AtoZ
novas práticas em informação e conhecimento

ISSN: 2237-826X
PAPER | ARTIGO | ARTÍCULO

Proposal for a model of research data management service

Proposta de modelo de serviço de gestão de dados de pesquisa

Luana Farias Sales[1], Luís Fernando Sayão[2]

[1] *Brazilian Institute of Information in Science and Technology – IBICT, Federal University of Rio de Janeiro - Postgraduate Program in Information Science – PPGCI, Rio de Janeiro, RJ, Brazil. ORCID: http://orcid.org/0000-0002-3614-2356*

[2] *National Nuclear Energy Commission – CNEN, Federal University of Rio de Janeiro – Postgraduate Program in Information Science – PPGCI, Rio de Janeiro, RJ, Brazil. ORCID: http://orcid.org/0000-0002-6970-0553*

Mail to/*Autor para correspondência/Correo a:* Luana Farias Sales, luanasales@ibict.br

Submitted/*Recibido*: 27 de abril de 2022; **Approved/*Aceptado***: 30 de agosto de 2022

Abstract

The planning, development, implementation, and operation of research data management platforms involves many variables that need to be properly addressed. The design of these platforms constitutes a complex and multifaceted problem, of which the solution needs to be manifested in the form of services that meet the needs of research flows in specific disciplinary domains, informational, technological and scientific parameters, varied professional profiles, political, legal and ethical constraints, and economic and temporal sustainability. In this multifaceted environment of contemporary science, the question we intend to answer is: What types of data management services can be offered based on the workflows of the scientific environment, considering the peculiarities of the disciplinary domains and serving as a support mechanism for the development of research and encouraging the sharing and reuse of scientific data? As an answer to the question, theoretical-exploratory research was developed that took as a methodological basis the analysis of the literature in the area, with special emphasis on articles, reports, manuals and data infrastructure projects, prepared by researchers and research institutions, giving prominence to the observation points of the actors most closely involved in the issue. From the interrelationship of the diverse variables raised, the result is the description of a set of services that can be offered, organized in a service model. It is hoped that this model can serve as a guide for research institutions that wish to implement data management platforms closer to their communities and that follow the standards and traditions of sharing and reusing scientific data from their researchers.

Keywords: Research data management services; FAIR data; Scientific knowledge management.

Resumo

O planejamento, desenvolvimento, implantação e operação de plataformas de gestão de dados de pesquisa envolvem muitas variáveis que precisam ser adequadamente articuladas. O projeto dessas plataformas constitui-se um problema complexo e multifacetado, cuja solução precisa ser manifestada na forma de serviços que atendam às necessidades dos fluxos de pesquisa de domínios disciplinares específicos, parâmetros informacionais, tecnológicos e científicos, perfis profissionais variados, condicionantes políticas, éticas e legais, e sustentabilidade econômica e temporal. Neste ambiente multifacetado próprio da ciência contemporânea, a questão que se pretende responder é: que tipos de serviços de gestão de dados podem ser oferecidos em uma articulação que permeie os fluxos de trabalhos do ambiente científico, considerando as peculiaridades dos domínios disciplinares e servindo como mecanismo de auxílio para o desenvolvimento de pesquisas e incentivo ao compartilhamento e reuso de dados científicos? Como resposta, desenvolveu-se uma pesquisa de cunho teórico-exploratório que tomou como base metodológica a análise da literatura da área, com ênfase especial em artigos, relatórios, manuais e projetos de infraestrutura de dados, elaborados por pesquisadores e instituições de pesquisa, dando proeminência aos pontos de observação dos atores mais proximamente envolvidos na questão. A partir do inter-relacionamento das diversas variáveis levantadas, apresenta-se como resultado a descrição de um conjunto de serviços que podem ser oferecidos, organizados em um modelo de serviços. Espera-se que esse modelo possa servir como norte para instituições de pesquisa que desejem implantar plataformas de gestão de dados mais próximas das suas comunidades e que sejam aderentes aos padrões e tradições de compartilhamento e o reuso de dados científicos de seus pesquisadores.

Palavras-chave: *Serviços de gestão de dados de pesquisa; Dados FAIR; Gestão do conhecimento científico.*

INTRODUCTION

Research data management is currently a focus of interest and one of the biggest challenges for research institutions. On a global scale, data management and curation are prominent in the 21st century research landscape, accompanying the ubiquity of digital technologies for collection, analysis and archiving in almost all disciplinary domains (Mayernik et al., 2012). As such, research institutions, to varying degrees, are reconceptualizing data management and identifying it as an integral part of research processes, reconsidering or expanding their data processing strategies, implementing management and curation platforms, acquiring analysis and visualization tools, and developing training programs for their teams.

We have come a long way in understanding and dealing with data management, and there is a growing interest on the part of the various stakeholders—researchers, funding agencies, research institutions and information professionals, scientific publishing houses and many others. However, the level of success in implementing and offering data management services within the various data-intensive research organizations has not been consistent and is generally still in the preliminary stages. The planning, development and implementation of research data management platforms, due to the number of variables that need to be addressed, is a complex and multifaceted problem. The solution to this issue needs to be articulated around workflows from specific disciplinary domains, information, technological, political, ethical and legal parameters for sustainability and expertise in an odyssey marked by constant change, the main characteristic of which is heterogeneity. In this sense, two questions stand out, which we tried to answer in the research, namely: what types of research data management services can be offered in the course of research development, taking into account the peculiarities of the disciplinary domains? Also, what types of data management services can be offered during workflows in the scientific environment, in order to help the development of research data?

This complex environment may be a suitable ground for adopting the FAIR Principles (Mons et al., 2017) as a horizon for implementing management services that make research data findable, accessible and interoperable, so that it can be reused over the long term. In this way, conditions are created for the transition from self-contained research to more collaborative, networked and relational research, which at the same time meets disciplinary requirements and benefits communities with specific cultures and constraints.

In an attempt to address this diversity, this paper aims to present a generic architecture to support the design of data services platforms, defining, realigning, aggregating and articulating the various conceptual modules—guidelines, policies, services, tools, infrastructures etc.—around a model of layers, which, like building blocks, can be adjusted according to the depth, scope, and philosophy of each institution or discipline.

Exploratory research, according to Gil (2002), is characterized by being a type of research carried out when there is little knowledge of the subject to be addressed, seeking to get to know the subject in greater depth in order to make it clearer for future hypothesis formulations. Because it focuses on new themes, exploratory research is not committed to rigid methods, and its planning is “quite flexible, so that it allows the consideration of the most varied aspects related to the fact studied” (Gil, 2002, p. 44). This research is characterized as exploratory and theoretical in nature. It explored the international literature on the subject of “research data management services”, with special emphasis on articles, reports, manuals, and data infrastructure projects drawn up by researchers and research institutions, giving prominence to the points of view of the actors most closely involved in the issue.

COMPREHENDING RESEARCH DATA MANAGEMENT SERVICES

In a scientific environment where there is intense generation and consumption of data due to the daily odyssey of scientists in the search for new knowledge, there is a critical need for devices to control and organize these assets, which ideally take the form of a set of services aligned with a coherent data management strategy. In this context of constant transition, the management of research data on a global scale is increasingly becoming an essential service that must be intermediated by the information units of research organizations (Tang & Hu, 2019).

The breadth of the actions provided by the technical and management systems of the data management platforms offers the various stakeholders that orbit around the flows of scientific knowledge generation a series of benefits that depend on their interest matrices, for example: for an academic coordinator or manager, the platform becomes a device for evaluating the productivity of a program, area, or research group; for a scientific editor and his body of reviewers, the platform serves as an instrument for evaluating and validating submitted records; for scientific policymakers and research funders, the platforms become a cartography that supports the ordering and levels of incentives needed for an area of study; and we could say that for the ordinary citizen and the mainstream press it would serve as an additional element of transparency in public investments in science. However, the researcher perceives the data as an instrument for research and the generation of new knowledge, which has transformational potential, as emphasized by Jones, Prior, and White (2013).

As such, data management is not an end in itself and takes the form, in the eyes of the researcher, of a broad spectrum of services and tools that support the entire life cycle of the data, within the scope of a research project, the benefits of which are directly perceptible by the researcher, such as: greater visibility for the research; more citations and prestige; recognition of the authorship of the data; a greater level of collaboration on a global scale; organization of the data for the researcher’s own use and that of his or her close colleagues; recognition in terms of promotion and funding; logical and physical protection of the data; and long-term preservation. This strongly indicates that in the field of data management, the organic connection with scientific communities takes the form of services that resonate with disciplinary dynamics and culture, but without losing the perspective of global and interdisciplinary insertion.

The in-depth study on the maturity assessment—a common approach to determining the level of sophistication of services and products—of data services conducted by Kouper, Fear, Ishida, Kollen, and Williams (2017) indicates that the most advanced data management services are likely to be those whose actions prioritize the needs of individual institution communities, but are also aware of the broader research communities in which individuals are embedded. The most consolidated research data management service programs are not necessarily those that offer the largest portfolios of services or employ the largest teams, "but those whose activities are most deeply connected to the mission of the library and the institution as a whole [. . .], in other words, where services are carefully chosen, organized, monitored and optimized," the authors conclude (Kouper et al., 2017, p. 164).

In this diversity of approaches, the FAIR principles offer a conceptual basis and a more substantive horizon for the design and achievement of data management services, since they are focused on ensuring that research objects are findable, accessible, interoperable, and reusable, thus summarizing the possible objectives of a data management service, or, at a higher level of abstraction, the scope of an Internet of FAIR Data and Services. The FAIR principles "describe characteristics and aspirations applied to systems and services aimed at supporting the creation of valuable research outputs that can then be rigorously evaluated and widely reused [. . .]" (Mons et al., 2017, p. 50). The principles deliberately do not specify technical requirements and are not standards, but constitute a set of guiding guidelines that establish a growing continuum of possibilities for reusing research objects through a variety of different implementations. In this direction, the FAIR principles describe the qualities or behaviors of data resources to achieve—possibly incrementally—an optimal level of discovery and academic reuse, opening up the possibility for many different approaches to service design and data reuse.

Taking the above arguments into account and using the FAIR principles as a conceptual aggregating core, we propose the following definition for **research data management services**: it is the set of informational, computational, scientific and administrative services offered within the scope of research data management and anchored in the specific needs of the academic and scientific communities, whose purpose is to make data findable, accessible, interoperable and reusable, so that it translates into benefits for science and all its stakeholders.

Based on this definition, below we propose a data management service model that seeks to represent the different aspects of research data management, without losing sight of the interrelated nature of the dynamics of the activities that take place in a data-intensive scientific environment, the aim of which is to make them FAIR.

OUTLINING THE SCOPE OF DATA MANAGEMENT SERVICES

As confirmed by Jones et al. (2013, p. 5), "to effectively support data management and sharing, an institution needs a coherent strategy and a set of services". But what could this set of data management services mean? Naturally, it has a continuous spectrum that varies in disciplinary, cultural and epistemological, institutional and political terms, and also depends on the technological bases available for data management. In fact, research institutions can offer data services in a great multiplicity, which varies not only in the types of service, but also in the depth and scope that these services are made available, the levels of specificity and commitment, and for whom and for what purposes these services are offered (Choudhury et al., 2018).

(2013) point out that data management services include the provision of information, consultancy, training and active involvement in data management planning, guidance during research (e.g., advice on data storage and file security), documentation and metadata, sharing of research data and curation (selection, preservation, archiving, citation) of completed projects and published data. From the perspective of Choudhury et al. (2018), data management services include the provision of the necessary infrastructure to carry out data curation through licenses for preservation, analysis, and access tools; the availability of space in storage systems funded by the organization for curated data; training and consultancy that allow the researcher to exploit the data services offered by the various units of the institution. Complementing this, Tang and Hu (2019) point out that in the research data management component diagram, the overarching activities include "data management policy and strategies" and "business plan and sustainability". Underlying the establishment of a research data management service, various levels of guidance, training, and support are required. For these authors, the focal point of the data management process should give prominence to the management service components of planning, active data management, selection and sharing, as well as data repositories and catalogs. In this sense, repositories or data catalogs are just one more service among many that a data management service platform can offer.

It is important to note that there are many differences between the more traditional management of resources and the level of technical requirements, infrastructures, and expertise needed to manage research data. A book, for example, has universal and standardized cataloging, the differences in treatment between disciplines are few, and its processes are focused on post-publication; the same cannot be said of research data and other digital research objects, such as databases and codes, whose management has to be concerned with the long and idiosyncratic life cycle that begins even in the planning phase—long before publication and archiving, up to post-publication, but in an even more complex process than was carried out in the management of bibliographic publications. Add to this all the peculiarities that require management to be linked to the life cycle of the research project. In this context, "The range of skills and knowledge required to deliver data management services is largely dictated by the individual phases of the project lifecycle," confirm Jones et al. (2013, p. 3). Thus, the scale of services that research institutions offer can vary not only in the types of services provided, but also in the level of depth at which they operate, and in the universe of users to whom the services are offered (Choudhury et al., 2018). Researchers, professors and graduate students are the most likely target customers of data management systems, but other stakeholders should be considered, such as S&T managers, funders and specific communities of practice—such as engineers and agronomists—who reuse data, especially data with a high degree of processing in their projects and endeavors, such as building the foundations of a nuclear power plant or selecting cultivars. The services can be distributed across several units of the institution or concentrated and coordinated by one unit, possibly the research library.

The fragmented and heterogeneous view of data management services—which ultimately reflects the multiple faces of research activity—creates an obstacle in delineating their contours and listing the diagram of their components. Possibly, a rationale based on researchers as users of these services could help in understanding and building a possible concept of data management services. This is what this essay attempts to do in the next section: propose a definition for a research data management service that takes into account the specific needs of its users.

FAIR RESEARCH DATA MANAGEMENT SERVICE MODEL

As a convenient abstraction of the reality to be understood, a model is a cultural creation, a "mindset" designed to represent a reality or some of its aspects, in order to make them qualitatively and quantitatively describable and sometimes observable (Sayão, 2001). From this point, it was decided to divide the model into four representational layers: 1) governance, where the guiding principles of the data management services project are discussed; 2) technical infrastructure, which also includes the categories of expertise required; 3) information, computational, scientific and administrative services; 4) the results of the implementation of these services manifested by the FAIRification of the data.

Research Data Governance: planning, policy, institutionalization, and sustainability

The organizational and institutional set-up in which data management is carried out can vary with regard to various aspects, such as the intensity of management support and the level of investment applied. Some institutions, such as scientific data reference centers and government statistical agencies, may be entirely dedicated to data management, with it as their main purpose. In other configurations, data management is part of a broader activity that connects to other research activities, as in the case of universities (National Research Council, 2015), whose data management activity stems from their teaching, research, and extension functions. However, even in the academic context, there are many ways of planning and carrying out data management tasks, which vary according to objective references, such as degrees of investment, available technical systems, volume, and type of data and how data management is integrated into their workflows and processes; and more subjective perceptions, such as disciplinary culture and academic prestige. In this model, these parameters are equated with a more administrative level comprised of the "data governance" category. On a more conceptual level, data governance outlines the principles, policies, and strategies that are commonly adopted in an environment that needs a coherent data management program; it also outlines the actions, functions, and roles that are needed to implement these policies and strategies. Within a research institution, the principles operationalized by governance govern the entire lifecycle of data—from conceptualization to archiving and possible disposal. As such, the data governance process deals with data not only in its spatial aspect, but also throughout its temporal dimension (Solomonides, 2019), and this requirement implies an increase in the degree of complexity and scope of governance commitments.

This structuring framework is necessary since digital research data can only be properly managed and preserved over time through a sustainable institutional commitment (Mayernik et al., 2012, p. 1). To a certain extent, the consolidation of data management services reflects the level of organizational acceptance incorporated into them and the degree of planning of the various actions required: sustainable budget in place, appropriate data policy, organic connection with target communities, compliance with ethical and legal codes, alignment with institutional strategic objectives and a development strategy that considers the possible paths for each institution. It is also necessary to consider the inevitability of the fact that the technological infrastructures for accessing, interpreting and preserving digital information are continually evolving; anticipating these problems and developing strategies to mitigate them is a relevant activity for governance commitments (National Research Council, 2015). On these pillars, advanced data services can be developed that can appropriately support the entire lifecycle of these information assets, according to the interests of the various stakeholders involved. Figure 1 below shows the pillars of data governance that will underpin the entire proposed model. This is followed by a description of each of the items that represent the pillars.

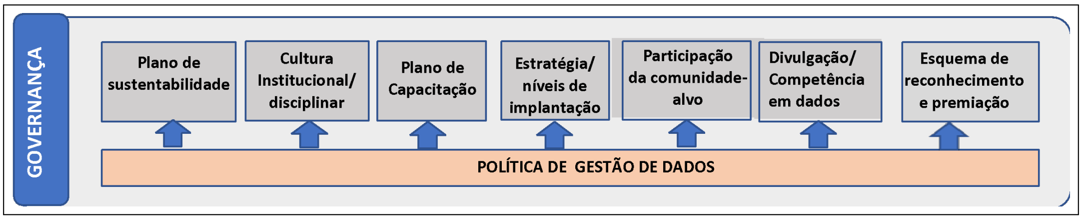

Figure 1. Pillars of Data Governance.

Institution's Data Management Policy – Establishes the institution's foundations, guidelines, and commitments concerning management, use, ownership, compliance with ethical and legal codes, adherence to the policies of funding agencies, national science, technology and innovation policies, international guidelines and practices, and lastly, but critically, the culture, practices and idiosyncrasies of disciplinary communities and domains. A comprehensive research data management policy should also identify the responsibilities of each of the actors—library, laboratories, information technology, administration, etc.—since data management involves different sectors of the institution (Mushi, Pienaar, & van Deventer, 2020) and the project is considered part of the institution's research activities. It is necessary to emphasize that the process of developing an institutional data management policy requires extensive consultation with all the agents involved and the approval of the relevant scientific communities and organizations (Wilson, Martinez-Uribe, Fraser, & Jeffreys, 2011). Policy guidelines should permeate the entire management cycle. "Policies can be an important motivating factor for FAIR data and other research objects (software, workflow, models, protocols, etc.). It is therefore essential that community-based bottom-up efforts are combined with top-down policies," add the European Commission (2020).

Institutional/disciplinary culture – The deployment of a research data management services platform should be preceded by a analysis of requirements that considers the institutional, community and disciplinary context and culture, as well as their unique characteristics. It is hoped that this process will help define a more effective portfolio of data management services to support the research practices of the institution and its communities (Coates, 2014, Reed, 2015, Mushi et al., 2020). It is also important to recognize that some disciplines require different types of technical solutions to obtain the same benefits from FAIR data (European Commission, 2020).

Sustainability plan – One of the major challenges of a data management infrastructure implementation and maintenance program is to ensure that each phase of the project is sustainable as an ongoing service over time (Wilson et al., 2011). Once research data management is recognized as something necessary for research activities, its costs must be estimated and its funding sources—especially perennial ones—identified. Choosing where to invest the typically limited resources and projecting future needs become essential (Choudhury et al., 2018). Thus, a project to implement research data management services needs to be associated with a sustainability plan that outlines a possible commitment to the now and the future. Creating and committing to a long-term strategy for the services can reveal more clearly the resources needed for the continuity of the services and the infrastructures required for this. This can therefore include a succession plan (Mushi et al., 2020).

Data dissemination/competence – In order to implement a FAIR research environment, it is necessary for the communities involved to develop a shared understanding of what is circumscribed by the FAIR concept and its principles. In general, researchers and other stakeholders have a low level of awareness of the importance of data management practices and the management and sharing requirements of funding agencies and the data deposit commitments signed with scientific publishers, as well as the ethical and legal issues involved in publishing data. For example, in relation to the FAIR concept, the European Commission (2020) notes that researchers do not know what FAIR data is and often think it is the same as open data. This indicates that planning and dissemination and awareness-raising actions are needed to bring these issues to light. A dissemination program in this direction should include the development of educational material (booklets and guides), courses, events, workshops, among others.

Knowledge/participation of the target community – As creators and users of research data, the engagement of researchers is crucial in the development of data management services. The provisioning of any service needs to be based on a close understanding of the patterns and flows of the research taking place in the institution, its motivations, characteristics, and priorities. Therefore, the precise definition of service requirements needs to be established with the commitment and input of the research community; without these considerations, the characteristics of the services may not be in harmony with the objectives of the researchers. The community must be accompanied in the changes of interest in the data, and its participation in the development and choice of shareable standards for FAIR practices and infrastructures must be recognized and institutionalized. The proximity, interaction, and alignment of communities with national and international organizations that deal directly with FAIR data management, such as GO FAIR, RDA, CODATA, DCC and others, should be encouraged.

Training plan – In order to offer comprehensive data management services, libraries need to have technologically qualified staff or greatly increase technological training opportunities for existing staff. (Tenopir, Birch, & Allard, 2012) Human sustainability is a critical condition to ensure the continuity and consistency of service provision over time. However, few formal programs in information studies include data management in their curricula, so research data managers are usually trained on the job in the specific disciplines where they work (Borgman, 2007, p. 155).

Deployment strategy/levels – The development and deployment of data management infrastructure, in addition to being resource-intensive, requires time to reach full maturity and reflect the demands of scientific communities. This implies the need to establish deployment levels for infrastructures and services. Research libraries, for example, have often proactively sought to meet the data management needs of their user communities. Often this happens without additional financial support for the development and provision of data services. Therefore, libraries have to start on a simpler scale, building a foundation on which to develop more sophisticated services (Erway, Horton, Nurnberger, Otsuji, & Rushing, 2016), starting with basic services that only require resources from the library itself, until they reach more complex services that require a high level of institutional commitment and more financial, technological and human resources Kouper et al. (2017).

Recognition and awards – Managing research data takes time, resources and requires a great deal of dedication from the researcher. However, this effort is rarely recognized by academic reward systems, except when linked to publications in scientific journals. Therefore, in order to encourage this new task for researchers and highlight its importance, it is essential that it is properly recognized and that it is taken into account in evaluation, promotion, and funding criteria.

Research Data Infrastructures

Infrastructure is a broad and multidimensional concept. It can have a technical, legal and organizational connotation and, in many cases, it is also essential to consider social, cultural and political aspects. In fact, this is the case in the field of science: the design of research infrastructure is simultaneously a technological question, a question of identifying research needs in specific disciplinary areas and a political question. This more general perspective applies to institutional research data management infrastructures that need to offer technologies and tools, processes, policies, resources, and training for the various and diverse stages of data management.

Just as institutions must provide basic infrastructures for research—such as laboratories, instrumentation, high-performance computing, networks, reagents and much more—they must also take measures for proper data management. This presupposes a broad spectrum of managerial, technological and informational activities, including information professionals trained to support researchers in planning and managing their data, access to secure storage devices and backups during project development, and the availability of long-term access and preservation platforms needed after the end of the research (Strasser, 2015). It is also essential to have a body of norms, standards and good practices that, above all, allow systems and services to interact at different levels, both locally and globally, which can be translated as interoperability. In this category, by way of example, there are **data model standards**—generally established by a disciplinary domain or repository—which determine the structure of the various components of a data collection, which ultimately have an effect on the interaction interfaces with human and computer users and on the levels of interoperability of the dataset (Choudhury et al., 2018).

When we compare traditional academic publishing with data publishing, we see that the infrastructures underlying academic publishing create an epistemological bridge between disciplines, with the aggregating point being research libraries that select, collect, organize and make accessible publications of all kinds and from all fields. By their nature, social institutions work to stabilize particular practices and forms of knowledge. In a sense, institutions are social infrastructures in themselves. In this sense, technical infrastructures are intertwined with the social infrastructures of institutions, often mediated by standards, protocols, documents, and artifacts that link the social and technical aspects of infrastructures (Leonardi, 2010). However, there is still no infrastructure of this magnitude for data. A few areas have consolidated mechanisms for publishing data; others are in the stages of developing standards and practices for aggregating their data and making it widely accessible. A key problem in research institutions, as noted by Mayernik et al. (2012, p. 158), "is the lack of a reliable infrastructure that can be deployed at an institutional level", this "lack of infrastructure for data amplifies the discontinuity in academic publishing", adds Borgman (2007, p. 155).

The infrastructural frameworks for data management are diverse and fragmented in terms of flows, complexity, application and topology, and organized differently by different disciplines and in different countries (Graaf & Waaijers, 2011). However, infrastructures are increasingly shaping data management standards and practices. Given this fact, knowledge about the origin, disciplinary domain, degree of processing, collection systems, workflows, etc. seems to be of critical importance when designing infrastructures for data management (Sayão & Sales, 2020). Figure 2 below shows the part of the model that represents the infrastructure needed for research data management, with each item explained below.

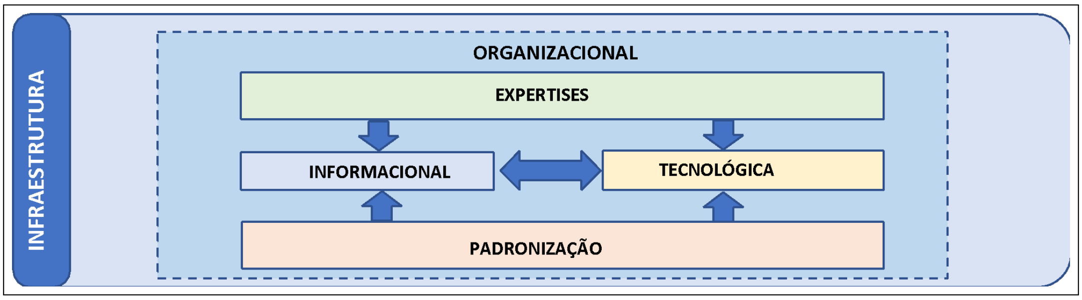

Figure 2. Infrastructure needed to implement Data Management Services.

In this model proposal, we consider five instances of infrastructure necessary for implementing data management systems: **standardization**—in which norms and standards are presented as consensual ways of codifying the knowledge that circulates across communities, ensuring uniformity and similarity to their products and processes across time and space; **technological**—which comprises a vast set of activities, equipment, processes, and expertise that can enable the operational technological requirements necessary for data management cyberinfrastructures; **informational**, which comprises the entire conceptual and theoretical framework embodied in the practices of Information Science, Librarianship and Archivology, which are fully applied to data management, such as selection, cataloging, indexing, classification and disposal, and the instruments that enable these practices; **professional**, which comprises support teams made up of different professionals, such as data administrators and data scientists, data librarians, data archivists, data informaticists, data curators, among others; **organizational**, which presupposes a research-oriented organizational structure, such as a university, research institute, or even a company whose business depends on data management.

These infrastructural aspects—which enable the interweaving of knowledge and practices that underlie equipment, facilities, methodologies and, above all, people—provide a vast portfolio of services, tools, and processes that continually bring research objects into line with the FAIR Principles.

What we have seen so far is that, in order to reach the category that represents research data management services, it is necessary to establish efficient data governance and build an adequate infrastructure for the formulation of services. The following section therefore describes the category that is implicitly at the heart of the model.

Research Data Management Services

In constructing this model, we considered a matrix of services based on two main axes: a temporal axis, which considers the development of data services over time, linking the data life cycle to the research life cycle; the second axis considers the anchor point of the services, meaning that they can be based on informational, computational, scientific or administrative processes. It is clear that these contours are not always well-defined and overlaps are present in both axes, which demonstrates the need for interconnection of various expertise in order to carry out research data management activities. From a temporal point of view, we can consider that management in the form of services takes place at three points in time: before the research begins, during the research and after the research has finished (Jones et al., 2013).

To meet the broad spectrum of data management needs, materialized in specific services, requires collaboration between different areas and the integration of expertise, infrastructures, practices and methodologies. While identifying that there are important—and desirable—overlaps, we have considered four types of services: scientific services, computational services, information services and administrative services (Figure 3).

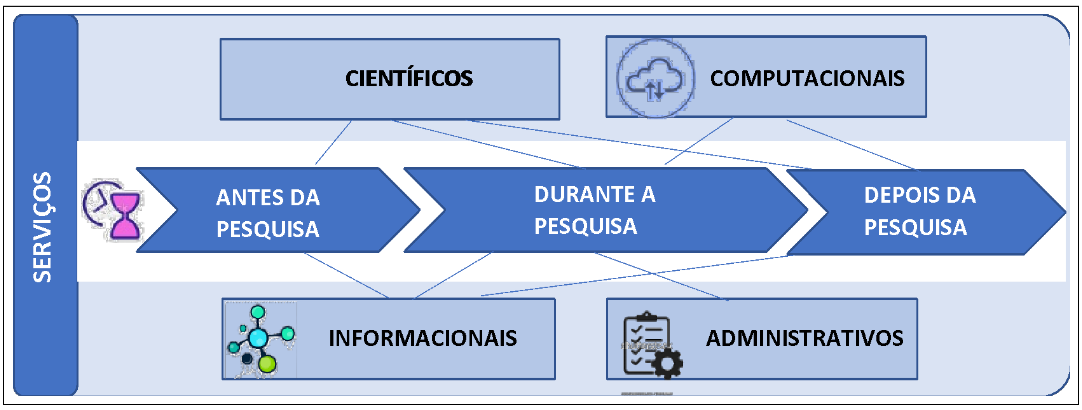

Figure 3. Research Data Management Services.

Scientific services comprise services that take place in predominantly scientific environments, such as laboratories and research centers, and are carried out by scientists, academics, or data management specialists with in-depth disciplinary knowledge, and can be: disciplinary description; appraisal of data collections; data cleaning; data organization; transformation; processing documentation; annotation; code documentation; workflow documentation; data analysis; data presentation and visualization; and data packaging.

Computing services are those that comprise the provision of software tools and computing equipment to support the processing, analysis and visualization of research data; support the processes of interoperability and machine activation of data and metadata; provide guidance on how data can best be structured and stored and work, if necessary, with researchers on structuring databases and text markup (Wilson et al., 2011); the services can also include specific training for the research team on the resources offered and, in more advanced situations, offer high-performance processing, high-volume cloud storage and grid computing. Examples of computing services are: storage systems; data protection; format standardization; backup services; and support for the safe disposal of data.

Information services are those offered by libraries and carried out with the support of professional librarians and archivists. Examples of information services that can be offered by scientific and university libraries and other research information centers are: scientific data portal; data reference desk; support in drawing up the data management plan; support in discovering and accessing the data collection; development of data collections; development of metadata; creation of standardized references; identification of data and researchers; cataloging/indexing of data collections; long-term archiving/preservation; publication of data; contextualization/linking; training for researchers.

Finally, **administration services** are those services that provide support, sustainability, and visibility to the others. They include guidance services on costs, budgeting, acquisition of data collections, ethical and legal compliance of data—especially sensitive data—with institutional, national and international rules and regulations; statistics on data use and reuse; this category also involves issues of intellectual property, licenses, and embargo time.

The idea of making data adhere to the FAIR Principles—expressed by the term "FAIRification"—does not happen on its own. This requires a multidimensional data management process, embodied in a list of services, which can effectively add value to research objects over time. This is the last layer of the proposed architecture, which will be presented below.

FAIRification of data

What can be seen is that the level of compliance of the various digital research objects with the FAIR Principles is linked to the scope and depth of the management to which they are subjected. As we have seen, this presupposes the need for a multi-layered framework—scientific, technological, informational and administrative—that addresses the multiple problems that stand between reuse, integrity, provenance, reproducibility, accountability, as well as the new needs and opportunities for large-scale analysis and re-analysis required for eScience flows (Wilkinson et al., 2016).

Therefore, the degree of adherence to the FAIR Principles—the FAIRification (Mons et al., 2017)—highlights a set of services, procedures, and tools which, even though this parameter is not the final objective, places it as part of a renewed scale for assessing the maturity of research data management systems. Figure 4 below shows what the top of the model looks like, indicating the target of any scientific data management service, which should be alignment with the 15 principles that are distributed among the four FAIR categories.

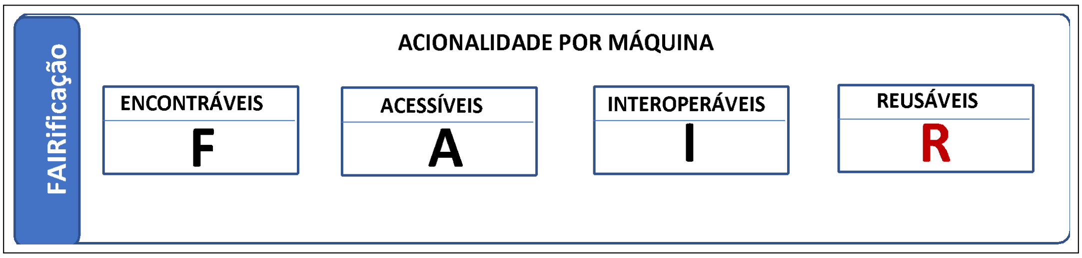

Figure 4. FAIRification of services as a Data Management target.

However, the alignment and implementation of the FAIR Principles in a research institution requires financial investments, cultural changes, training, and the construction of an appropriate technical infrastructure (Graaf & Waaijers, 2011). These factors can be coalesced around the concept of a "research data management platform", the implementation of which has the potential to operationalize the various layers of management and establish a growing infrastructure of informational, scientific, computational and administrative services that enables a scale of adherence to the FAIR principles of research objects, whether they are data themselves or algorithms, codes, workflows or other physical or conceptual devices that lead to data, as well as metadata and infrastructures for data management.

The architecture represented in figure 5 below is proposed based on a combination of the semantic models of aggregation, association, and classification - in which it presents the general categories of the semantic model of aggregation from a bottom-up perspective, namely: governance, infrastructure, services and FAIRification. These categories are associated based on a viability relationship, i.e., one component makes the other viable. Within the governance category, its components are associated based on a steering relationship, showing how one main element can direct the other components, using a management policy as a structuring parameter. The infrastructure category reveals the aggregation of components needed to build the platform's infrastructure, which needs to be permeated by an organizational infrastructure, namely: expertise, technologies, information resources and norms and standards. In this category, once again from a bottom-up perspective, we see norms and standards underpinning the technological and informational instances, which relate to expertise from a top-down perspective, showing who operates them. In the Services category, the model used to build the architecture is one of aggregation, but its components are related in a timeline, i.e., the scientific, computational, informational and administrative services are found in the timeline that surrounds the entire research process. Finally, the FAIRification category, also in a semantic model of aggregation, shows the categories in which the FAIR principles are organized as components necessary to promote FAIRification and, at a higher level, an ecosystem of FAIR data and services.

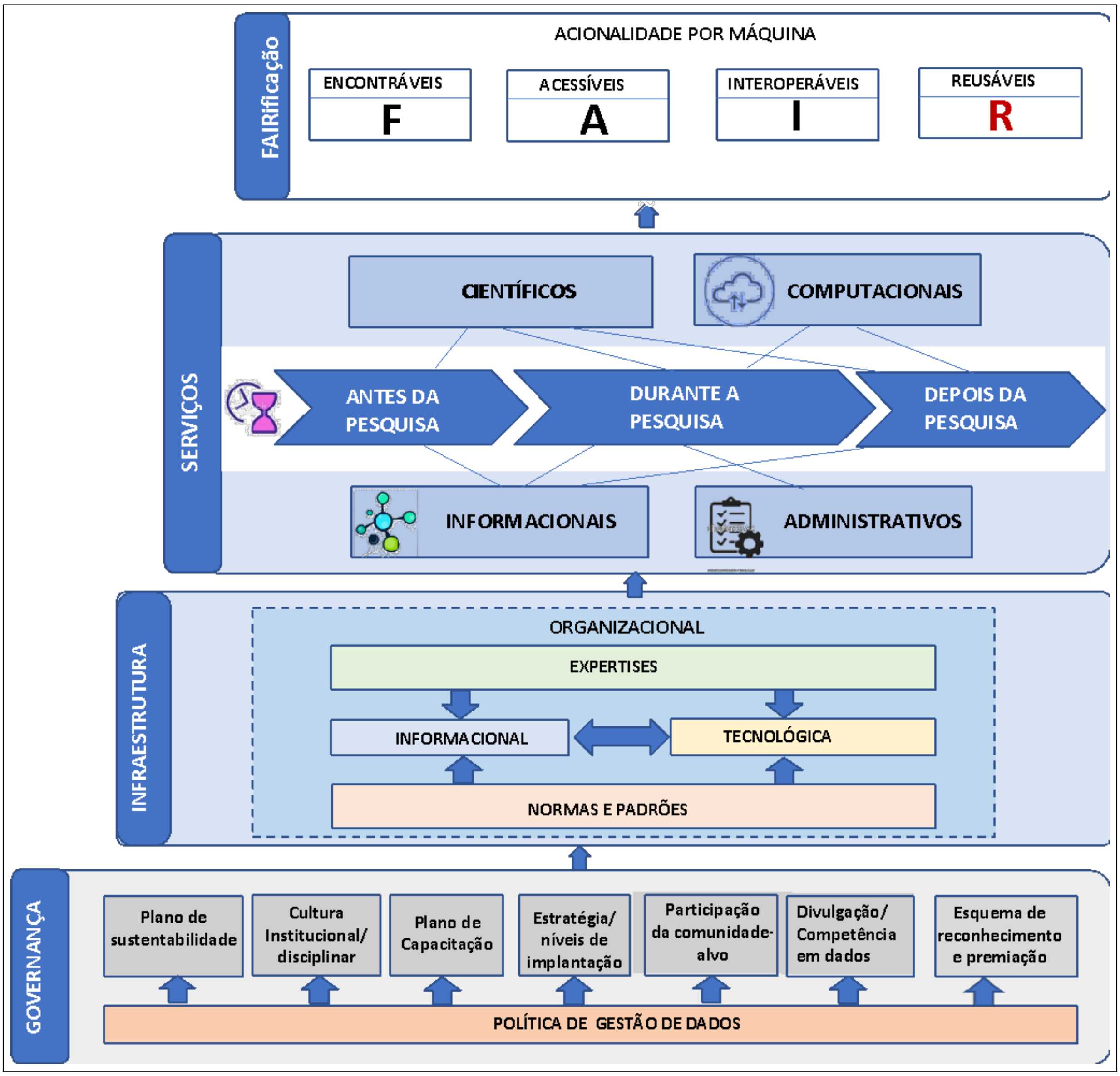

Figure 5. FAIR Data Management Service Model.

IN CONCLUSION

Figure 5 provides an overview of the components of the proposed data management service model, which have been grouped into layers, showing their interrelationships, based on the discussion in the text.

From this perspective, the proposed model sought to deconstruct the blocks that make up a generic architecture for a data management service platform, articulating the various conceptual modules—guidelines, policies, services, tools, infrastructures, etc.—in the form of pieces that can be adjusted according to the depth, scope, and philosophy of each institution or discipline. In order to organize these sometimes entropic and diffuse efforts, the application of FAIR, scaled by its 15 guiding principles, points to a horizon for the efficient construction of data management platform architectures.

Even taking into account the generalist approach of the model, it must be considered that when implementing data management practices and infrastructures, the specific context of scientific communities and the possibilities of adoption must be taken into account. The importance of each service may depend on the priorities and the generation and use of certain research objects.

With this in mind, this model can serve as a guide to be used to create a possible scale for measuring the level of maturity of management service projects, given that the proposed architecture aims to make data adhere to the FAIR Principles, making the idea of an Internet of FAIR Data and Services more concrete.

A model can never be an end in itself. Models must be applied, evaluated and improved. New knowledge obtained after the development of this research on other works related to the topic of data management services can be considered in a possible continuation of studies aimed at improving this proposal, as is the case with the DIAMANT Model (Design and Information Architecture for Data MANagement Technologies), which, as Marin-Arraiza and Vidotti (2019, p. 268) explain, "links the services of different sections within an institution with the phases of the research project and the consequent need for services for researchers in each phase". This article is the result of ongoing research carried out within the BRIET research group (Librarianship, Representation, Interoperability, E-science and Technology), and is the result of the research project "FAIR research data management: a proposed model for accelerating scientific research in the state of Rio de Janeiro" funded by FAPERJ and the development project "Support Infrastructure for the Management and Preservation of Brazilian Nuclear Knowledge", funded by CNEN and CNPq.

AGRADECIMENTOS

Ao CNPq e à FAPERJ pelo financiamento dessa pesquisa. À Teodora Marly Gama pela revisão, sugestões e preciosa contribuição.

REFERENCES

Borgman, C. L. (2007). Data: the input and output of scholarship. In *Scholarship in the digital age: Information, infrastructure, and the internet.* London: The MIT Press. doi: 10.7551/mitpress/7434.003.0009

Choudhury, S., Cowles, E., Croft, H. R., Estlund, K., Fary, M., Faustino, G., . . . Waters, N. (2018). *Research data curation: a framework for an institution-wide services approach.* EDUCAUSE Working Group on Data Curation. Retrieved from https://hsrc.himmelfarb.gwu.edu/libfacpubs/35/

Coates, H. L. (2014). Building data services from the ground up: Strategies and resources. *Journal of eScience Librarianship*, *3*(1), 52–59. doi: 10.7191/jeslib.2014.1063

Erway, R., Horton, L., Nurnberger, A., Otsuji, R., & Rushing, A. (2016). *Building blocks: laying the foundation for a research data management program.* Dublin: OCLC. Retrieved from https://files.eric.ed.gov/fulltext/ED589141.pdf

European Commission. (2020). *Six recommendations to implementation of fair practices.* Bruxelas: European Commission. doi: 10.2777/986252

Gil, A. C. (2002). *Como elaborar projetos de pesquisa* (4th ed.). São Paulo: Atlas.

Graaf, M. v. d., & Waaijers, L. (2011). *A surfboard for riding the wave: towards a four country action programme on research data.* Copenhagen: Knowledge Exchange. Retrieved from http://hdl.voced.edu.au/10707/183860

Jones, S., Prior, G., & White, A. (2013). *How to develop research data management services – a guide for heis.* Edinburgh: Digital Curation Centre. Retrieved from https://www.dcc.ac.uk/guidance/how-guides/how-develop-rdm-services

Kouper, I., Fear, K., Ishida, M., Kollen, C., & Williams, S. C. (2017). Research data services maturity in academic libraries. In *Curating research data: practical strategies for your digital repository* (Vol. 1, pp. 153–170). Chicago: Association of College and Research Libraries. Retrieved from https://experts.illinois.edu/en/publications/research-data-services-maturity-in-academic-libraries

Leonardi, P. M. (2010, June). Digital materiality? how artifacts without matter, matter. *First Monday*, *15*, 6–7. doi: 10.5210/fm.v15i6.3036

Mayernik, M. S., Choudhury, G. S., DiLauro, T., Metsger, E., Pralle, B., Rippin, M., & Duerr, R. (2012). The data conservancy instance: Infrastructure and organizational services for research data curation. *D-Lib Magazine*, *18*, 9–10. doi: 10.1045/september2012-mayernik

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud. *Information Services & Use*, *37*(1), 49–56. doi: 10.3233/isu-170824

Mushi, G. E., Pienaar, H., & van Deventer, M. (2020). Identifying and implementing relevant research data management services for the library at the university of dodoma, tanzania. *Data Science Journal*, *19*(1), 1–9. doi: 10.5334/dsj-2020-001

National Research Council. (2015). *Preparing the workforce for digital curation.* Washington: The National Academies Press. doi: 10.17226/18590

Reed, R. (2015). Diving into data: Planning a research data management event. *Journal of eScience Librarianship*, *4*(1). doi: 10.7191/jeslib.2015.1071

Sayão, L. F. (2001). Modelos teóricos em ciência da informação - abstração e método científico. *Ciência da Informação*, *30*(1), 82—91. doi: 10.1590/s0100-19652001000100010

Sayão, L. F., & Sales, L. F. (2020). Afinal, o que é dado de pesquisa? *Biblos*, *34*(2). Retrieved from https://www.seer.furg.br/biblos/article/view/11875

Solomonides, A. (2019). Research data governance, roles, and infrastructure. In *Clinical research informatics* (p. 291—310). Springer International Publishing. doi: 10.1007/978-3-319-98779-8_14

Strasser, C. (2015). *Research data management.* Baltimore: NISO. Retrieved from https://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf

Tang, R., & Hu, Z. (2019). Providing research data management (rdm) services in libraries: Preparedness, roles, challenges, and training for rdm practice. *Data and Information Management*, *3*(2), 84—101. doi: 10.2478/dim-2019-0009

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future.* Chicago: Association of College and Research Libraries. Retrieved from https://trace.tennessee.edu/utk_dataone/20/

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. doi: 10.1038/sdata.2016.18

Wilson, J. A. J., Martinez-Uribe, L., Fraser, M. A., & Jeffreys, P. (2011). An institutional approach to developing research data management infrastructure. *International Journal of Digital Curation*, *6*(2), 274–287. doi: 10.2218/ijdc.v6i2.203

How to cite this article (APA):
Sales, L. F. & Sayão, L. F. (2022). Proposal for a model of research data management service. *AtoZ: novas práticas em informação e conhecimento, 11*, 1 – 13. Retrieved from: http://dx.doi.org/10.5380/atoz.v11.85765

NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	**Luana Farias Sales**	**Luís Fernando Sayão**
Concepção do manuscrito		X
Escrita do manuscrito	X	X
Metodologia	X	
Curadoria dos dados	X	
Discussão dos resultados	X	X
Análise dos dados	X	

FINANCIAMENTO

O(s) autor(es) declara(m) que esta pesquisa recebeu financiamento conforme dados indicados a seguir e o documento comprobatório foi anexado como documento suplementar: **Chamada: Universal 2018 – Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – nº 430553/2018-8**

O(s) autor(es) concorda(m) em interagir diretamente com pareceristas responsáveis pela avaliação do manuscrito, dessa forma tornando a revisão por pares aberta.

Disponibilidade de Dados Científicos da Pesquisa

Os conteúdos subjacentes ao texto da pesquisa já estão disponíveis.