





Research Data Sharing: Framework of Factors that Can Influence Researchers

Elizabete Cristina de Souza de Aguiar Monteiro^(✉) 
and Ricardo César Gonçalves Sant'Ana 

São Paulo State University, Av. Hygino Muzzi Filho, 737, Marília 17525-900, Brazil
ricardo.santana@unesp.br
<http://www.marilia.unesp.br>

Abstract. Health emergencies contribute to a greater sharing of research data among the scientific community, however, there are other factors that can influence researchers to share or retain their data. The objective is to identify the factors that can influence the perception and attitude of researchers in sharing data sets. The specific objective was to present a framework with the identified factors. Documentary and exploratory research using the method of content analysis and application of the institutional theory, the theory of planned behavior and the research model for data sharing behaviors. The data revealed that indicators of the Cognitive, Normative, Career, Resources and Social pillars influence the perception and attitude of sharing data. It is hoped that the results will allow a perception of the factors that influence researchers to share their data.

Keywords: Research data sharing · Data management · Open Science

1 Introduction

The pandemic caused by Covid-19 motivated the collaboration between agents of the scientific community to share data and research results for the diagnosis, treatment, and development of vaccines with the efficiency that the situation requires. These resources were published in preprints, journals, institutional pages, and open access data repositories.

In the health area, two dimensions of interest can be perceived: social and commercial. Health is a sensitive and strategic area for society living with intense market potential, profitable and competitive, supported by a structure marked by patrimonial law (patents; publication in high impact journals) as well as by moral law (quotation, recognition, and prestige among peers) [1, 2].

In this scenario, premises such as open science and open access to research data contributed to the acceleration of discussions and collaborations by the scientific community. Open Science is an international movement of scientific practice based on open publication, open metadata, open data, open source, open educational resources, open peer review, impact and open metrics, open repositories, including data repositories,

under FAIR Principles (Findable, Accessible, Interoperable, Reusable). This movement provides the enhancement of the reproducibility of science, greater transparency of public funding, reuse of data in line with society's needs, considering the ethical and legal aspects [1, 3–5].

Open Science seeks to promote the transformation, openness and democratization of science and research. It is a complex concept that requires the adoption of new practices, data sharing and research results, which requires a change in the culture of the actors involved and the behavior of researchers in relation to the development of research, thus enabling the expansion of social and innovation impact [6, 7].

The presented context reinforces the relevance of the collaboration between agents of the scientific community, but also the conflicts and benefits that can emerge with the sharing of data and research results.

It is necessary to identify the different factors that can influence researchers to share or retain their data sets, which include, but are not limited to, institutional, cultural, ethical, social and personal. Data sharing contributes to increasing the transparency of research and, consequently, the credibility of its results, as well as maximizing the replication of data, accelerating scientific collaboration, testing different hypotheses, reducing costs, increasing citations, among other benefits. Once the data is retained, the benefits are not enjoyed.

The general objective of this article was to identify factors that can influence the perception and attitude of researchers to the act of sharing data sets, also proposing a framework with the identified factors.

To achieve the objective and support this discussion, the e-cienciaDatos repository was analyzed. It is a data repository resulting from the cooperation agreement between libraries in the Community of Madrid composed of six universities: Universidad de Alcalá, Universidad Autónoma de Madrid, Universidad Carlos III de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos and UNED, members of the Madroño Consortium.

This research presents an initial view of the institutional, cultural and individual factors that can influence the researchers' perception of data sharing and their attitude between sharing or retaining their data sets.

From the results of this study, it is expected to contribute to the enrichment of theoretical discussions about factors such as institutional demands and individual perceptions that can instigate or inhibit the attitude of sharing data sets. The results can also be considered in the training of researchers and in the dissemination of products and services on data management.

2 Methodological Procedures

This article uses documentary research of a descriptive nature, in which laws, regulations, manuals, guidelines, audiovisuals, infographics, statements and reports made available by the Madroño Consortium and its libraries, as well as questionnaires and interviews applied to four librarians who work in the repositories. The questionnaire was composed of 30 questions, 20 of which were closed and 10 open, the last ones following as a script for the interview. The questions addressed aspects of the activities performed with data

management and its life cycle, as well as the competencies and skills required to work in the repositories, both in technical services and in user support and training. The documentary analysis was performed between the months of January to May 2020.

With the application of the Content Analysis method and the Categorical Analysis technique, six categories were defined: Cognitive, Normative, Regulatory, Career, Resources and Social. Then, for each category, thirteen indicators were identified that made up a framework [8–10]. The definition of the indicators was defined a posteriori. In order to have reliable categories and indicators for measuring and benchmarking, the search for them was conducted on themes, concepts and characteristics of the research universes. The selection and delimitation of the indicators and categories was supported by the Institutional Theory [11] Theory of Planned Behavior [12] and the Model for data sharing behavior.

3 Framework for Data Sharing Perspectives and Attitudes

Institutional Theory comes from sociological and organizational studies and helps to connect different factors intertwined with institutions, infrastructure and people, in addition to pointing out how actors perform acceptable behaviors to be legitimized in the face of institutional pressures [11, 13].

The Institutional Theory contemplates three pillars [11] that served to define the categories and organize the indicators selected for analysis. The three pillars are:

- Regulator: establishes rules, inspects them and establishes rewards or punishments in an attempt to influence future behaviors.
- Normative: introduction of the prescriptive, evaluative and mandatory dimension in social life, specify how its basic elements should be realized, defining legitimate meanings to the values and rules adopted.
- Cognitive: presence and interaction between the actors, internalized understanding of each actor based on his interpretation of the social reality in which he participates together with the active cultural system.

The Theory of Planned Behavior (TPB) is “[...] a theory designed to predict and explain human behavior in specific contexts [...]” pointing as a factor to center that there is “[...] the intention of the individual to perform a certain behavior.” [12].

The Theory of Planned Behavior presents a clear relationship of how an individual's attitudes, subjective norms and perceived behavioral controls influence his behaviors, mediated by behavioral intentions [12, 14]. A central factor in the Theory of Planned Behavior is the individual's intention to perform a certain behavior [12]. It is noteworthy that the performance of a behavior is a joint function of intentions and perceived behavioral control.

The Data Sharing Behaviors research model for scientists is a model proposed by Kim (2017) which, according to the author, presents an extensive map of data sharing behaviors of scientists [13]. To support his model, the author used as a basis Scott's Institutional Theory and Ajzen's Theory of Planned Behavior. The model explains how scientists decide to share their data and how the factors that involve sharing differ in different forms of behavior.

The elaboration of the framework presented in this research results from the contribution of the combination of these three theories.

4 Results and Discussions

Data analysis, using the Categorical Analysis of the Content Analysis method and based on the theories presented, resulted in the identification of six categories: Cognitive, Normative, Regulatory, Career, Resources and Social, and 13 indicators: Cultural, Data Management Plan (DMP), Open Science, FAIR Principles, Funding agencies, Repository, Rules and Laws, Benefits, Risks, Efforts, Support from librarians, Benefits to society and Ethics, which made up the framework presented in Fig. 1.

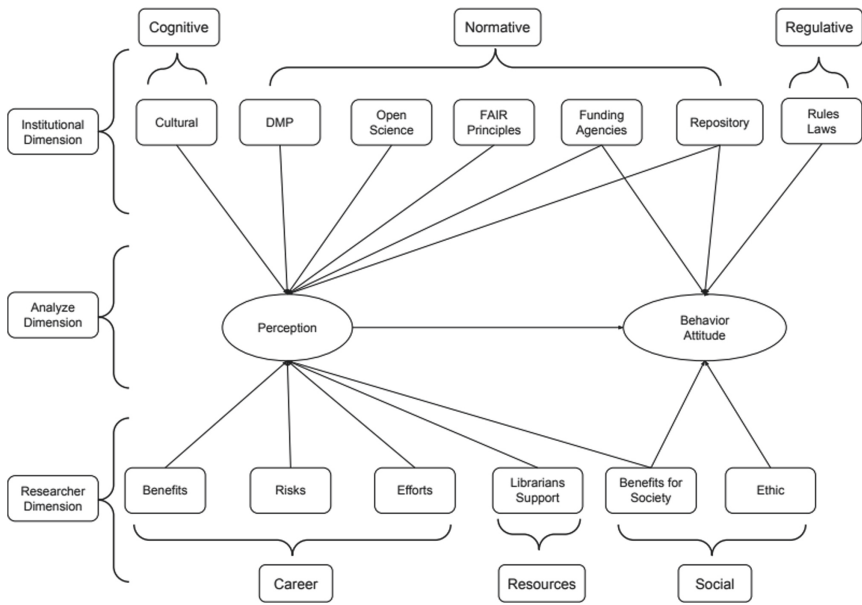


Fig. 1. Data sharing perspective and attitude framework

As illustrated in Fig. 1, the framework is made up of three dimensions: Institutional Dimension; Analysis Dimension; Researcher Dimension.

The Institutional Dimension is based on Scott’s Institutional Theory (1995) being composed by the Cognitive, Normative and Regulatory categories, forming the pillars of this dimension and include the indicators:

- Cultural: points out the relationship and interaction between the actors at institutional, national and international level, highlighting the organizational system and social life with the rights, rules and responsibilities;
- Data Management Plan (DMP): document requested by funding agencies and which articulates aspects of data management by researchers;

- Open science: encompasses open scientific practices and changing the behavior of researchers related, among other aspects, to open data sharing;
- FAIR Principles: guiding principles in which they establish and regulate the procedures so that the meta (data) are Findable, Accessible, Interoperable and Reusable (FAIR), in addition to requiring changes in terms of culture and research practice;
- Development Agencies: their guidelines are being articulated according to institutional and social needs, involving the establishment of rules and how they should be applied;
- Repository: formulates policies and guidelines related to data management, conducting data sharing behavior and articulating the demand for other elements of the proposed theoretical model;
- Laws and Norms: related to the researchers' understanding of the legislation and regulations of international management spheres, their country, the university and the repository.

The focus of the analysis of the framework is the Analysis Dimension, where the researcher is. The Analysis Dimension is based on Ajzen's Theory of Planned Behavior (1991) with the perception and attitude of sharing data from researchers in which they are structured according to their context of performance, interweaving all 13 identified indicators [12]. It has two dimensions of analysis:

- Perception: corresponds to the resources and opportunities available to the researcher that impact his intentions and actions and will dictate the probability of sharing the data sets; researchers' perception of the ease or difficulty in carrying out the behavior of interest;
- Data sharing attitude: corresponds to behavior, the act of sharing your data sets. After assessing and perceiving the advantages and disadvantages of the scenario and the social and institutional pressures, the researcher decides between the available alternatives.

The Researcher Dimension is based on Kim's Data Sharing Behavior model (2017) [13] with the Career, Resources and Social categories composed of the indicators:

- Career Benefits: are the benefits perceived and expected from the sharing of data resulting from the researcher's work;
- Career Risks: are the risks perceived in the process of sharing your data;
- Career efforts: are the activities, time and energy spent on managing the data and preparing it for sharing;
- Support from Librarians: corresponds to support from librarians in data management, working with repositories, training researchers and guiding researchers on their data needs;
- Benefits for Society: it is in line with the needs of making data available for the agility to support a certain demand from society, speed in situational analysis and decision making. It includes the scientific community as it contributes to maximize the reuse of data;

- Ethics: indicates the set of values of an area of knowledge and its appropriate behavior or not along with the duties, principles and norms that are compatible with the groups of researchers in their context.

The results show that the framework is suitable for explaining the data sharing behavior of researchers in various fields of knowledge. The framework is going to be validated with application to university researchers. It can serve as an instrumental tool to quantitatively measure effects in specific researcher populations or environments.

5 Final Considerations

This article presents theoretical and practical contributions to the study of institutional and personal factors that can influence researchers in sharing their data sets. A framework of factors that could potentially influence the attitude of researchers in sharing or retaining their data sets was developed and presented. The composition of these factors considered elements such as: career, resources, cognitive perceptions, normative and regulatory issues, and social aspects.

The presented framework provides a scenario of basic factors of probable influences that can instigate or inhibit researchers of the various areas of knowledge. Therefore, this study aims to contribute to the direction of future studies on various aspects of data management.

The perception and sharing attitude of researchers may change depending on their region of operation, policies, institutional norms and characteristics of their community. It is likely that studies of the same nature will find different indicators, highlighting the intrinsic particularities of each social context.

References

1. Monteiro, E.C.S.A., Sant'ana, R.C.G., Hernández-Pérez: Direitos autorais de dados científicos no contexto da Ciência Aberta: estudo do repositório de dados do Consórcio Madroño. In: 20th Encontro Nacional de Pesquisa em Ciência da Informação, pp. 1–30, UFSC, Florianópolis (2019). <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/1128/756>. Accessed 10 Jan 2020
2. Jorge, V.D.A.: Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika. Universidade Federal do Rio de Janeiro, Rio de Janeiro (2018). https://www.arca.fiocruz.br/bitstream/icict/32604/2/Tese_VanessaJorge.pdf. Accessed 29 Apr 2020
3. Mora, F. (coord.): Compromisos de las universidades ante la Open Science. Madrid: CRUE Universidades Españolas (2019). https://www.crue.org/Documentos%20compartidos/Informes%20y%20Posicionamientos/2019.02.20-Compromisos%20CRUE_OPENSCIENCE%20VF.pdf. Accessed 21 Sept 2020
4. Bezjak, S., et al.: Manual de formação em Ciência Aberta. FOSTER (2018). book.fosteropen-science.eu/pt. Accessed 21 Sept 2020
5. Santos, P.X. (coord.): Livro Verde - Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Fiocruz, Rio de Janeiro (2017). <https://www.arca.fiocruz.br/handle/icict/24117>. Accessed 21 Sept 2020

6. Pontika, N., et al.: Fostering open science to research using a taxonomy and an eLearning portal. In: 15th International Conference on Knowledge Technologies and Data Driven Business, The Open University, Milton Keynes (2015). https://oro.open.ac.uk/44719/2/kmi_foster_iknow.pdf. Accessed 21 Sept 2020
7. Ramjoué, C.: Towards Open Science: the vision of the European Commission. *Inf. Serv. Use* **35**, 167–170 (2015). <https://pdfs.semanticscholar.org/c47af7edc0607c6561390bba2a993c89711297c8.pdf>. Accessed 21 Sept 2020
8. Bardin, L.: *Análise de conteúdo*. Rev. e atual. Lisboa: Edições 70 (2010)
9. Fonseca Júnior, W.C.: *Análise de conteúdo*. In: Duarte, J., Barros, A. (eds.) *Métodos e técnicas de pesquisa em comunicação*, 2nd edn., pp. 280–304. Atlas, São Paulo (2014)
10. Franco, M.L.P.B.: *Análise de conteúdo*. 2nd edn. Brasília, DF, Liber Livros (2005)
11. Scott, W.R.: *Contemporary institutional theory*. In: Scott, W.R. (ed.) *Institutions and Organizations*, Chap. 3, pp. 32–62. SAGE, London, Thousand Oaks (1995)
12. Ajzen, I.: The theory of planned behavior. *Org. Behav. Hum. Decis. Process* **52**(2), 179–211 (1991)
13. Kim, Y.: Fostering scientists' data sharing behaviors via data repositories. *J. Suppl. Pers. Commun. Methods Inf. Process. Manage.* **53**(4), 871–885 (2017). <https://linkinghub.elsevier.com/retrieve/pii/S0306457316305908>. Accessed 21 Sept 2020
14. Kim, Y., Zhang, P.: Understanding data sharing behaviors of STEM researchers: the roles of attitudes, norms, and data repositories. *Libr. Inf. Sci. Res.* **37**(3), 189–200 (2015). <https://linkinghub.elsevier.com/retrieve/pii/S0740818815000584>. Accessed 21 Mar 2020
15. Kim, Y., Adler, M.: Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *Int. J. Inf. Manage.* **35**(4), 408–418 (2015). <https://www.sciencedirect.com/science/article/pii/S0268401215000432>. Accessed 21 Mar 2020