11. DEVELOPMENT AND APPLICATION OF INTEROPERABILITY STANDARDS FOR SCIENTIFIC DATA REPOSITORIES: IBICT AND CNPQ REPOSITORIES

Lucas N. Paganine[140]
Washington L. Ribeiro de Carvalho Segundo[141]
João L. R. Moreira[142]

Abstract: *In the 4th Plan of the Open Government Partnership (OGP), specifically in Commitment 3 (Establish governance mechanisms for scientific data for the advancement of Open Science in Brazil), proposed by the Brazilian Agricultural Research Corporation (Embrapa) and coordinated by the Comptroller General of the Union (CGU), comes the landmark 8, "Proposal of interoperability standards for research data repositories", this coordinated by the Brazilian Institute of Information in Science and Technology (IBICT), which resulted in the document Interoperability standards for data repositories of research search. The document aims to develop a minimum set of metadata descriptions for research data, making extensions to specific domains of knowledge. The methodology comprises two stages: the first characterized by the establishment of a general multi-thematic metadata scheme standard from the OpenAIRE Guidelines v4 and the metadata requirements of the Fair Data Point specification; and the second through the development of extensions by area of knowledge, from the Metadata Directory maintained by the Research Data Alliance (RDA), and also through the records of thematic research data repositories in the Registry of Research Data Repositories (re3data). The final results showed a core of 13 mandatory metadata and 3 extensions for specific areas: Biology, Agriculture and Social Sciences. An analysis of the document in question will be carried out and the application cases of the Aleia and LattesData repositories, under development, will be presented. It is noteworthy then that this is not a job that is exhausted in the final application, on the contrary it needs constant development.*

Keywords: *Research Data repositories. Scientific data. Metadata standards.*

140 Bachelor in Library Science with graduate degree in Public Management, Brazilian Institute of Information on Science and Technology (Ibict) - lucaspaganine@ibict.br

141 PhD in Computer Science, Brazilian Institute of Information on Science and Technology (Ibict) - washingtonsegundo@ibict.br

142 PhD in Computer Science, University of Twente - j.luizrebelomoreira@utwente.nl

11.1 INTRODUCTION

Open Science (OS) is an umbrella term that covers several aspects such as Citizen Science and notebook science sharing. It emerges as the development of Open Access Movement (OA), dealing with the opening of scientific processes to the population in general, thus following collaboration and transparency principles. In OS context, it is emphasized the increasing demand for scientific data sharing, using several tools for this purpose, among which scientific data repositories stand out.

These repositories are tools for treating, organizing, dissemination and preserving digital objects, in this case the scientific data. However, due to the needs of the several areas of knowledge and different institutional realities that implement the repositories, a lot of standards for describing datasets stored emerge.

The importance of a research on these different existing patterns and models is highlighted, aiming at developing a central descriptive scheme that also allows meeting the specific needs of different areas of knowledge, through extensions of this central pattern, thus enabling the interoperability among repositories from different thematic domains.

This study was then developed within the scope of the Open Government Partnership (OGP), an international initiative initiated in 2011 with the purpose of encouraging transparency as a government practice, in particular, the access to public information and the active cooperation with society. Such purpose is very much in line with OS.

OGP acts through National Actions Plans committed to Open Govern practices. As a result of the execution of these plans, reports expressing the progress in meeting the proposed goals are elaborated.

In 2018, Brazil developed its 4th National Action Plan, with 11 commitments, among which the Commitment 3, which aimed at "Stablishing mechanisms for research data governance for the development of Open Science in Brazil" (RNP, 2018). This commitment, known as Commitment for OS, was under Embrapa's coordination, but with the participation of several institutions, most of them governmental.

The commitment was organized in nine landmarks, with Landmark 8 described as a "Proposition of interoperability standards for scientific data repository" (RNP, 2018), coordinated by Ibict, but with the collaboration of the National Education and Research Network, Twente University, National Nuclear Energy Commission and National Council for Scientific and Technological Development (CNPq). One of the results from the Landmark was a Guide with the proposal of "Interoperability Standards for research data repositories" which are applicable to any research data repository that wants to promote the interoperability and opening stored scientific data. Interoperability criteria are defined in the document, guiding the building or improvement of scientific data.

Considering this scenario, it is worth it to deeply explore the product of the landmark in question, in special When considering its purpose of "develop and apply a minimal set of descriptions for scientific data, making appendices for specific knowledge domains, based on existing international standards and guidelines" (Paganine *et al.* , 2020).

Therefore, the methodology used herein will begin with the description of the development of the result in question as well as the presentation of results, and at the end, a description of the stages and application process of this document in Ibict and CNPq scientific data repositories is provided.

The document in Paganine *et al.* (2020) is divided in 2 parts, one with a general metadata set, and the other with appendices for specific areas of knowledge. Documents from well-established international guidelines for scientific data repositories are used as main references. They were: the OpenAIRE Guidelines for Data Repositories, concomitantly with the OpenAIRE Guidelines for Repository of Scientific Publications; and the metadata set described by *Fair Data Point* (FDP) *framework.* OpenAIRE guidelines are extensively adopted internationally, however, research dealing with semantic interoperability indicates the importance of extensions to meet FAIR principles FAIR (Wilkinson *et al.*, 2016). This extension also cited by Santos *et al.* (2016) is explored in the FDP *framework*.

For the classification on the extensions to specific areas of knowledge, for organizational reasons internal to Brazil, the CNPq tables of knowledge areas from Frascati (OECD, 2015) research manual and the area division one used by Data Curation Centre (DCC) and RDA, the *Deutsche Forschungsgemeinschaf* (DFG, 2020). The complete comparative table is found in the original document resulting from the landmark.

Regarding the survey of standards for the extensions for specific knowledge areas, it started with Metadata directory[143], a tool maintained by RDA, followed by subsequent analysis of metadata standards used in thematic repositories (they cover only a certain knowledge area) and institutional and/or multi-thematic data found in the Registry of Research Data Repositories (re3data).

At the end, the standards found in the surveyed repositories were checked, in search of which appendices can complement OpenAIRE added to the FDP standard for the 4 areas initially selected: Biology (due to its behavior in the knowledge trees surveyed), Agriculture (due to the importance of the area nationally), Health (due to the importance and pioneering action in Open Access Movement and other aspects related to the research) and Social Sciences (due to the general object of study of the institute where the research was developed, the Ibict). It is noteworthy that during the development of the proposal, specialists in the area were consulted, especially Fiocruz (Health) and Embrapa (Agriculture).

11.2 DEVELOPMENT

The guidelines presented in Paganine *et al.* (2020), also have as references the old DRIVER guidelines that were published in 2007 by the *Digital Repository Infrastructure Vision for European Research (DRIVER)* project and the *Guidelines for content providers: Exposing textual resources with OAI-PMH*, containing initial recommendations for interoperability. These recommendations were complemented by *OpenAIRE Guidelines for Literature Repositories*, which, at the moment this text was being written, were found in version 4 (OpenAIRE, 2018).

143 rd-alliance.github.io/metadata-directory/

OpenAIRE guidelines are organized in three sections: The first one is introductory; the second describes the use of *OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)*, with orientations about it and; finally a general view of a profile for application. This guideline is composed by 4 metadata standards: *Dublin Core;* and its qualified version; *Datacite* and; Oaire (standard elaborated by *OpenAIRE* itself). Some vocabularies controlled for use are also specified, as for example the *Confederation of Open Access Repositories (COAR)*.

Regarding OAI-PMH protocol, it is a tool for exposing metadata through *Hypertext Transport Protocol (HTTP)* and *Extensible Markup Language (XML)* languages that allow the communication and interoperability among databases. It is noteworthy that regardless its use in the interoperability among systems, and its recommendation in OpenAIRE guidelines, it is noted a certain limitation of this tool, in its metadata set (15 elements of *Dublin Core*), which has been used since the beginning of 2000 (Garcia; Sunye, 2003). Other initiatives are increasingly more prominent, especially on Semantic Interoperability, such as W3C: PROV-O (model of generic data of *World Wide Web Consortium*) and the Data Catalog Vocabulary (DCAT) 2.0 (a vocabulary in Resource Description Framework-RDF), which adds classes for describing data services and putting them in line with FAIR principles.

OpenAIRE V4 guidelines establish 4 levels of mandatory fields to be filled out: Mandatory (M), when filling out is mandatory (Applied to 6 fields); Mandatory if applicable (MA), if filling out is mandatory only if the information in the field is part of registry (for instance, the name of the funding body is mandatory, in the case of the dataset results from financing) that is applied to 8 fields; Recommended (R), that is relevant and important but not essential (applied to 15 fields) and; Optional (O), which would only add value to the description, even if it is not necessary (applied to 3 fields). The guideline application profile is, in short, represented by the table that relates the thirty-two Fields of the guideline, with the instructions for filling in and the controlled vocabularies selected to specific fields

The second tool in analysis is the FDP, an independent and open-code web application, developed as implementation of specification reference of FDP itself[144]. These specifications guide softwares for repositories, dealing with metadata management, in particular about semantic Technologies such as RDF, thus being a complementary tool to a data repository software. A repository based on FDP delas with interoperability issues, enabling findability, accessibility, interoperability and reuse (FAIR principles).

The implementation of FDP reference uses an API REST with several functions: creation, storage, release of metadata thus allowing these metadata to be exposed, provided and available in accordance with FAIR principles. It also allows finding metadata of sets available and access them when they have an open-use license. Any data repository can adopt FDP metadata, thus also working as an FDP instance.

One of the main specifications of FDP is the specification of levels of metadata[145], which guides the application of a profile in RDF, reusing standardized semantic models. Therefore, the specification of levels of FDP metadata introduces an organization in four levels of metadata: first, the metadata repository itself; second, the catalog; third, the dataset

144 github.com/FAIRDataTeam/FAIRDataPoint-Spec

145 github.com/FAIRDataTeam/FAIRDataPoint-Spec/blob/master/spec.md

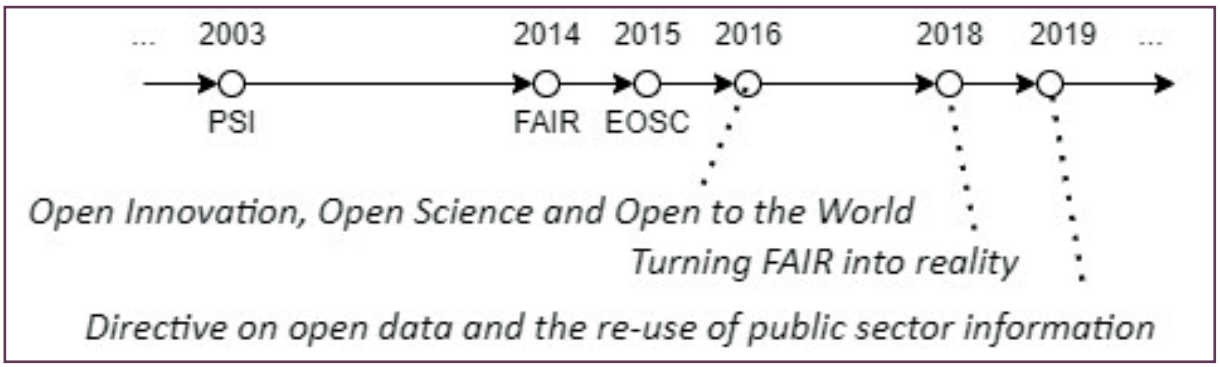
and; fourth, data distribution (the archives belonging to the set). A metadata repository can have one or more catalogs, each catalog can have one or more datasets, and each dataset contains one or more distributions.

FDP metadata standard is based on re3data schema[146] and DCAT vocabulary[147]. As previously noted, the metadata standard is organized in four levels and each property has two possibilities to be filled out: Mandatory, applied to ten fields, and Optional, applied to twelve fields. The levels describe, each one, a type of complex digital object that is possibly described, they are: the level of metadata repository, containing information on the data repository and the FDP itself; the level of metadata in the catalog, containing information on the collection, where each catalog represents a category (generally defined by domain); the level of dataset metadata, containing information on possible serializations of datasets, for instance, the individual archives that compose the datasets.

For example, the data repository B2Share (https://b2share.eudat.eu/) approaches catalogs through communities. Kinder Corona Studies (KiCoS) is one of the communities (catalogs) of B2Share and contains a series of datasets (in the tool represented as *records*); and a dataset can contain a series of distribution (*files*). The implementation of FDP metadata specification as a "semantic *proxy* (*wrapper*)" can add the aforementioned functionalities to the data repository software (Moreira *et al.*, 2019).

Also noteworthy in the European scenario actions and programs of the European Commission (EC) that deal with scientific data sharing and opening, such as: the European Open Science Cloud (EOSC), which dates from 2015; the Evolution of FAIR principles, which started in 2014. In 2016, the document Open Innovation, open Science and Open to the World was published; in 2018, the report turning FAIR into reality is released: Final Report and Action Plan on FAIR Data, and an increasing participation of EC in RDA is noticed; in 2019 there is the transformation of the guidelines of the Public Section Information (PSI), which had been edited in 2003, in the Guidelines for open data and reuse of public sector information (see Figure 1 below).

Figure 1 – Timeline of Other EC initiatives



Source: Designed by author.

146 https://www.re3data.org/schema

147 https://www.w3.org/TR/vocab-dcat-2/

The elaboration of general guidelines for scientific data repositories begins with a comparative analysis between OpenAIRE and FDP sets in Search of differences and similarities among the requirements. The full comparison can be found in the original document resulting from the landmark.

It was noted the possibility of equivalence among most of the fields compared, only the differences in definitions of mandatory levels in the standards on equivalent fields were highlighted. Based on the clarification of these mandatory differences, the search for definitions of minimum mandatory metadata began.

Difficulties were encountered in defining equivalences especially in OpenAIRE Fields related to *Type* subtypes (such as *dateType*), in addition to the adoption of different controlled vocabularies. In particular, OpenAIRE recommends a controlled vocabulary for *resourceType*, named *Controlled Vocabulary for Resource Type Genres (Version 2.0)*, which is a taxonomy for classifying resource genre typologies.

It was decided not to adopt this taxonomy because it was identified that it presents a series of semantic problems in its hierarchical relation, once it is not possible to identify which type of relation is used, for instance, if it is a relation of specialization, such as l *rdf:Type* (or *"is a"*), or if it is another type of relation. Another example is the case When the taxonomy describing an interview is a dataset, which does not seem to make sense once the interview is an intentional action (an event, or *perdurant*), while a dataset is a substance (an *endurant*) that can have different identity principles. This taxonomy also presents reasoning problems in relation to the principles of identity, rigidity and logical disjunction of the categories. For example, the taxonomy presents, in the same level, the *learning object* and *text* elements, which can be (or not) disjoined, and not share the same identity principles.

It is important to point out that adopting a taxonomy of this type can cause a series of difficulties in interoperability, since the machines need a precise description of the types of resources available in data repositories. An incompatibility was also found in the *dcat:distribution* field of the FDP, this field asks for a description of information about the individual archive that composes the dataset (*dcat:Dataset*).

The minimum mandatory central standard developed covers 13 fields with examples of completion and application that can be found in the original document resulting from the landmark in Paganine *et al.* (2020).

Starting from the definition of the core, the design of thematic metadata is approached. A difficulty was the issue of frequent multidisciplinary even in monothematic repositories. For that purpose, the table comparing knowledge areas was used as a guide when choosing and organizing these areas.

It started with the Search for thematic metadata schemes. A page published by DCC in 2020 was used as an initial tool, with different metadata schemes, divided by fields of knowledge. This list was analyzed and deduplicated for the standards of the chosen areas (Social science, Biology and Agronomy). However, the result did not present satisfactory specificity and coverage. Therefore, a list of standards maintained by a group interested in RDA[148]

148 rd-alliance.github.io/metadata-directory/

metadata was used (RDA, 2020). Finally, the result obtained was refined by checking the metadata obtained with the list displayed in re3data[149] record, measuring which of the selected standards are efficiently used in thematic data repositories related to the selected areas.

Following the level of complexity from the simplest to the most difficult to be treated, the analysis was followed in the tables in the Biology area. Three metadata schemes were analyzed: MIBBI, Darwin Core and ABCD. MIBBI has 40 categories or modules, as they are named, with 23 main fields, but only 17 of them are completely developed, and the remaining were still being designed when the table was collected. Darwin Core has 12 description packets with semantic upload or categories, while ABCD has 38 categories of expandable metadata. While browsing in re3data it is noted that, excluding generic schemes and schemes from different or general knowledge areas, from the 42 results, 2 used MIBBI and only one used Darwin Core.

After the comparative analysis, the metadata of apparent importance in the area were selected, according to their frequency in the standards, eliminating redundancies and generalizing similar terms. The result is also found in the original document, containing 8 fields using Darwin Core, which was chose for the ease of translation and comparison with the traditional Dublin Core:

When focusing on the Social Sciences, which is an area of greater coverage in the chosen standards, a range of standards is also observed. They are: METS, MODS, MARC, CERIF and Dublin Core (DC). METS has 7 classes, MODS, has 20, DC, has 15, MARC, has 9 and CERIF has 22. These classifications and respective Fields or elements of the standards were compared following the same process applied in the Biology standards. The selected result was then adapted to Datacite language, due to its development towards scientific data and the extensive application and compatibility.

Finally, in regard to Agriculture, the same process applied in Biology is used. The schemes AGRIS and AgMES were selected. AgMES has 21 fields based on DC and covers semantic standardizations in agriculture on description, finding resources, interoperability and data exchange in several informational resources. AGRIS has 16 fields and is aimed at the international system of information on agricultural sciences and technologies guidelines, on good practices for information. In re3data record it was not identified the intended use of any of the 2 schemes in thematic repositories of the agriculture area. In order to describe the fields herein, AgMES scheme was chosen due to its proximity with the DC.

Also noteworthy are the negative results obtained in Health, the area presented an unforeseen high complexity of sets description standards. There was a lack of documentation on these, thus, metadata schemes specific to the area were not identified. In the attempt to capture some of the fields frequently used in the area, queries were carried out with thematic data repositories and with some experts from Fiocruz, but the results obtained were still not satisfactory and sufficient for the elaboration of a proposal that encompass the entire Health area.

149 https://www.re3data.org/

The general multi-thematic metadata standard developed herein has already been applied in the development of Ibict scientific data repositories (named *Aleia*) and CNPq (named *LattesData*).

Aleia and LattesData creations were also motivated by Commitment 3, of the 4th National Action Plan of OGP Brazil, through a technical cooperation agreement (TCA) between CNPq and Ibict, in December 2019. Aleia aims at providing a tool with the functionality of recording, gathering, organizing, disseminating, sharing and preserving scientific data from research carried out by Ibict collaborators and scientific datasets external to the body, but from specific scientific communities. LattesData aims at working as official tool that enables its funded researchers to make inputs in datasets that emerged as a result of projects developed with CNPq resources, being part of the accountability procedures, as well as non-client and partner institutions of CNPq that sign agreements for the collaborative use of the space created.

Both repositories are multi-thematic and intend to cover datasets from researchers from different institutions, from several areas and realities. With this in mind, it was decided to start by just the application of the minimum central description standard with just a few minor changes and additions to better fit its supporting institution.

Aleia additional metadata are presented as follow:

Table 1 – Aleia additional metadata

Field	Description
Author's curriculum in the Lattes platform	Address to access curriculum in the Lattes platform (this field was just adapted from the field "Identifier" of the original standard)
Author's institute of origin	Full name of the institution to which the researcher is linked
Description	General description of dataset and its content
Alternative title	Title of dataset in another alternative language
Contact	Email of the responsible for the dataset.
Dataset language	Language in which the dataset was developed
Reading software and data manipulation	Program used to access and manipulate the dataset archives

Source: Designed by the authors

The additional metadata of LattesData are presented in the following table.

Table 2 – LattesData additional metadata

Field	Description
Contact	Email, preferably institutional, of the responsible for the data
Author's Internal Identifier (IDLattes)	Definition: Identifier number of the curriculum in the Lattes platform (this field was just adapted from the field "Identifier" of the original standard)
Author's external identifier	ORCID identifier number
Author's institution	Name of the institution to which the author is affiliated
Dataset alternative identifier	Another persistent identifier of dataset obtained differently from the main one used in LattesData repository
Notes	Free text that can be used to list/describe archives (related each archive, its type, description and it also informs if it needs a specific software), comments or guidelines for access among other details
Project summary	An explanatory text describing the project and the dataset in a general way, encompassing conclusions, methodology, collection, etc.
Value received	Value, in reais (Brazilian currency) received by the Project from the funding body
Project validity	Start and end date of the project
Materials and other related products	Any product or material related to the dataset Other than scientific publication in formal standard

Source: Designed by authors

Once the additional fields were defined in both repositories, great difficulties were encountered to change them and the form in the chosen software (Harvard Dataverse). As an attempt at a solution, a communication form is being developed with API REST of the Dataverse, for retrieving and filling in metadata externally to the software. Another difficulty encountered is the change of Fields prefixes that comes with the Data Documentation Initiative (DDI) standard instead of the DCAT and Datacite scheme recommended. In the future, it is planned to integrate the filling in of metadata with the development of a Data Management Plan (DMP), so that it also have a machine-readable format.

When analyzing the objectives and results obtained up to now, it is noted that the general core of minimum metadata presents information enough for the interoperability of the chosen guidelines. However, it is interesting to highlight the importance of adding other relevant institutional information to contribute with a more qualified description of the sets deposited, and its association to financed projects, in the case of the CNPq.

Finally, it is also highlighted the magnitude of the complexity of treatment and description of different knowledge areas, in multi-thematic repositories and in special, When considering the possibility of also covering institutions with very diversified realities and contexts.

11.3 FINAL CONSIDERATIONS

The main difficulty faced in the execution of this work was the fact that the repositories do not adopt well-known metadata standards, and the fact that the standards adopted in different areas are not compatible with each other. The content presented as a result from the work carried out in OGP is adequate when dealing with areas that have efforts towards the representation of scientific products, but the description of datasets, which potentially covers any area of knowledge, shows itself to be a work that is constantly changing and updating. Thus, it is noteworthy that this is not a study that ends in its application, on the contrary, it needs constant development in Search of extensions and adaptations to Other areas of knowledge. It is also intended to develop a corpus composed of a set of metadata that describes research data for each specific areas of knowledge. This information will be collected in data repositories of unique themes registered in *re3data* which allow communication. The collection will be automated and will cover: title, keywords, abstract and subject. The Corpus will then be organized and its visualization generated with the VosViewer program for identification of key issues and specific areas highly populated by datasets.

It is noteworthy that the work presented here was prepared in collaboration with IBICT, CNPq and the University of Twente. For future developments, it is intended to continue studies in other areas of knowledge, in special exact sciences and health, as well as in the ontological analysis of genre types of resources to address the problems identified in the classification recommended by OpenAIRE. Future activities also include widespread adoption of FAIR principles and, particularly, the evolution of semantic interoperability among data repositories through application of well-founded ontologies in developing repositories.

REFERENCES

CNPQ. **Tabela de Áreas do Conhecimento**. Brasília: CNPQ, 2020. Available from: http://lattes.cnpq.br/web/dgp/arvore-do-conhecimento. Access on: 28 May 2020.

DCC. **Disciplinary Metadata**. [*S. l.*], 2020. Available from: dcc.ac.uk/resources/metadata-standards. Access on: 28 May 2020.

DFG. **Classification of Subject Area, Review Board, Research Area and Scientific Discipline**. Bonn, 2020. Available from: dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp. Access on: 28 May 2020.

FDP. **FAIR Data Point Specification**. [*S. l.*], 2016. Available from: github.com/FAIRDataTeam/FAIRDataPoint/wiki/FAIR-Data-Point-Specification. Access on: 28 May 2020.

GARCIA, Patrícia de Andrade Bueno; SUNYE, Marcos Sfair. O protocolo OAI-PMH para interoperabilidade em Bibliotecas Digitais. Ponta Grossa, p. 1-12, 2003. Available from: https://www.researchgate.net/publication/229039114_O_protocolo_OAI-PMH_para_interoperabilidade_em_Bibliotecas_Digitais. Access on: 28 May 2020.

MOREIRA, João Luiz R.; BONINO, Luiz.; PIRES, Luís F.; SINDEREN, Marten van; HENNING, Patricia. Repositórios para dados localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR): adaptando um repositório de dados para se comportar como um FAIR Data Point. **Liinc Em Revista**, v. 15, n. 2, 2019. Available from: http://revista.ibict.br/liinc/article/view/4817. Access on: 28 May 2020.

OPENAIRE. **DRAFT:** OpenAIRE Guidelines for Literature Repository Managers v. 4. European Union, 2018. Available from: openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/. Access on: 28 May 2020.

OECD. **Frascati Manual 2015**: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities. Paris: OECD Publishing, 2015. Available from: oecd.org/publications/frascati-manual-2015-9789264239012-en.htm. Access on: 28 May 2020.

PAGANINE, Lucas Nóbrega; CARVALHO SEGUNDO, Washington Luís R. de; MOREIRA, João Luiz R.; SAYÃO, Luís F.; BARRETO NETO, Vanderlino Coelho; CIUFFO, Leandro Neumann; FELICÍSSIMO, Carolina Howard; DIAS, Gustavo Neves; **Padrões de interoperabilidade para repositórios de dados de pesquisa**. Brasília, DF: IBICT , RNP; Rio de Janeiro: CNEN; Enschede: University of Twente, 2020. 44 p. Available from: https://www.gov.br/cnen/pt-br/acesso-rapido/centro-de-informacoes-nucleares/material-didatico-1/padroes-de-interoperabilidade-para-repositorios-de-dados-de-pesquisa.pdf. Access on: 28 May 2020.

RDA Metadata Standards Directory. 2020. Available from: https://rd-alliance.github.io/metadata-directory/. Access on: 28 May 2020.

RE3DATA. **Registry of research data repositories**. [*S. l.*], 2020. Available from: service.re3data.org/about. Access on: 28 May 2020.

RNP. Wiki **Ciência Aberta na OGP Brasil**. 2018. Available from: wiki.rnp.br/pages/viewpage.action?pageId=107315238. Access on: 30 Apr. 2020.

SANTOS, L. FAIR Data Points Supporting Big Data Interoperability. *In*: MERTINS, Kai; JARDIM-GONÇALVES, Ricardo; POPPLEWELL, Keith; MENDONÇA, João P.(org.). **Enterprise Interoperability in the Digitized and Networked Factory of the Future**. London: ISTE Press, 2016. p. 28

WILKINSON. Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle; AXTON Myles; BAAK, Arie; BLOMBERG, Niklas; BOITEN, Jan-Willem; SANTOS, Luiz Bonino da Silva; BOURNE, Philip E.; BOUWMAN, Jildau; BROOKES, Anthony J.; CLARK, Tim; CROSAS, Mercè; DILLO, Ingrid; DUMON, Olivier; EDMUNDS, Scott; EVELO, Chris T.; FINKERS, Richard; GONZALEZ-BELTRAN, Alejandra; GRAY, Alasdair J. G.; GROTH, Paul; GOBLE, Carole; GRETHE, Jeffrey S.; HERINGA, Jaap; HOEN, Peter A. C 't; HOOFT, Rob; KUHN, Tobias; KOK, Ruben;

KOK, Joost; LUSHER, Scott J.; MARTONE, Maryann E.; MONS, Albert; PACKER, Abel L.; PERSSON, Bengt; ROCCA-SERRA, Philippe; ROOS, Marco; SCHAIK, Rene van; SANSONE, Susanna-Assunta; SCHULTES, Erik; SENGSTAG, Thierry; SLATER, Ted; STRAWN, George; SWERTZ, Morris A.; THOMPSON, Mark; LEI, Johan van der; MULLIGEN, Erik van; VELTEROP, Jan; WAAGMEESTER, Andra; WITTENBURG, Peter; WOLSTENCROFT, Katherine; ZHAO, Jun; MONS, Barend. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, England, v. 3, 2016. Available from: https://www.nature.com/articles/sdata201618. Access on: 28 May 2020.

How to cite this chapter: PAGANINE, Lucas N.; CARVALHO SEGUNDO, Washington L. Ribeiro de; MOREIRA, João L. R. Development and application of interoperability standards for scientific data repositories: Ibict and CNPQ repositories. *In*: SALES, Luana Faria; VEIGA, Viviane Santos de Oliveira; VIDOTTI, Silvana Aparecida Borsetti Gregório; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **FAIR Principles Applied To Research Data Management**: Brazilian Experiences. Brasília, DF: Editora Ibict, 2024. cap. 11, p. 133-144. DOI: 10.22477/9788570131959.cap11.