

# Some considerations about research data repositories: a summary report

L. F. Sales<sup>1</sup>, L. F. Sayão<sup>2</sup>  
 e-mail: luanafsales@gmail.com,  
[lsayao@cnen.gov.br](mailto:lsayao@cnen.gov.br)

<sup>1</sup> *SEBICT, IEN*

<sup>2</sup> *CIN, CNEN*

**Keywords:** Research data, Digital repository, Research data management.

Because of the commitment to seek new knowledge and new discoveries, contemporary scientific research produces and intensively uses digital research data. In this changing scenario, data are no longer simple byproducts of research activities, but have become first-rate information resources, characterizing a new scientific paradigm based on data sharing, access and reuse.

The objective of this research was to identify the role of digital data repositories in the new scientific research scenarios, and to present an overview of their main characteristics, categories, benefits, functions and infrastructures. The methodology consisted in the analysis of the related literature and the main systems that give support to infrastructures for accessing and managing research data. This research verified that data and collections of research data transmit knowledge in time and space and are reused, it is necessary to implement a permanent and sustainable technological and managerial infrastructure that allows them to be cared throughout life cycle. At the heart of this infrastructure are digital repositories of research data, which are systems designed to support the selection, cataloging, archiving, access and sharing of research data. As a result, it was obtained a generic model of the flow of research data management steps (Figure 1) and a systematization of the data research repository functions (Table 1).



Figure 1. Ranking to fuzzy top event

Table 1 - Data research repository functions

<b>DATA INGEST</b>
Insertion of data sets - raw or derived - from experiments, simulations, observations, questionnaires, surveys, etc. The data can be incorporated into the repository by the authors themselves - by self-submission - or by library services

<b>Functions</b>	<ul style="list-style-type: none"> <li>• Selection of data that can be archived;</li> <li>• Verification of the suitability of the data in the scope of the repository;</li> <li>• Verification of acceptable file formats for submission;</li> <li>• Verification of rights associated with collections (copyright and licenses);</li> <li>• Verification of sensitive data (non-anonymous, confidential, personal);</li> <li>• Verification of the volume and quantity of files;</li> <li>• Verification of the general and disciplinary metadata that accompany data;</li> <li>• Normalization of the list of standardized formats accepted for archiving and dissemination;</li> <li>• Data quality control;</li> <li>• Definition of embargo period.</li> </ul>
<b>CATALOG OF DATA SETS</b>	
Data description, metadata assignment and attachment of documentation to ensure that data can be accessed and interpreted in time and space. Attribution of: descriptive, structural, administrative, technical metadata (which includes those relating to the technical dependencies of digital objects)	
<b>Functions</b>	<ul style="list-style-type: none"> <li>• Assignment of preservation metadata, which ensures the provenance, authenticity and integrity of data over time;</li> <li>• Use of specialized and disciplinary taxonomies;</li> <li>• Persistent identifier assignment (DOI, Handles, UNF, URN, etc.) that data can be persistently cited as academic publications;</li> <li>• Author identification (ORCID ID, Scopus Author ID, ResearcherID, etc.);</li> <li>• Inclusion of documentation on the data, including project description, files and parameters; laboratory and field notebooks, research protocols or methodology, etc.;</li> <li>• Linking data to publications and to related internal and external data.</li> </ul>
<b>ARCHIVING AND PRESERVATION</b>	
Archiving to ensure short- and long-term management of data sets guided by a plan / policy of digital preservation	
<b>Functions</b>	<ul style="list-style-type: none"> <li>• Storage in protected systems;</li> <li>• Management of short-term preservation (backups, redundant backups offsite; integrity checking, secure storage, encryption, compression);</li> <li>• Long-term management (migration, emulation, reformatting to standardization, application of relevant standards (OAIS, TRAC), of fixity aimed at validating the authenticity and integrity of an object digital (checksums, digital signature)</li> <li>• Implementation of audit trails.</li> </ul>
<b>INTEROPERABILITY</b>	
Exchange and sharing and linkage with other data repositories and other information systems (institutional repositories, digital libraries of academic publications, scientific publishing)	
<b>Functions</b>	<ul style="list-style-type: none"> <li>• Availability of metadata according to the OAI-PMH protocol;</li> <li>• Aggregation for formation of extended publications according to the standard OAI-ORE;</li> <li>• Use of standards, web service, linked data and others</li> <li>• Metadata packaging for interchange according to METS standards.</li> </ul>
<b>DISCOVERY, ACCESS AND REUSE</b>	
Web interface for discovering, accessing and downloading collections of data relevant to the user or for computational applications, such as visualization and mapping, which can provide services from these collections; linked to an access policy established by the institution that includes: period of embargo, right of access, payments, restrictions on certain collections, access to metadata only; user records and terms of use of the data.	
<b>Functions</b>	<ul style="list-style-type: none"> <li>• Providing web interfaces for retrieval, access and downloading;</li> <li>• Availability of applications and services on collections</li> </ul>

In a conclusion, it can be stated that because of their importance as information resources, data repositories have quickly become an essential part of the research infrastructures on a global scale, by allowing an important part of the research activity visible and open to society as a whole. In this way, they pose some relevant challenges for Information and Library Science. To read more about this work, consult the reference below. [1]

## References

[1] SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa *Informação & Informação*, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016.