

PROV FAMILY AND DATA REPOSITORY IN THE CONTEXT OF RE3DATA

Família Prov e repositório de dados no contexto do Re3data

Felipe Ivo da Silva

Attending Masters Degree in Information Science
at Universidade Federal de São Carlos,
São Carlos, São Paulo, Brasil.
felipe_ivodasilva@hotmail.com
<https://orcid.org/0009-0005-1379-4692>

Felipe Augusto Arakaki

PhD and Master in Information Science
at Universidade de Brasília,
Brasília, Distrito Federal, Brasil.
felipe.arakaki@unb.br
<https://orcid.org/0000-0002-3983-2563>

ABSTRACT

The purpose of this study is to investigate the adoption of PROV Family metadata standards in research data repositories listed in the Research Data Repositories Information. The Re3Data database was used to search and retrieve data repositories, and 3,302 repositories were analyzed, of which only 8 adopt these standards. The low use reveals challenges in data management, since provenance is essential for tracking, integrity, and reuse of data over time. The research highlights the need to expand the adoption of PROV standards. The use of these standards contributes to more effective management of scientific data, strengthening the transparency and reliability of science. The adoption of these metadata can improve research reproducibility, ensuring that data are understood and reused correctly. Therefore, it is essential that institutional policies encourage the implementation of these standards in scientific data repositories.

Keywords: Metadata. Provenance. Data repository. PROV.

RESUMO

Objetivo: Investigar a adoção dos padrões de metadados da Família PROV em repositórios de dados de pesquisa listados no Research Data Repositories Information. Para busca e recuperação de repositórios de dados, utilizou-se a base Re3Data e foram analisados 3.302 repositórios, dos quais apenas 8 adotam esses padrões. O baixo uso revela desafios na gestão de dados, pois a proveniência é fundamental para rastreamento, integridade e reutilização dos dados ao longo do tempo. A pesquisa ressalta a necessidade de ampliar a adoção dos padrões PROV. O uso desses padrões contribui para uma gestão mais eficaz dos dados científicos, fortalecendo a transparência e a confiabilidade da ciência. A adoção desses metadados pode aprimorar a reprodutibilidade da pesquisa, garantindo que os dados sejam compreendidos e reutilizados corretamente. Assim, é essencial que políticas institucionais incentivem a implementação desses padrões nos repositórios de dados científicos.

Palavras-chave: Metadado. Proveniência. Repositório de dados. PROV. PROV-O.



1 INTRODUCTION

Information plays a fundamental role in global development, and with the advancement of technology, the amount of digital information has significantly expanded. As asserted by Brandt et al. (2019), the discourse surrounding metadata has witnessed a marked escalation in prominence within the domains of information and technology, consequent to the exponential increase in the volume of available information. In light of the foregoing, the study of Provenance is indispensable for the characterization and reliability of information. As Hayens (2018, p. 134) emphasises, the availability of information about provenance is essential to prove the authenticity of a record. Thus, ensuring it has not been altered, and that the evidence presented is reliable. As a result, the validation and credibility of information resources, such as metadata, is ensured, especially in the digital context.

Furthermore, provenance is a term that encompasses several areas of study, such as data science, library science, and information management. In this regard, the concept of provenance adopted in this article is that described by Moreau and Groth (2013), which characterizes a set of information about individuals, institutions, entities, and actions linked to the creation, influence, or dissemination of specific data or objects.

As Arakaki and Santos (2021) have demonstrated, the term 'provenance' is employed to identify the individual or entity that is responsible for the generation, storage, and management of information and resources in a variety of domains. The utilisation of metadata within this identification process is advantageous, as it facilitates the evaluation of information regarding a particular information resource with greater precision. Consequently, the reliability and precision of the data can be assured.

Consequently, data repositories have emerged as instrumental tools to facilitate the execution of the necessary processes for research data management (Sayão; Sales, 2016) and play a pivotal role in open science, enabling researchers to deposit their data so that it can be accessed and reused by others. However, it

should be noted that not all repositories adopt provenance metadata, which has the potential to impact quality and usability of said data.

Metadata can thus be considered a practical application that concerns the description of objects, the development of databases, and the digital recording of transactions. Joudrey, Taylor, and Wisser (2018) also emphasise that metadata can comprise descriptive information pertaining to the context, quality, condition, or characteristics of the data.

This definition suggests that metadata encompasses not only descriptive information, such as that found in conventional resource discovery tools, but also information necessary for management, utilisation and preservation of the information resource (e.g. location data, online display information, ownership information, condition data). In this sense, metadata can be considered part of the description process and may be guided by cataloguing principles.

It is therefore evident that the establishment of a definition and standardisation of information resource metadata is pivotal in the creation of tools that ensure the identification and preservation of information resources. Such tools are also essential for the facilitation of search, access, use, and reuse of information. Consequently, provenance metadata emerges as a solution for users to more accurately evaluate the selection of a specific information resource, thereby ensuring the quality and accuracy of the data in question. Consequently, metadata is regarded as a pragmatic application associated with cataloguing, indexing, database development, and digital transaction recording.

Despite the extensive discourse on metadata within the domain of Information Science, the implementation of provenance in research data repositories remains underdeveloped. This is evidenced by Arakaki's (2019) observation that "research on provenance in Brazil remains underdeveloped, lacking both theoretical exploration and practical or professional initiatives." Furthermore, existing literature lacks comprehensive studies that systematically explore the adoption of PROV Family standards and their effects on data reliability and reuse. The objective of this

research is to address this gap by identifying repositories that utilise these standards, conducting a thorough analysis of the challenges and implications for data management, and contributing to the advancement of open science and the enhancement of digital curation.

In light of the aforementioned context, the objective of this study is to ascertain which data repositories utilise PROV Family metadata standards, according to Research Data Repositories Information (R3DATA). It is imperative that this investigation is undertaken in order to comprehend both the necessity of adhering to these standards and the significance of provenance in research data repositories, particularly within the paradigm of open science.

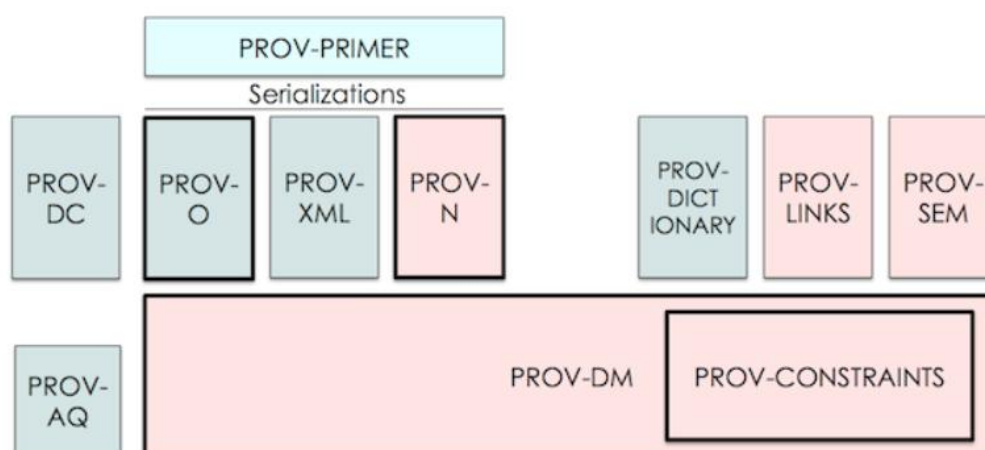
The endmost goal of this analysis is to emphasise the significance of employing provenance metadata standards, particularly the PROV Family. It aims to showcase how these standards play a vital role in advancing information frontiers, enhancing data quality and accuracy, and strengthening its technological relevance.

2 PROV FAMILY

In 2013, the World Wide Web Consortium (W3C) established a working group with the objective of developing a set of specifications known as the PROV Family. The purpose of this initiative was to establish a model for ensuring data interoperability on the Web. The family comprises four primary recommendations: the PROV Data Model (PROV-DM), the PROV Ontology (PROV-O), the Provenance Notation (PROV-N), and the Constraints of the PROV Data Model (PROV-CONSTRAINTS). In addition to these guidelines, eight supplementary notes have been published that offer additional guidance and information, facilitating the understanding and application of the PROV model.

As illustrated in Figure 1, the documents that comprise the PROV Family are characterised by a specific structure and interconnectedness.

Figure 1 - PROV Family



Source: Groth and Moreau (2013, not paginated).

As stated by Groth and Moreau (2013), the PROV family of documents establishes a model, serialisations and other supporting definitions that facilitate the exchange of provenance information in heterogeneous environments, such as the Web. Accordingly, the W3C document pertaining to the PROV family offers a non-normative synopsis, in conjunction with the establishment of guidelines and recommendations for its utilisation.

It is evident that the PROV family, developed by W3C, offers a set of specifications that promote interoperability and the exchange of provenance information on the Web. This guarantees that metadata encompasses not only contextual descriptions, but also critical details required for data management, usage, and preservation. The PROV metadata standards established by W3C support detailed tracking of provenance information, thereby promoting confidence in the data shared within repositories.

In order to facilitate the use of the PROV model, W3C mapped it to Dublin Core, enabling the provenance metadata standard to be more easily incorporated into data repositories that already use Dublin Core as their main metadata standard.¹

Table 1 illustrates how the 'prov:wasGeneratedBy' property from the PROV family can be applied. This example uses prefixes, such as RDF and XSD, to define

¹ <https://www.w3.org/TR/2012/WD-prov-dc-20121211/>

namespaces, and establishes relationships between entities and activities. Following W3C specifications for the PROV Family, this model illustrates how `prov:wasGeneratedBy` property can be used to indicate that an entity was generated by a particular activity.

Table 1 – Example of `prov:wasGeneratedBy` property.

```
1 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
2 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .
4 @prefix prov: <http://www.w3.org/ns/prov#> .
5 @prefix : <http://example.com/> .
6 :bar_chart
7   a prov:Entity;
8   prov:wasGeneratedBy :illustrating;
9 :illustrating a prov:Activity .
```

Source: Conducted by W3C².

Table 2 illustrates the application of PROV Family's `prov:wasDerivedFrom` property. It illustrates how this property can be used to represent the derivation relationship between entities in accordance with the W3C PROV Family specifications.

Table 2 – Example of `prov:wasDerivedFrom` property.

```
1 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
2 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .
4 @prefix prov: <http://www.w3.org/ns/prov#> .
5 @prefix ex: <http://example.com/vocab#> .
6 @prefix : <http://example.com/> .
7 :bar_chart
8   a prov:Entity, ex:Barchart;
9   prov:wasDerivedFrom :aggregatedByRegions;
10 .
11 :aggregatedByRegions
12   a prov:Entity, ex:Dataset;
13 .
```

Source: Conducted by W3C³.

The models outlined in Tables 1 and 2 were proposed by W3C to support the implementation and optimize the use of metadata standards. These models

² <https://www.w3.org/TR/prov-o/#wasGeneratedBy>

³ <https://www.w3.org/TR/prov-o/#wasGeneratedBy>



emphasise key characteristics for the process of representing information resources, thereby facilitating their implementation. Both models emphasise the importance of clearly and structurally representing the relationships between entities, activities and agents in accordance with W3C specifications.

Re3data (2024) reports that Dublin Core stands as the most commonly adopted metadata standard across data repositories, utilised by approximately 619 repositories. In addition, its compatibility with PROV enables the exchange of detailed provenance information within these systems, significantly enhancing the scope and impact of provenance metadata applications.

Standardised and well-defined metadata are fundamental to the development of cataloguing tools that improve search, access, use and reuse of information, thus consolidating their relevance in technology.

3 RESEARCH DATA REPOSITORIES

The intense generation of data and information within academic and scientific scenarios fosters the development of digital information environments, such as research data repositories. These repositories are "[...] a necessary tool for storing and managing data used and produced during research" (Vidotti et al., 2017, p. 8).

Within the framework of open science, the Registry of Research Data Repositories (Re3data) serves as a worldwide directory of research data repositories. It consolidates data from various fields of knowledge, offering it in multiple formats and ensuring universal accessibility.

This global directory, which originated in Germany in 2012, facilitates the consultation of research data by including repositories that enable permanent storage and access to data sets for researchers, funding agencies, publishers and academic institutions. The founding partners of Re3data include the Berlin School of Library and Information Science, the Helmholtz Open Science Office at the German Research Centre for Geosciences (GFZ), the Karlsruhe Institute of Technology (KIT) Library, and Purdue University Libraries (Re3data, 2024).

The directory offers multiple search options, including by country, topic and data type. Consequently, Re3data covers a wide range of knowledge areas, promoting sharing, access, and greater visibility of research data. Each result retrieved from the directory includes a description or summary containing information about the corresponding research, metrics from various directories, and details such as country of origin, topic or keywords, and document type.

In order for research data to be reused, it must be preserved and easily accessible for location and retrieval. According to Hedstrom (1997), digital preservation is a process that encompasses “planning, resource allocation, and the application of preservation methods and technologies necessary to ensure that digitally stored information of continuing value remains accessible and usable.” In this sense, provenance metadata is part of the activities required to ensure that digital objects can be located, reproduced, and understood over time.

It can therefore be said that data repositories are fundamental to data sharing, as their structure is designed for this purpose. Thus, detailed description and retrieval of stored data are fundamental operations in these environments. In addition to providing the system, effective management must be implemented in order to allow the use and reuse of data.

In terms of provenance, this management also ensures long-term data preservation and facilitates access for those seeking to reuse it effectively, providing indispensable information for informational reliability.

4 METHODOLOGICAL PROCEDURES

This study is characterized as an exploratory and descriptive qualitative review of the literature. As such, this research approach was chosen as it allows for contextualizing the problem related to cataloguing provenance metadata sets, with a particular focus on the PROV family. Additionally, this methodology plays a significant role in fostering the development of analytical perspectives that underpin the establishment of the theoretical foundation for the study (Alves-Mazzotti, 2002).

Consequently, the data analysis is predominantly qualitative and exploratory, based on categorising and synthesising the obtained information. To achieve the proposed objective, Re3data was used as a methodological tool to identify which repositories use metadata from the PROV family. This method consists of identifying patterns and categorising the responsible institutions, as well as the name of the repository and the predominant study topic within the repository.

As it examines and comprehends the various theoretical aspects inherent to the studied phenomenon, it is also exploratory research – based on information obtained from literature review (Gil, 2017), in addition to describing the use of provenance metadata in Data Repositories. Re3data was used to select data repositories that use the provenance metadata proposed by the World Wide Web Consortium (W3C) for data collection.

5 RESULTS ANALYSIS

This section provides an analysis based on information obtained from Re3Data, highlighting research data repositories that utilize the PROV family for describing their collections. To ensure quality, the search filters available in Re3Data were employed to identify metadata standards implemented within these data repositories, identifying 29 standards in 3,302 repositories. This mapping focused exclusively on identifying repositories that adopt PROV Family standards, with no investigation into the existence of other metadata standards that could potentially perform similar or compatible functions with provenance. It is evident that these repositories often employ multiple metadata standards. Nonetheless, it should be noted that only eight repositories have adopted the PROV family⁴.

Table 3 categorises information based on an analysis of research data repositories found in Re3Data that utilise the PROV family to describe their collections. Three main aspects were considered: the institution responsible for the repository; its geographical location; and the subject area to which it belongs.

⁴ <https://www.re3data.org/metrics/metadataStandards>



This categorisation enables us to understand how repositories adopt provenance as part of their metadata, demonstrating its applicability in various fields of study.

Table 3 – Characterization sheet for data repositories using PROV.

Repository name	Responsible institution/ Country	Theme/Area
Rotterdam Ophthalmic Data Repository	Rotterdam Ophthalmic Institute – Netherlands	Medicine and Life Sciences
DANDI	Dartmouth College – USA	Medicine, Neuroscience, and Life Sciences
MorphoSource	Duke University, Trinity College of Arts & Sciences – USA	General
Jülich DATA	Forschungszentrum Jülich – Germany	Life Sciences, Natural Sciences, and Engineering Sciences
DesignSafe-CI Data Depot Repository	National Science Foundation – USA	General
TROLLing	CLARIN-ERIC – European Union	Linguistics, Humanities, and Social Sciences
NSF Arctic Data Center	DataONE – USA	General
DataverseNO	Inland Norway University of Applied Sciences – Norway	General

Source: Conducted by the authors (2024)⁵.

Although the selected repositories are maintained by institutions of different nationalities, including the United States, the Netherlands, Norway and Germany, the majority of the content they host is predominantly in English. Another aspect is that all repositories contain metadata identifying the origin of the content. The records also contain additional information, such as the title, author, date and place of creation or collection, format, size, permissions, usage, and system requirements.

The metadata standards implemented by the eight repositories were identified as those established by the Digital Curation Centre (DCC), an internationally recognized organization specializing in digital information curation. The DCC is

⁵ <https://www.re3data.org/search?query=PROV>



dedicated to fostering expertise, developing capacities, and enhancing skills in the effective management of research data.

This institution promotes oversight in the informational description of repositories that use PROV Family metadata. Rusbridge et al. (2005, p.5) demonstrate the importance of implementation regarding data provenance and quality:

The value of digital curation is lessened if the evidence as to the origin and integrity of the data is lost or unknown. Current DCC work in this area includes developing formal models to support the description of the state of the problem and to extend our understanding of the problems that arise when data are copied from one database to another or when tracking where the data has come from that has not been adequately documented (its provenance not suitably clear). As these formal models emerge they will assist those developing software tools and standards for tracking, exchanging and managing the provenance of data as it is transferred between databases.

The survey revealed that, according to Research Data Repositories Information, only eight research data repositories claim to use PROV Family metadata standards. According to Re3Data, all of these repositories are linked to the Digital Curation Centre (DCC), an institution responsible for implementing PROV Family metadata standards schemas in these repositories.

The DCC and Re3data identify the application of PROV family metadata standards in eight repositories, yet no specific manuals demonstrating their implementation were discovered. Despite this, the DCC's metadata standard application profiles reference the original W3C manuals, which outline the use of PROV⁶.

The analysis focused on the repositories outlined in Table 3, facilitating the identification of TROLLing and DataverseNO as repositories utilizing PROV standards. Specifically, the analysis identified those that use the prov:wasGeneratedBy and prov:wasDerivedFrom standards.

⁶ <https://www.dcc.ac.uk/resources/metadata-standards/prov>



In Table 4, the mapping of the presence or absence of metadata standards is presented, as well as the format of the analyzed files. The metadata standards of the first ten documents published in 2024 were investigated, selected based on the recency of their publications.

Table 4 – Mapping of repositories to verify the presence of PROV.

Repository name	Documents from January 1, 2024, to November 15, 2024	Analyzed documents	PROV standards	File format
Rotterdam Ophthalmic Data Repository	10	10	It was not possible to identify the use of any metadata standards.	-
DANDI	131	10	0	Json
MorphoSource	141	10	The repository has internally structured metadata, although it does not explicitly display it to the user ⁷ .	-
Jülich DATA	99	10	0	HTML
DesignSafe-CI Data Depot Repository	273	10	0	XML
TROLLing	27	10	10 files prov:wasGeneratedBy prov:wasDerivedFrom.	HTML
NSF Arctic Data Center	189	10	0	XML e HTML
DataverseNO	179	10	4 files prov:wasGeneratedBy prov:wasDerivedFrom.	HTML

Source: Conducted by the authors (2024).

Given the aforementioned, in the sample observed, four repositories that did not use PROV metadata standards in their document descriptions were identified; along with one repository that has internally structured metadata, but does not explicitly display it to the user, limiting transparency and verification of standards such as the PROV Family; and one repository where it was not possible to locate any metadata standards application profile, which points to the need for greater transparency in the availability of metadata in the analysed repositories.

⁷ <https://www.morphosource.org/terms/ms#metadata>



The two PROV properties identified in the repositories are defined as follows:

1) prov:wasGeneratedBy: generated by the completion of the production of a new entity by an activity; 2) prov:wasDerivedFrom: derived from the transformation of one entity into another, which may also involve updating or constructing a new entity based on a pre-existing one (Lebo et al., 2013).

In addition, none of the first ten documents analysed from the Rotterdam Ophthalmic Data Repository provided a data format that would allow the extraction of metadata, making it impossible to verify the use of PROV Family standards and other standards. Although the MorphoSource repository does not provide user-visible data formats that allow direct metadata extraction in the first documents analysed, it has internally structured metadata. However, this metadata is not displayed explicitly in the user interface.

These two repositories lack structured information, which makes it difficult to assess the traceability and interoperability of metadata, representing a gap in data transparency.

In repositories that presented “0” PROV standards, other metadata standards were identified, albeit without explicit use of the PROV Family provenance namespace.

Overall, analysis of this scenario points to significant adversity: although there are other sources that catalog research data repositories, such as OpenDOAR⁸ and DataCite⁹, the number of repositories that adopt PROV metadata remains low. This is a cause for concern given the importance of provenance metadata standards for identifying, preserving, discovering, accessing, using and reusing information.

Therefore, the low adoption of these standards suggests an absence of effective data management practices, which could compromise the long-term integrity and usefulness of research data. It is important to disseminate and implement PROV metadata more widely to improve reliability of data repositories and ensure that the scientific community can trust the data made available (Moreau and Missier, 2013).

⁸ <https://v2.sherpa.ac.uk/opensoar/>

⁹ <https://v2.sherpa.ac.uk/opensoar/>



6 FINAL CONSIDERATIONS

From an open science perspective, Re3data is a notable online directory of data repositories that facilitates access to research data in various fields of knowledge. Such initiatives play a vital role in scientific communication by increasing the visibility of research and maximising its societal impact.

Metadata standards are fundamental to Research Data Repositories, as they enable a standardized and structured description of information (Sanchez et al., 2018). PROV Family use, in particular, enables data provenance to be recorded, providing detailed information about its origin and transformations, which strengthens the reliability and accessibility of data. Furthermore, provenance metadata promotes interoperability between systems by documenting and referencing all stages involved in the production and reuse of data.

Nevertheless, this research has certain limitations. Firstly, the analysis was based exclusively on the repositories listed in Re3data, which may not cover the full diversity of existing repositories. Additionally, while the methodology enabled the mapping of PROV standards adoption, it did not explore the underlying reasons for limited adherence in detail. Future studies could address these issues by identifying institutional, technical, and cultural barriers that hinder the implementation of provenance metadata.

The findings partially align with the study's predictions, indicating that the use of PROV Family standards in research data repositories continues to be restricted. This is consistent with literature pointing to challenges in implementing provenance metadata. However, the results also contradict some expectations, as despite the growing appreciation of open science, only a small number of PROV Family-using repositories were identified.

In terms of its theoretical and applied implications, this study reinforces the importance of PROV family in research data management, and suggests that its adoption should be encouraged through institutional policies and educational initiatives. Moreover, future research could explore strategies to promote its adoption,

as well as assess the impact of using provenance metadata on the reproducibility and integrity of scientific data. The analysis could also be expanded to include a larger number of documents within the existing repositories, allowing for a more comprehensive investigation of the presence and use of PROV Family standards.

Ultimately, the scientific community must make a continuous effort to integrate and promote provenance standards in data repositories to ensure data provenance and integrity. Effective data management and strengthened transparency and reliability of science depend on institutional support, such as that provided by the DCC, combined with awareness of the benefits of these standards (Davidson et al., 2017).

REFERENCES

ARAKAKI, Felipe Augusto; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Proveniência e contexto digital: contribuições da ciência da informação. **Palavra Chave (La Plata)**, La Plata, v. 10, n. 2, 1 abr. 2021. DOI: <http://dx.doi.org/10.24215/18539912e124>.

ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família PROV. Tese (Doutorado) – Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, 2019.

ALVES-MAZZOTTI, A. J. A “revisão bibliográfica” em teses e dissertações: meus tipos inesquecíveis - o retorno. In: BIANCHETTI, L.; MACHADO, A. M. N. (org.). **A bússola do escrever**: desafios e estratégias na orientação de teses e dissertações. Florianópolis: Editora da UFSC, 2002. p. 25-44.

BRANDT, Mariana Baptista; VIDOTTI, Silvana Aparecida Borsetti Gregório; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ZAFALON, Zaira Regina. Catalogação de metadados: descrição de metadados de negócio a partir dos princípios e objetivos bibliográficos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 3, p. 3-18, 2019 Available at: <https://periodicos.ufmg.br/index.php/pci/article/view/22627>. Accessed on: 3 fev. 2025.

DAVIDSON, S. B.; KHANNA, S.; MILO, T.; ROY, S.; TANNEN, V. Provenance views for module privacy. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2011, Athens, Greece. **Proceedings...** New York: ACM, 2011. p. 557–568. DOI: 10.1145/1989284.1989305. Available at: <https://dl.acm.org/doi/10.1145/1989284.1989305>. Accessed on: 11 mar. 2025.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.



GO FAIR. FAIR principles. [2016]. Available at: <https://www.go-fair.org/fair-principles/>. Accessed on: 11 mar. 2023.

HAYNES, David. **Metadata for Information Management and Retrieval: Understanding metadata and its use.** [S.l.]: Facet Publishing, 2018.

HEDSTROM, M. **Digital preservation: a time bomb for digital libraries.** Computers and the humanities, Países Baixos, V31, mai 1997.

JOUDREY, Daniel N.; TAYLOR, Arlene G.; WISSER, Katherine M. **The organization of information.** Fourth edition ed. Santa Barbara, California: Libraries Unlimited, 2018. (Library and information science text series).

LEBO, Timothy; SAHOO, Satya; MCGUINNESS, Deborah. **PROV-O: The PROV Ontology.** Available at: <https://www.w3.org/TR/2013/REC-prov-o-20130430/>. Accessed on: 16 dez. 2024.

MOREAU, L. e GROTH, P. (2013). Provenance: an introduction to PROV. **Synthesis lectures on the semantic web: theory and technology**, San Rafael, v. 3, n. 4, 1-129. DOI: <https://10.2200/s00528ed1v01y201308wbe007>.

MOREAU, L. *et al.* The Open Provenance Model core specification. **Future Generation Computer Systems**, Amsterdã, v. 27, n. 6, p. 743-756, 2011. Available at: <https://doi.org/10.1016/j.future.2010.07.005>. Accessed on: 11 mar. 2025.

MOREAU, L; MISSIER, P. **PROV-DM: The PROV Data Model.** (2013). Available at: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>. Accessed on: 25 dez. 2023

RE3DATA. **About - re3data.org.** Available at: <https://www.re3data.org/about>. Accessed on: 13 nov. 2024.

RE3DATA. **Metadata Standards - re3data.org Metrics.** Available at: <https://www.re3data.org/metrics/metadataStandards>. Accessed on: 7 nov. 2024.

RUSBRIDGE, C. et al. O Centro de Curadoria Digital: uma visão para a curadoria digital. In: SIMPÓSIO INTERNACIONAL IEEE SOBRE SISTEMAS E TECNOLOGIA DE ARMAZENAMENTO EM MASSA, 2005. **Anais [...]**. [S.l.]: IEEE, 2005. DOI: 10.1109/LGDI.2005.1612461. Accessed on: 11 mar. 2025.

SANCHEZ, F. A.; DA SILVA, N. B. P.; VECHIATO, F. L. Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa. **Informação & Tecnologia**, João Pessoa: UFPB, v. 5, n. 1, p. 37–51, 2019. DOI: 10.22478/ufpb.2358-3908.2018v5n1.38350. Available at: <https://periodicos.ufpb.br/index.php/itec/article/view/38350>. Accessed on: 11 mar. 2025.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, PR, v. 21, n. 2, p. 90-115, 2016. DOI: <https://doi.org/10.5433/1981-8920.2016v21n2p90>. Available

at: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Accessed on: 11 mar. 2025.

SILVA, N. R. Diretórios de dados abertos de pesquisas: o caso do Re3Data. **Ciência da Informação Express**, Lavras, v. 2, n. 5, p. 1-4, 3 maio 2021. DOI:

<https://doi.org/10.60144/v2i.2021.84>. Available at:
<https://cienciadainformacaoexpress.ufla.br/index.php/revista/article/view/84>
Accessed on: 11 mar. 2025.

VIDOTTI, S. A. B. G *et al.* Repositório de Dados de Pesquisa para Grupo de Pesquisa: um projeto piloto. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017. **Anais...** Marília - SP. Anais... Marília - SP: PPGCI, UNESP. Available at:
<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/view/388>.
Accessed on: 11 mar. 2025.



Notes and Credits of the Article

Acknowledgements: Not applicable.

Funding: Not applicable.

Conflicts of interest: Not applicable.

Ethical approval: Not applicable.

Data and material availability: The datasets analyzed during the present study are available in Re3data, accessible via the following link
<httpswww.re3data.orgmetricsmetadataStandards>
<httpswww.re3data.orgsearchquery=PROV>

Manuscript published as a preprint: The manuscript was originally published as a full paper at the National Seminar on Cataloging and Technology (SNCat) in 2024. It was subsequently subject to a new double-blind peer review, in addition to receiving adjustments and content updates.

Author Contributions:

Contributions	Silva, F. I.	Arakaki, F. A.
Study conception	X	X
Conceptualization	X	X
Methodology	X	
Data collection / Investigation	X	
Data curation	X	
Data analysis	X	
Discussion of results	X	X
Visualization (charts, tables, and others)	X	X
Original draft	X	
Final revision and editing	X	X
upervision and administration		X

License of Use

The authors grant Ciência da Informação Express - CIExpress exclusive rights for first publication, with the work simultaneously licensed under the Creative Commons



Ciência da Informação *Express*

ISSN 2675-8717

Attribution (CC BY) 4.0 International License. This license allows third parties to remix, adapt, and build upon the published work, while giving appropriate credit to the author and initial publication in this journal.

Scientific Dissemination

Does it authorize posts on CIExpress for scientific dissemination?

(X) Yes.

() No

Publisher

Federal University of Lavras (UFLA).

The ideas expressed in this article are the responsibility of the authors and do not necessarily reflect the opinion of the editors or the university.

Editor of the Communication and Scientific Dissemination Channel, *Ciência da Informação Express*

Nivaldo Calixto Ribeiro, Federal University of Lavras (UFLA).

Linguistics Reviewer

By the authors.

References Reviewer

By the authors.

English version

Ana Sílvia Bondance Cotrim Dias

History

Received on: 19/11/2024

Accepted on: 11/03/2025

Published on: 19/03/2025

