomes, to the benefit of society. This viewpoint is becoming mainstream among many funders, publishers, scientists, and other stakeholders in research, but barriers to achieving widespread publication of open data remain. (MOLLOY, 2011, p.1).

In the *Open Data Handbook*, by Open Knowledge International: “Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.” (OPEN KNOWLEDGE INTERNATIONAL, s.d.). The Open Knowledge Foundation (OKF) is an important organization that has been supporting these initiatives since its foundation in 2004, and according to its website, its mission is:

[...] to see enlightened societies around the world, where everyone has access to key information and the ability to use it to understand and shape their lives; where powerful institutions are comprehensible and accountable; and where vital research information that can help us tackle challenges such as poverty and climate change is available to all. (OKF, s.d.,).

Due to the continuous evolution of ICT, the stages of data production, such as those concerning its treatment, storage, sharing, dissemination and ethical use, needed to be described and analyzed in detail (SAYÃO; SALES, 2014). Researchers often spend considerable time collecting and analyzing vast amounts of data, so managing them becomes essential to ensure research efficiency, viability, and productivity. In this context, the software “Laboratory Information Management System” is an example of a system that works to improve the performance of a laboratory, manages documents, provides data entry manually or automatically, assists in the preparation of reports, etc. It is known mostly in the exact sciences and its functionality may vary. According to Tolle, Tansley and Hey (2011, p. 21), it is a system that reorganizes information in order to organize data so that it can be deposited in a database where it can be published for the general public.

In addition to software and tools, there has been a proliferation of principles and guidelines regarding the open data management cycle. For example, the Panton Principles, developed in 2010 by a group of researchers at the University of Cambridge in England, recommend some guidelines for genuinely opening up scientific data, including:

- When publishing data make an explicit and robust statement of your wishes.
- Use a recognized waiver or license that is appropriate for data.

- If you want your data to be effectively used and added to by others it should be open as defined by the Open Knowledge/Data Definition – in particular non-commercial and other restrictive clauses should not be used.
- Explicit dedication of data underlying published science into the public domain via PDDL or CCZero is strongly recommended and ensures compliance with both the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition. (PANTON PRINCIPLES, s/d).

There are also the FAIR Principles, an acronym for data that is Findable, Accessible, Interoperable and Reusable, drawn up in 2014 by Lorenz Workshop and published by Future of Research Communications and e-Scholarship - FORCE11. The four properties prescribe facets for open data, and determine that:

1. To be Findable any Data Object should be uniquely and persistently identifiable
2. Data is Accessible in that it can be always obtained by machines and humans;
3. Data Objects can be Interoperable only if (Meta)data is machine-actionable and (Meta)data formats utilize shared vocabularies and/or ontologies, and (Meta)data within the Data Object should thus be both syntactically parseable and semantically machine-accessible.
4. For Data Objects to be Re-usable they should be compliant with principles 1-3 and (Meta)data should be sufficiently well-described and rich that it can be automatically (or with minimal human effort) linked or integrated, like-with-like, with other data sources and published Data Objects should refer to their sources with rich enough metadata and provenance to enable proper citation (FORCE 11, s.d.).

Noteworthy here is the work of the Open Science Framework (OSF), an effective proposal for good scientific practice, offering a solution for how it can be attained. This system features three badges which are: Open Data, Open Materials, and Preregistered, as seen in Figure 1 below:

Figure 1. The badges developed by the OSF for good practices in scientific outputs



Source: OSF (2016).

The Open Data badge is obtained when digitally shareable data is publicly available and is required to reproduce the reported results, and should be made available in an open access repository, with the description of the data and the application of open licenses. The Open Materials badge is obtained by making publicly available the components of the research methodology required to reproduce the reported procedure and analysis, and the components should be described in order to understand the relationship of materials with the reported methodology. Finally, Preregistered would be a plan of how the research will be conducted, its

method design, assumptions, sample size, detailed description of variables, among other information (OSF, 2019), all pre-registered in an open access institutional system (for example ClinicalTrials.gov). This last badge resembles the record of research on the Ministry of Health's "Plataforma Brasil", although this platform is not yet open access, as it sometimes contains sensitive information about human participants.

As with open access to publications, designing and implementing an open data policy requires a change that influences the way the results of scientific research are disseminated, and is a challenge for researchers as well. Scheliga and Friesike (2014) raised some obstacles to Open Science based on interviews that the authors conducted with researchers, dividing the obstacles into individual and systemic ones. For individual obstacles, they point out that the "(...) fear of free-riding is a recurring topic in the interviews. Scientists fear that if they release their research materials early on in the research process, they expose themselves to intellectual property abuse." (SCHELIGA; FRIESIKE, 2014). As for systemic obstacles, they noted that "(...) Many researchers stated that open science efforts are not rewarded by the current academic system." (SCHELIGA; FRIESIKE, 2014), which points to the evaluation of scientific productivity still being very much based on the publication of articles, only, and especially in journals with high impact factors that often do not permit open access to the article on publication. But there were also positive responses in the study: "Many interviewed researchers have expressed their endorsement for the idea of open science. They describe that sharing knowledge via the Internet provides them with expanded opportunities to exchange feedback and to collaborate with scientists internationally." (SCHELIGA; FRIESIKE, 2014).

After this brief outline of the context of Open Science and open data, the importance of elaborating a Data Management Plan (DMP) becomes clear, given that it can be done from an understanding of its relevance in the academic environment, resulting in good scientific practice.

4 Data Management Plan

The Data Management Plan (DMP) is a document that aims to describe the treatment of data during a research project and what will happen to this data after the research is completed; that is, these plans should address the data lifecycle from discovery, collection and organization up until how it will be preserved. (MICHENER, 2015). The DMP template will usually be developed and suggested by educational institutions, public agencies and research funding agencies, which in turn require their researchers to include a DMP in their proposal, which details the lifecycle of their research data, thereby aiming to develop a standard that can be made available with details of how data management will be carried out, thus constituting a practice in scientific communication.

Michener (2015) presents ten guidelines for how to create a good DMP:

1. Determine the research sponsor requirements;
2. Identify the data to be collected;
3. Explain how the data will be documented;
4. Explicar como os dados serão documentados;
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy
7. Define the project's data policies
8. Describe how the data will be disseminated
9. Assign roles and responsibilities
10. Prepare a realistic budget.

The above guidelines really need to be considered when designing a DMP according to the policies established by the institution to which the researcher is linked, as they serve as an excellent rule of thumb, avoiding misconceptions and offering what is essential to draw up the DMP in an understandable way. It is also important to note that the data to be made available must be certified with an open license to ensure that it can be legally reused.

Because many research funding agencies have started demanding DMPs, the Institute for Social Research (ISR) at the University of Michigan, USA, also provides guidance to assist researchers in how to create them. The document “Guidelines for Effective Data Management Plans” (2012) gives some recommendations for what to consider when developing a DMP, as well as examples, which resources can be used, and some other information. According to these recommendations, it is paramount to include in a DMP: data description, access and sharing, domain repository such as the ISR, metadata, intellectual property rights, ethics and privacy; format, archiving and preservation, storage and backup, security, liability, existing data, selection and retention periods; audience and users; data organization, quality assurance, expenses and legal requirements.

The drawing up of a DMP occurs for several reasons, and one of the most important today is to embrace good practices in the scientific community. When we refer to “good practices”, responsible conduct is assumed to ensure reliability and integrity.

Analyzing the guidelines and recommendations outlined above, the following are some of the elements/components that can be considered as fundamental when developing a DMP, as they recur in several examples researched here. These are:

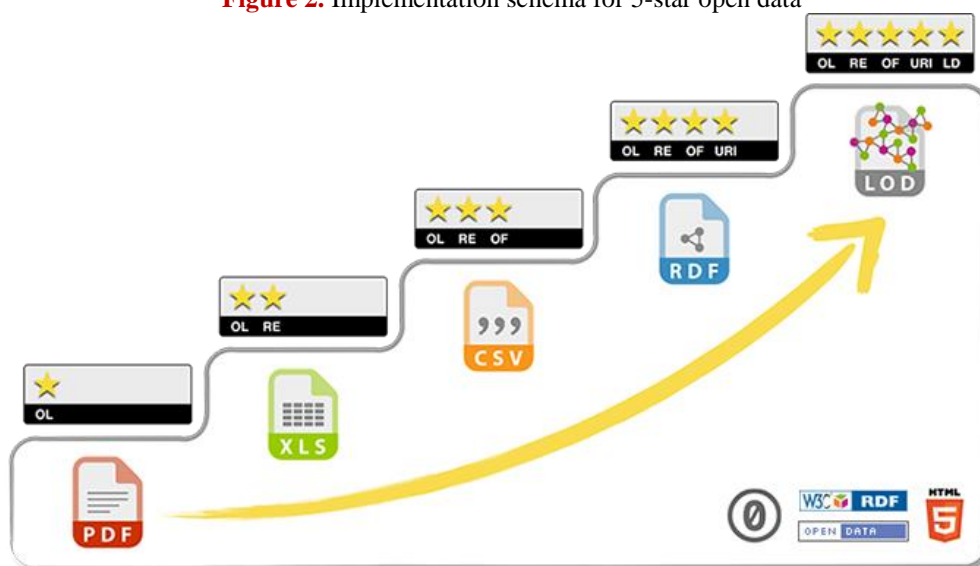
Identify what types of data present in the research and under which licenses they should be “protected”: This topic can vary greatly depending on the area of knowledge as there are different types of data such as observational, computational, experimental, etc. Thus, it is necessary to observe what data will be worked on in order to choose how they will be preserved, and it is important to choose a corresponding license to obtain copyright so that in the future, the data may be reused with due responsibility.

Data format: In this item, it is intended that the data really are open. As mentioned earlier, one purpose of open data is that it can be reused. Data formats (textual, numeric, multimedia, software language, etc.) may vary and thus their preservation also changes, and in the age of “digital information”, it is necessary that they can be interpreted by machines to ensure their preservation. As SAYÃO and SALES (2015, p.59) observe:

Even considering the retrospective compatibility of many software packages - which allow data created in previous versions to be read in current versions of the software - and interoperability between competing software, the safest option to ensure long-term access is to convert data to standardized formats.

Este item também nos lembra das “5 estrelas dos dados abertos”², esquema proposto por Tim Berners-Lee, e que tem como objetivo apontar o quanto os dados estão disponíveis e conectados na *web*, conforme demonstrado na figura 4. This item also reminds us of the “5 stars of open data” scheme proposed by Tim Berners-Lee, which points to how data is available and connected on the web, as shown in Figure 4.

Figure 2. Implementation schema for 5-star open data



Source: HAUSENBLAS; KIM (2012).

In this schema, one star means that data is available on the web regardless of its format, but with an open license. With two stars, the data is more structured to be machine readable, and with three stars, there is already the possibility of a non-proprietary format. Four stars indicate that URIs have been used for resource identification, and finally, five stars indicate that in addition to machine readable data, there is also has the ability to connect for interoperability.

² Disponível em: <https://5stardata.info/pt-BR/>. Viewed: 26 nov. 2018.

Metadata: Metadata are important because they have the function of describing the data: how they were used in the research, through which paths they passed, if they were processed, and how.

Realising the benefits of open data requires effective communication through a more intelligent openness: data must be accessible and readily located; they must be intelligible to those who wish to scrutinise them; data must be assessable so that judgments can be made about their reliability and the competence of those who created them; and they must be usable by others. For data to meet these requirements it must be supported by explanatory metadata (data about data). (ROYAL SOCIETY, 2012, p.7).

Metadata are also referred to as data documentation, which is considered to be “one of the best practices in data creation, organization and management, and is an important strategy for the digital preservation of data.” (SAYÃO; SALES, 2015, p. 27).

Choice of repository or database to archive data: There are currently several repositories for storing data. Some recommended in the literature consulted were: <https://figshare.com/> and <https://datahub.io>. Many others are logged in the Registry of Research Data Repositories (www.re3data.org), which is a global repository of data repositories for researchers, funding agencies, editors, and academic institutions. It was founded in 2012 and is funded by the German Research Foundation (DFG), and is an important tool for identifying which repository to select for which research data. Its content is licensed under an International Creative Commons License, and is of great relevance to the concept of open data. Data Portals (<http://dataportals.org>) is also a similar option.

Analyzing the whole context presented, there is an understanding that DMPs play an important role in scientific communication and influence the relationship between authors and organizations. It can be considered that the DMP is an activity that aids good scientific practices because it proposes that the data be used and made available in a coherent way within the digital medium that today structures scientific processes and output. DMPTool is an online service that provides information to help researchers prepare a free and open source Data Management Plan. In addition, the site also offers support aimed at the researcher meeting the requirements of funding agencies. The development of DMPTool was conceived in the context of perceived demand for DMPs required by research funding agencies, and then emerged in 2011 with the help of contributing institutions. Examples of DMPs can be found on the site provided by several institutions. Another tool that also offers data management plan templates is DMPonline, created in partnership with the Digital Curator Center (DCC) and the University of California Curator Center (UC3) to propose collaboration between funders and universities to provide up-to-date DMPs that meet the requirements throughout the scientific project lifecycle.

5 Information Science in Open Scientific Data Policies

Research data in the current digital context generates a flow that grows exponentially, being difficult to store correctly for reuse in new projects, so data management is important in the scientific field. Given this, there are new perspectives on scientific output that require efficient data management. Sayão and Sales (2012, p. 182) cite a fourth paradigm:

In the fourth paradigm we have science unifying experiments, theories and simulations through the intensive use of data captured by increasingly sophisticated instruments or generated by simulation, processed by software and stored in computers in the form of databases.

The term “Digital Curation” emerges in this scenario of digital resource management as a practice aimed at preservation and access, involving a set of techniques. (SAYÃO; SALES, 2012). Research data increasingly require the elaboration of a DMP, as discussed earlier. Thus, it can be understood that scientific data management and e-Science are also related to Digital Curation.

Generally speaking, scientific data curation adds speed to the cycle of scientific communication by providing researchers with data ready for reuse, i.e. treated data, accompanied by semantic and structural metadata - which ensure the reliability of its meaning. and the correct reconstruction of its presentation, coupled with metadata that ensure integrity, accuracy and authenticity. (SAYÃO; SALES, 2012, p. 188).

It is recommended that scientific data be stored in specific data repositories if the institution offers the option of a digital repository, or that a search for the most appropriate one be made, depending on the type of research. Performing a search for Brazil on re3data.org, we retrieved eight recommended data repositories: Exploration and Production Database; WorldClim - Global Climate Data; GLOBE (Global Collaboration Engine); IODP (International Ocean Discovery Program); Biological Survey Data Repository (PPBio Data Repository); Brazilian Institute of Information Science and Technology (IBICT) Dataverse Network; Scientific Database of the Federal University of Paraná and Center for Documentation and Digital Collection of Research of the Federal University of Rio Grande do Sul (CEDAP Research Data Repository).

Scientific publications here analysed also point to a topic that is also very important for good practices in scientific communication involving information professionals: attention to copyright.

Among the various academic practices and ethical and legal norms that structure the practice of scientific publication, copyright is today a fundamental theme for understanding the challenges of scientific communication and, above all, the role that libraries play in this mission. (SIQUEIRA, 2015, p.30).

Copyright influences innovation strategies, with the movement of open data and development of management of this data in the scientific environment becoming an essential aspect, mainly because scientific production and outputs are fully inserted in the digital environment. According to Siqueira (2015, p. 31):

Author rights are based on the idea that the author of a work should be granted the privilege of exclusive exploitation of its economic benefits, as well as the right to moral recognition of its authorship, for a limited period of time, so that it encourages the creation and stimulates the circulation of intellectual works.

With the development of open scientific data management, scientific communication benefits by bringing the opportunity for improvements in knowledge generated by research. Thus, copyright helps researchers share their data more securely and can also prove to be a barrier considering some obligations and the lack of a more specific law in this regard. "There is a feeling in the academic and scientific community that the balance between access to knowledge and protection for the author has been lost." (SIQUEIRA, 2015, p. 32).

Research Data Services (RDS) began to emerge in academic libraries due to increased demands for data management and sharing by funding agencies (TENOPIR et al., 2014). Tenopir et al. (2014) address the need for data management and how libraries are prepared when developing and planning RDS, and how library policies also need to be aligned with the perception of librarians, thus enabling a good RDS service for the academic community. Focusing on North American academic research libraries, the authors raise several questions about how RDS have been conceived and offered, and they reinforce how the library can offer more informative and advisory services such as assisting researchers in finding repositories to deposit data and giving examples of DMPs, or more practical services such as working in the institution's repository and assisting with the writing of DMPs. The main focus of the study is to identify which libraries (in the US and Canada) already offer data management related services, and also if librarians are inserted in this medium, and to compare library policy on RDS and the perceptions of librarians regarding their implementation. The authors conclude that:

It is clear that some academic research libraries are offering a variety of research data management services and more plan to do so within the next two years. Most commonly these services are extensions of traditional informational or consultative services, such as helping faculty and students locate datasets or repositories. A small, but growing, number of libraries are becoming more involved with research data, from helping with data management plans to preparing and preserving research data for deposit in data repositories. (TENOPIR et al., 2014, p. 89).

"Research Data Librarian" is a term that has appeared in some job descriptions for librarians. A vacancy published at the London School of Economics and Political Science (LSE) in May 2018 describes that the practitioner: "[...] is responsible for providing expert advice and training to researchers and staff in both the use of resources. research data as well as the creation and preservation of their own research data." (LSE, 2018). There is a notable

tendency for information professionals to be increasingly involved with open scientific data, because it is a subject that needs management and elaboration of practices for its future retrieval and use. Just as libraries can be inserted into open data projects, according to the opinion of Ferrer-Sapena, Peset and Aleixandre-Benavent:

Our opinion is that libraries cannot remain indifferent to the movement of open data. In recent years, open standards, open source software, free access to publications and currently free access to data have been treated almost daily. An example in which its active role can be seen is the DISC-UK Datashare (Data Information Specialist Committee) project in the United Kingdom, whose objective is that the country's research libraries serve as conservers of management support data and to research activities through the creation of institutional open repositories and web 2.0 technologies. (FERRER-SAPENA; PESET; ALEIXANDRE-BENAVENT, 2011, p. 268).

In the “Research Data Management Guide for Librarians and Researchers,” the authors indicate that the guide is also intended for librarians because:

Librarians are well-positioned to work with data because of their expertise in information management, metadata, resource discovery, digital preservation, and they have always established a long and productive relationship with researchers. (SAYÃO; SALES, 2015, p.6).

It is therefore undeniable that librarians are well-prepared to be involved in data management, especially in higher education institutions, promoting their skills related to the representation, retrieval and access of information, as well as being up-to-date concerning the demands of open access in the digital environment.

6 Concluding Remarks

From the study here presented, which focussed on the concepts of Open Science, Open Scientific Data and Data Management Plans, we observed that there is really an awareness on the part of educational institutions regarding the organization of research data, made available for reuse, by means of requiring DMPs from researchers. However, there is also a need for support from these institutions, either in the form of guidance or providing ready-made templates, and further development of this culture of open science initiatives so that it becomes more widespread and normalized, giving researchers greater confidence.

It is also important to highlight the fact that librarians can be professionally involved in this area, and the DMPs involve decisions on how to manage the research data, implying that the information professional can perform an informational support function to the user and still exercise their skills actively in the fields. educational institutions and may be part of the scientific programs and policies of their digital repositories.

A policy is required advocating widespread use of scientific Data Management Plans, giving priority to data openness to guarantee their ethical reuse. Thus, it would be interesting to have an agreement between institutions for a possible standardization of such plans, but which also respect the different knowledge areas, since not all data are generated in the same way. In addition, it is evident that there is a new paradigm in scientific publication and dissemination, which also influences its communication, either with academia or with society.

References

- ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. L. E-science, ciência aberta e o regime de informação em ciência e tecnologia. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 7, n. 1, 2014. Available at: <http://ridi.ibict.br/bitstream/123456789/854/1/124-540-1-PB.pdf>. Viewed: 19 April, 2018.
- ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília; Rio de Janeiro: IBICT; UNIRIO, 2015. p. 9-26.
- ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, p.434-450, 5 dez. 2014. Available at: <http://revista.ibict.br/liinc/article/view/3593>. Viewed: 26 November, 2018.
- AVENTURIER, P. **Plano de Gestão de Dados**: uma introdução. 2017. Available at: <https://publicient.hypotheses.org/1660>. Viewed: 7 March, 2018.
- BOAI15. Budapest Open Access Initiative. Available at: <https://www.budapestopenaccessinitiative.org/boai15-1>. Viewed: 19 June, 2018.
- CHAN L.; OKUNE, A.; SAMBULI, N. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento? In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília; Rio de Janeiro: IBICT; UNIRIO, 2015. p. 91-120.
- DELFANTI, A.; PITRELLI, N. Ciência aberta: revolução ou continuidade? In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília; Rio de Janeiro: IBICT; UNIRIO, 2015. p. 59-70.
- DUDZIAK, E. A. **Gestão de dados de pesquisa**: o que precisamos saber hoje! 2018. Available at: <http://www.sibi.usp.br/noticias/gestao-de-dados-de-pesquisa-o-que-precisamos-saber-hoje/> Viewed: 17 January, 2018.
- FECHER, B.; FRIESIKE, S. **Open Science: One Term, Five Schools of Thought**. 2014. p.17-47.

FERRER-SAPENA, A.; PESET, F.; ALEIXANDRE-BENAVENT, R. Acceso a los datos públicos y su reutilización: open data y open government. **El profesional de la información**. p. 260-269, 2011, v. 20, n. 3.

FURNIVAL, A. C. M.; SILVA-JEREZ, N. S. Percepções de pesquisadores brasileiros sobre o acesso aberto à literatura científica. **Inf. & Soc.: est.**, João Pessoa, v. 27, p.153-166, 2017.

FORCE 11. **Guiding Principles For Findable, Accessible, Interoperable And Re-Usable Data Publishing Version B1.0**. Available at: <https://www.force11.org/fairprinciples> Viewed: 30 October, 2018.

HAUSENBLAS, M; KIM, J.G. 5-star deployment scheme for Open Data.png. Available at: https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png Viewed: 30 October, 2018.

HYLA SOFT. **Sistemas de Gerenciamento de Informação Laboratorial (LIMS)**. Available at: https://www.hylasoft.com/pt_br/solution/laboratory-information-management-system-lims. Viewed: 30 October, 2018.

ICPSR. **Guidelines for Effective Data Management Plans**. 2012. Available at: <https://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>. Viewed: 4 April, 2018.

ICPSR. **Data Management & Curation**. Available at: <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html>. Viewed: 4 April, 2018.

MACHADO, J. Dados abertos e ciência aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília; Rio de Janeiro: IBICT; UNIRIO, 2015. p. 201-228.

MICHENER, W, K. Ten Simple Rules for Creating a Good Data Management Plan. **PLOS Computational Biology**. 2015. Available at: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004525>. Viewed: 7 March, 2018.

MOLLOY, J. C. The Open Knowledge Foundation: open data means better science. **PLoS Biology**, United Kingdom; v.9, December, 2011. Available at: <https://doi.org/10.1371/journal.pbio.1001195>. Viewed: 7 March, 2018.

OPEN DATA COMMONS. **Legal tools for Open Data**. Available at: <https://opendatacommons.org/licenses/>. Viewed: 20 August 2018.

OPEN DATA GOVERNMENT WORKING GROUP. **8 Principles of Open Government Data**. 2007. Available at: https://public.resource.org/8_principles.html. Viewed: 18 February 2018.

OPEN DEFINITION. **The Open Definition**. Available at: <http://opendefinition.org>. Viewed: 29 March 2018.

OPEN KNOWLEDGE BRASIL. **Guia de Dados Abertos**. Available at: http://opendatahandbook.org/guide/pt_BR/. Viewed: 1 March, 2018.

OPEN KNOWLEDGE INTERNATIONAL. Open Definition. **Conformant Licenses**. Available at: <http://opendefinition.org/licenses/#Dat>. Viewed: 20 August, 2018.

OPEN SCIENCE FRAMEWORK. Available at: <https://osf.io/tvyxz/wiki/home/>. Viewed: 15 February, 2018.

PANTON PRINCIPLES. **Principles for open data in science**. Murray-Rust, Peter; Neylon, Cameron; Pollock, Rufus; Wilbanks, John; (19 Feb 2010). Available at: <https://pantonprinciples.org>. Viewed: 29 March, 2018.

PARRA, H. Z. M. Ciência cidadã: modos de participação e ativismo informacional. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília; Rio de Janeiro: IBICT; UNIRIO, 2015. p. 121-142.

ROYAL SOCIETY. **Science as an open enterprise**. London: The Royal Society, 2012.

RE3DATA.ORG. **About**. Available at: <https://www.re3data.org/about>. Viewed: 8 May, 2018.

RE3DATA.ORG. **Browse by country**. Available at: <https://www.re3data.org/browse/by-country/>. Viewed: 20 August, 2018.

RESEARCH INFORMATION NETWORK. NATIONAL ENDOWMENT FOR SCIENCE, TECHNOLOGY AND THE ARTS. **Open to All? Case studies of openness in research**. 2010. Available at: <http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies>. Viewed: 8 February, 2018.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v.22, n.3, 2012, p. 179-191.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**. 2014. p.76-92.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. 90 p. Available at: <http://www.cnen.gov.br/component/content/article?id=160>. Viewed: 1 June, 2018.

SCHELIGA, K; FRIESIKE, S. **Putting open science into practice: a social dilemma?** First Monday, v. 19, 2014. Available at: <https://uncommonculture.org/ojs/index.php/fm/article/view/5381/4110> Viewed: 21 August, 2018.

SIQUEIRA, L. P. B. P. Direitos autorais e comunicação científica: desafios para bibliotecas. **Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, v. 2, n. 1, 2015. Available at: <https://www.brapci.inf.br/v/a/21564>. Viewed: 21 August, 2018.

TENOPIR, C. et al. Research data management services in academic research libraries and perceptions of librarians. **Library & Information Science Research**, v. 36, n. 2, p.84-90, abr. 2014. Elsevier BV. Available at: <https://www.sciencedirect.com/science/article/pii/S0740818814000255>. Viewed: 20 August, 2018.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE. LSE. **Job description**. Research Data Librarian. 2018.

TOLLE, K.; TANSLEY, S.; HEY, T. (Org.). Jim Gray e a eScience: um método científico transformado. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. São Paulo: Oficina de Textos, 2011. p. 17-29.

