Evaluating FAIR Principles Compliance in the Capes Open Data Portal: a case study of Open Government Data

Avaliação da conformidade com os princípios FAIR no Portal de Dados Abertos da Capes: um estudo de caso de dados governamentais abertos

Francis Bento Marques (1)

Benildes Coura Moreira dos Santos Maculan (2)

Renato Rocha Souza (3)

(1) Universidade Federal de Minas Gerais (UFMG), Brasil, fbmarques@gmail.com

(2) Universidade Federal de Minas Gerais (UFMG), Brasil, benildes@gmail.com

(3) Fundação Getúlio Vargas (FGV), Brasil, rsouza.fgv@gmail.com

Abstract

The article examines the Open Government Data (OGD) movement, emphasizing applying the FAIR (Findable, Accessible, Interoperable, Reusable) principles to government data, using Brazil's Capes open data portal as a case study. The study underscores the significance of Open Government Data (OGD) in promoting transparency and democratizing access to information within the public sector. It addresses issues such as typographical errors, the absence of standardization, duplicated data, and emphasizes the necessity for a process model that integrates quality control mechanisms The paper also explores the economic and political potential of OGD, analyzing the challenge of managing open government data with high quality and interoperability. The study's results on the Capes portal show mixed compliance with the FAIR principles, with recommendations for future improvements.

Keywords: Open Government Data; FAIR Principles; Government Transparency; Capes Open Data

MARQUES, Francis Bento; MACULAN, Benildes Coura Moreira dos Santos; SOUZA, Renato Rocha. Evaluating FAIR Principles Compliance in the Capes Open Data Portal: a case study of Open Government Data. *Brazilian Journal of Information Science: research trends*, vol. 18, publicação contínua, 2024, e024035. DOI: 10.36311/1981-1640.2024.v18.e024035.

Resumo

O artigo examina o movimento Open Government Data (OGD), enfatizando a aplicação dos princípios FAIR (Findable, Accessible, Interoperable, Reusable) aos dados governamentais, usando o portal de dados abertos da Capes do Brasil como estudo de caso. O estudo ressalta a importância dos Open Government Data (OGD) na promoção da transparência e na democratização do acesso à informação no setor público. Ele aborda questões como erros tipográficos, ausência de padronização, dados duplicados e enfatiza a necessidade de um modelo de processo que integre mecanismos de controle de qualidade. O artigo também explora o potencial econômico e político dos OGD, analisando o desafio de gerenciar dados abertos do governo com alta qualidade e interoperabilidade. Os resultados do estudo no portal da Capes mostram conformidade mista com os princípios FAIR, com recomendações para melhorias futuras.

Palavras-chave: Dados abertos governamentais; Princípios FAIR; Transparência governamental; Dados abertos Capes

1 Introduction

Over recent years, the Open Government Data (OGD) movement has emerged as a critical pillar in advancing transparency and democratizing access to information within the public sector. This movement aims to make processable digital data available in an open way, allowing citizens, researchers, and developers to access previously restricted or complex information. As argued by Buckland (1991, p. 352), data can be considered a form of "information as thing" as they are tangible or digital objects that, when processed, provide relevant evidence and knowledge. In the context of open government data, these data must be properly structured and accessible in order to fulfill their role as a basis for knowledge and public transparency.

Moreover, the Open Government Data initiative plays a pivotal role in enhancing digital democracy. By allowing citizens and organizations to freely analyze and utilize these data, it fosters increased civic participation and collaboration in governance. This capability not only enhances transparency but also promotes innovation and the development of technological solutions aimed at improving the efficiency of public services and enhancing the quality of life for citizens.

The release of government data is intrinsically linked to the principles of transparency and accountability, which are essential for building a more informed and participatory society. In the Brazilian context, the Coordination of Improvement of Higher Education Personnel (Capes) plays

a crucial role in making available data for evaluating postgraduate courses in the *stricto sensu*, providing a valuable basis for research in various areas of knowledge (Roche, 2020).

An essential aspect of the Open Government Data initiative is the delicate balance between data openness and privacy protection. As highlighted by Borgesius et al. (2015), while open government data can significantly enhance transparency and public engagement, it raises concerns regarding the privacy of individuals whose data might be included in these datasets. The authors propose a framework that respects privacy rights without impeding the benefits of open data. This framework emphasizes the need to carefully consider the data types released and the conditions under which they are made available.

It is imperative to foster collaborative and multidisciplinary initiatives to address these challenges. Such efforts should involve partnerships between government agencies, academic institutions, and the private sector. These collaborations can lead to developing innovative solutions and standards that enhance data quality and usability. By working together, these diverse stakeholders can create a more robust and effective open data ecosystem, which is crucial for the advancement of digital democracy and public service improvement (Zuiderwijk & Janssen, 2014).

However, the effectiveness of opening up government data faces significant challenges related to the quality of the available data. In a study by Marques et al. (2023), problems such as typos, lack of standardization, missing records, and duplicate data in different fields were common obstacles that affected data quality and reuse. These issues highlight the importance of adhering to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, which aim to ensure that the available data are findable, accessible, interoperable, and reusable (Borgesius et al., 2015).

Implementing data repositories that follow FAIR principles, especially in big data contexts, is a central challenge, as De Castro et al. (2020) discussed. In addition, producing high-quality open government data requires a process model incorporating quality control activities, as Penteado et al. (2021) proposed. These studies highlight the need for a structured and methodical approach to improve the quality and utility of open data.

Against this backdrop, this research seeks to investigate whether the research data made available by Capes meets the FAIR principles and to verify the quality of the data available.

2 Open Government Data

The age of digitization and the rapid development of internet technologies have transformed how governments manage and make data available. Dymora et al. (2018) discuss the emergence of open data as a significant new area, highlighting the publication of databases from various public institutions and private sectors on the global internet. This data, available at no extra cost, allows information to be analyzed and processed freely, representing a milestone in transparency and governance.

Nikiforova (2021) explores the economic potential of OGD, pointing out that despite its value in the millions, the effective reuse of data still needs to be improved. The research suggests the need for more "intelligent" government data to drive a sustainable economy and ICT innovations, acting as a bridge of creativity in developing new ecosystems in Industry 4.0 and Society 5.0.

Ruijer et al. (2019) approach OGD from a political perspective, arguing that Open Government Data is a strategic resource that organizations may be hesitant to share. They develop an analytical framework to study the politics of OGD, showing that organizational responses to these initiatives can range from compliance to active resistance. This study provides insights into why governments may release datasets in specific policy domains but not others, producing a "strategically opaque transparency" (Ruijer et al., 2019, p. 263).

Ruijer and Meijer (2019) investigate OGD as an innovation process, identifying different phases in the open data innovation process. Using a living lab in a province in the Netherlands, the study analyzes how interventions stimulated the use of OGD and raised awareness within the government. However, several mechanisms inhibited the achievement of OGD ambitions, indicating that the challenge for open data innovation lies in the scale of its provision and use.

On the other hand, Ding et al. (2012) address the challenges associated with public access to government data, emphasizing the heterogeneity and complexity of the public information ecosystem. They report a community consensus on the architecture of the interconnected open government data ecosystem, reviewing the key technologies and challenges for opening, connecting, and reusing those data. This study highlights the adoption of linked data-based

solutions by leading practitioners, such as Data.gov in the US and data.gov.uk in the UK, to deliver an open and incremental ecosystem that interconnects providers, consumers, and contributors of open government data.

Open Government Data represents a significant advancement in sharing and using information, promoting transparency and innovation. However, fully realizing their potential depends on a balanced approach considering the opportunities and the challenges inherent in this evolving field.

3 Open Government Data in Brazil

In the Brazilian context, Open Government Data has become a tool to promote transparency and citizen participation. Breitman et al. (2012) discuss the current status of OGD in Brazil, summarizing the lessons learned from publishing Brazilian government data as linked data. This study highlights the importance of adapting open data practices to the specificities of the Brazilian context. Matheus, Ribeiro, and Vaz (2012) explore new perspectives for e-governance in Brazil, emphasizing the adoption of OGD. They argue that, despite a relatively late start, Brazilian state and federal governments are creating new standards and practices that should spread throughout the Brazilian public sector.

Oliveira et al. (2016) performed a comprehensive analysis of OGD portals in Brazil, evaluating a range of criteria including data volume and quality. The study revealed considerable diversity in the size and quantity of datasets, with the CSV format being the most commonly used. This study provides a detailed overview of OGD in Brazil at the time.

Corrêa et al. (2014) presented a more comprehensive evaluation of transparency portals by surveying 20 Brazilian municipalities. They use Brazil's Access to Information Law requirements as evaluation criteria, thus covering the principles of OGD. The results show a gap between local transparency portals and the effective implementation of OGD principles, leading to a misconception that transparency portals are disconnected from the open government initiative.

As elucidated by Gonçalves and Araujo (2023), the effective management of OGD necessitates a strategic approach, focusing on critical success factors such as robust information systems, stakeholder engagement, and policy frameworks. In the Brazilian context, particularly in the State of São Paulo, the successful implementation of OGD initiatives has demonstrated the potential for these data to revolutionize public administration. By ensuring that government data are open, effectively managed, and utilized, OGD initiatives can lead to more informed decision-making, foster public trust, and encourage civic engagement, reinforcing the foundations of a democratic society.

The analysis of Open Government Data in Brazil reveals an evolving scenario with unique challenges and opportunities. Research highlights the importance of adapting OGD practices to the Brazilian context, the emergence of new perspectives for e-governance, diversity in open data portals, and the need to align transparency portals with open government principles (Matheus et al., 2012; Corrêa et al., 2014). This overview suggests a promising path for transparency and citizen participation in Brazil, although significant challenges remain.

In addition to open data initiatives, Brazil has been making significant strides in aligning its practices with the FAIR principles, with the GO FAIR Brazil initiative serving as a key example of this effort. As highlighted by Sales et al. (2020), the establishment of the GO FAIR Brazil Office marks an important milestone in coordinating thematic networks aimed at implementing the FAIR principles across strategic sectors such as health, agriculture, and nuclear energy. The first active network, GO FAIR Brazil-Health, demonstrates how FAIR principles are being applied in practice, promoting interoperability and reusability of research data. These initiatives not only strengthen data governance in Brazil but also bring the country in line with global open science standards, contributing to the improvement of the quality and accessibility of both scientific and governmental data.

4 FAIR Principles

The FAIR principles were established by the participants of the Lorentz Workshop entitled Jointly Designing a Data FAIRport, held in 2014 and maintained by GO FAIR. They were created

to improve the findability, accessibility, interoperability, and reuse of digital research objects for humans and machines. Wilkinson et al. (2016) pioneered formalizing these principles, highlighting the urgent need to improve the infrastructure supporting the reuse of scholarly data. They argue that the FAIR principles can serve as guidelines for those wishing to increase the reusability of their data.

The FAIR Principles represent a milestone in scientific data management, establishing essential guidelines for effectiveness and efficiency in research. These principles are divided into four fundamental categories: Findable, Accessible, Interoperable, and Reusable, each of which plays a vital role in promoting open science and maximizing the value of research data (Wilkinson et al., 2016; Lamprecht et al., 2020; Silva et al., 2023). Buckland (1991) argues that information, understood as a "thing" to be effectively accessed and reused, must be appropriately structured. The implementation of the FAIR principles aligns with this perspective, ensuring that data are findable, accessible, interoperable, and reusable, thus maximizing their value as a source of information and evidence.

The Findable principle is the first step, essential to making data more accessible and usable. It ensures that humans and computer applications easily locate and identify data. In this principle, the data is achieved by assigning unique and persistent identifiers to datasets and creating rich and detailed metadata. Metadata should include clear and precise information about the content, context, and structure of the data, allowing users to quickly find the information they seek (Wilkinson et al., 2016; Zupancic et al., 2021).

The Accessible principle guarantees that data can be efficiently and reliably accessed once located. In this principle, the data must be stored in reliable and accessible repositories using standards-based, open, and free protocols. In addition, it is essential that data remain accessible in the long term, even after initial publication, and that access conditions are clearly defined and communicated (Wilkinson et al., 2016; Jacobsen et al., 2020).

The Interoperable principle refers to the ability of data to be integrated and used simultaneously with other data sets, systems, and applications. To achieve interoperability, data must be structured and formatted in accordance with open standards and controlled vocabularies.

This principle allows different systems and tools to understand, process, and use the data effectively (Lamprecht et al., 2020; Molinaro, 2021).

Finally, the Reusable principle focuses on the need to document data comprehensively to ensure that it can be reused effectively in different contexts and by different users. This principle includes providing clear information on the license, provenance, standards adopted, and any other relevant information that may influence or facilitate data reuse. The efficient reuse of data maximizes its value and impact, promoting innovation and collaboration in scientific research (Lamprecht et al., 2020; Hempelmann et al., 2021).

Together, the FAIR Principles establish a robust and comprehensive framework for data management in scientific research. These Principles facilitate data discovery and access and promote effective integration and reuse across different domains and platforms. By following these principles, researchers and institutions can significantly contribute to transparency and collaboration in the scientific field, strengthening open science and innovation (Lamprecht et al., 2020).

Zupancic et al. (2021) discuss how data and metadata can be prepared in laboratories to align with FAIR principles. This chapter provides practical guidance for researchers and laboratories on how to make data more findable, accessible, interoperable, and reusable, emphasizing the importance of proper formatting and guidance. Applying FAIR principles is not only limited to data but also extends to research software. Barker et al. (2022) introduce the FAIR Principles for Research Software (FAIR4RS), adapting the original FAIR principles to reflect the specific needs of research software. They outline the content and context of the FAIR4RS Principles, providing groundwork for discussions about their adoption.

In his 2023 study, Delfino uses the FairDataBR tool, which has simplicity and intuitiveness as its main features, to make it accessible to many users, regardless of their level of technical expertise. This tool lets users quickly assess whether a dataset meets the FAIR criteria, supporting effective data management following international best practices.

The application of the FAIR Principles in Open Government Data is exemplified by the study of De Figueiredo et al. (2019), which conducts a systematic analysis of data management

across different contexts, including research, corporate, and government data. This study highlights how the FAIR Principles can be integrated and adapted to enhance government data management, emphasizing the importance of interoperable and reusable data structures. The research demonstrates that applying the FAIR Principles to government data facilitates the discovery and access to these data and promotes effective integration and reuse across various contexts, contributing to more transparent and accountable governance.

Lamprecht et al. (2020) discuss how the FAIR principles can be adapted for the research software, highlighting the desired characteristics of the research software that go beyond the FAIR principles. They analyze how existing principles can be applied directly to software and how they need to be adapted or reinterpreted.

The FAIR principles represent an important milestone in the management of the research data and software, promoting practices that increase the efficiency and effectiveness of scientific research (Wilkinson et al., 2016). They encourage the creation of a more collaborative and transparent research ecosystem where data and software can be easily shared and reused. Implementing these principles involves challenges but offers significant opportunities to advance open science and collaborative research (Lamprecht et al., 2020; Zupancic et al., 2021).

4 Methodology

The methodology of this study has an applied nature and exploratory objectives and uses a quantitative-qualitative approach to analysis. The quantitative evaluative analysis utilizes the FairDataBR software to assess datasets containing information on Brazilian postgraduate programs available on the Capes Open Data portal. The software assesses the platform's data based on the FAIR criteria, generating a detailed report that shows the degree of compliance. This report identifies patterns, gaps, and opportunities for improvement in data management in a qualitative assessment that discusses the implications of the results for data management practices and suggests strategies for improving adherence to the FAIR principles

4 Quantitative analysis: Software application

The FairDataBR assessment thoroughly analyzes the datasets available on the CAPES portal to assess their adherence to the FAIR principles. This evaluation provides valuable insights into the current state of open government data management, identifying potential areas for improvement and aligning with international best practices in open science and data transparency.

The evaluation process using the FairDataBR consists of a set of 20 questions based on the FAIR principles. The set of questions is primarily multiple-choice, providing a systematic and standardized approach to the evaluation. Depending on the previous answers, some questions are followed by open questions, allowing for a more in-depth and contextualized analysis of certain aspects of the data. After completing the questionnaire, the FairDataBR tool presents the results graphically to offer a holistic, clear, and intuitive view of the level of adherence of the CAPES data repository to the FAIR principles. This graphical representation helps interpret the results and highlights specific areas for potential improvement, providing a solid basis for future action actions in open government data management (Barker et al., 2022; Delfino, 2023).

Here are the most significant results for each FAIR principle: Findable, Accessible, Interoperable, and Reusable.

2.1 Findable principle

In the context of the FAIR Principles, the "Findable" principle is crucial to ensuring that data is easily locatable and identifiable by humans and machines. This principle emphasizes assigning unique identifiers to datasets and providing rich, detailed metadata that facilitates data discovery and access (Wilkinson et al., 2016; Zupancic et al., 2021). Götza (2023) highlights that the systematic application of the FAIR principles, especially findability, helps address the reproducibility crisis in science and enables scientific data to be available for verification and use beyond its original purpose. Adopting this principle is fundamental to transparency and efficiency in research, allowing the data used in scientific publications to be made available to the community for verification, reproduction, and the derivation of new results. Adherence to the Findable

principle is an essential step towards maximizing the value and impact of data in scientific research (Götz, 2023).

The application of the FairDataBR questionnaire to assess the Findable principle is shown in Table 1.

Table 1- Questions and answers on the Findable Principle

Id	Question	Answer	Comment
F1	Do (meta)data sets have a unique, global and persistent identifier?	Web address	Data are only identified by URLs (web addresses), they may not be considered persistent, as URLs can change over time.
F2	Are the data sets described with metadata?	Simple metadata	Basic metadata are found, such as a title and brief description, but without in-depth details or an organized structure.
F3	Is the identifier included in all the records / metadata files that describe the data?	Yes	A URL with the metadata is found in every collection. The identifier is easy to find and exists in all datasets.
F4	Can the digital resource be found on web search engines?	Yes	The resource is easily found by searching the web.
F5	Is the (meta)data set published in a repository?	General Public	The repository is accessible to anyone, with no specific domain or institution restrictions.

Source: Research data (2023)

2.2 Accessible principle

Within the FAIR Principles, the Accessible principle is crucial in ensuring that the data is easily accessible by humans and machines. This principle emphasizes the importance of making the data accessible under clearly defined conditions using open and standardized protocols (Zupancic et al., 2021). Jacobsen et al. (2020) argue that although the FAIR Principles are formulated at a high level and can be interpreted and implemented in different ways, they are essential to support convergence in the implementation choices that are widely accessible and reusable for true interoperability between systems.

Table 2 illustrates the application of the FairDataBR questionnaire on the Accessible principle, providing an analysis of how the datasets on the CAPES portal adhere to this specific principle.

Table 2- Questions and answers on the Accessible Principle

Id	Question	Answer	Comment
A1	How accessible are (meta)data?	Accessible to the Public	All data and metadata are publicly available and unrestricted.
A2	Are (meta)data accessible online without the need for intermediary protocols or specialized tools from the moment access is allowed?	Online file download	Data and metadata can be downloaded directly from the portal as files (CSV, XLS, PDF).
A3	Is it possible to access the (meta)data set by the identifier provided?	Yes	Identifiers consistently lead directly to the corresponding data and metadata.
A4	Are the metadata available even when the dataset is no longer accessible?	No	Metadata becomes inaccessible when the data are not available because metadata depends on the URL of the data.
A5	Is the protocol (e.g. HTTP, SAML, OAI-PMH) open, free and universally implementable?	No	There is no possibility of accessing the data and metadata through the protocols identified, with or without the need for software or paid licenses.
A7	Is it possible to download the (meta)data sets?	Yes	Both datasets and general metadata can be downloaded from the portal. The metadata can include information such as title, description, etc. and the data is in CSV and/or XLS formats.

Source: Research data (2023)

2.3 Interoperable principle

In the context of the FAIR Principles, the Interoperable principle is essential to ensuring that data can be integrated and used in different systems and contexts. This principle emphasizes the importance of structuring data to be compatible and can be combined with other datasets using open standards and controlled vocabularies (Wilkinson et al., 2016; Zupancic et al., 2021). A study by Molinaro et al. (2021) highlights the need for interoperable systems, especially in automated contexts, and the importance of open standards to prevent adverse effects on the user experience.

Thus, adherence to the Interoperable principle is crucial for the effectiveness and impact of data on scientific research, promoting integration and collaboration between different fields of knowledge (Molinaro, 2021).

Table 3 demonstrates the application of the FairDataBR questionnaire on the Interoperable principle, providing a detailed assessment of the ability of the CAPES portal datasets to integrate effectively with other systems and contexts.

Table 3- Questions and answers on the Interoperable Principle

Id	Question	Answer	Comment
I1	Are (meta)data sets available in preferred formats?	Yes	The datasets are available in common formats: CSV, XLS, RDF, and the metadata in PDF format.
I2	Are the datasets structured using a metadata schema or data models approved by the community?	Explicit Schema	There is an explicit metadata schema, but it is not necessarily a declared community standard (Dublin Core, DDI (Data Documentation Initiative), MARC (Machine-Readable Cataloging)).
I3	Are (meta)data sets linked to other (meta)data using identifiers?	No Links	No links in the metadata, such as DOIs or URIs, connect different datasets or metadata.
I4	Did you use terminology control resources/instruments to represent the data or data sets?	No resources/ instruments	There is no evidence of the use of terminology control tools, controlled vocabularies, thesauruses, ontologies or other terminology control tools.

Source: Research data (2023)

2.3 Reusable principle

In the context of the FAIR Principles, the Reusable principle is essential to ensuring that the data are not only accessible and interoperable but can also be effectively reused in different contexts and by different users. This principle emphasizes the importance of documenting the data clearly and comprehensively, including information on the license of use, provenance, and standards adopted (Lamprecht et al., 2020; Wilkinson et al., 2016). A study by Hempelmann et al. (2021) discusses the application of the FAIR Principles in climate service information systems, highlighting the need for metadata models for efficient consumption and reuse due to the large

volumes of data. These, therefore, require different, more efficient processing approaches for the data. Thus, adherence to the "Reusable" principle is critical to maximizing the value and impact of the data on scientific research, promoting the efficient and responsible reuse of data.

Table 4 of the study illustrates the application of the FairDataBR questionnaire on the Reusable principle, providing an assessment of the capacity of the datasets on the CAPES portal.

Table 4- Questions and answers on the Reusable Principle

Id	Question	Answer	Comment
R1	Are the (meta)data sets licensed?	Yes	The datasets and metadata have clear and accessible information about the license of use, Creative Commons Attribution Open Data.
R2	What is the license to use the (meta)data sets?	Open	Data includes Creative Commons Attribution Open Data licenses.
R4	The (meta)data sets have detailed provenance?	Yes	The portal provides detailed and transparent information on the provenance of datasets and metadata, including details on collection, authorship, modifications, and updates.
R5	Are the (meta)data sets accompanied by or linked to a description of the origin of the workflow that produced the (meta)data?	No Information of Provenance	There need to be details about the process that led to the data creation, including the specific steps, tools used, and decisions made during data collection and processing.
R6	Do the (meta)data comply with standards relevant to the Domain?	No standards used	There needs to be more information on a specific research domain standard established for educational, scientific or government data or the use of recognized standards such as Dublin Core, METS, MODS, DataCite, etc.

Source: Research data (2023)

4 Qualitative data analysis

Based on the answers provided for the Findable principle of the FAIR Principles, it is possible to see that the CAPES portal has some positive characteristics regarding accessibility and

data identification. The presence of a Web Address for the datasets (F1) eases access, but it does not offer the robustness of a persistent identifier. The Simple Metadata (F2) shows that the data are described, but these metadata have room for enrichment and detailing. The fact that each dataset has an identifier in its metadata (F3) and the ability to find the resource in web search engines (F4) are strengths that enhance the visibility and accessibility of the data. Finally, publishing the data in a repository accessible to the General Public (F5) reflects a commitment to transparency and data availability to a wide range of users. However, improving the metadata quality and adopting more robust and persistent identifiers would be beneficial to align fully with the FAIR Principles.

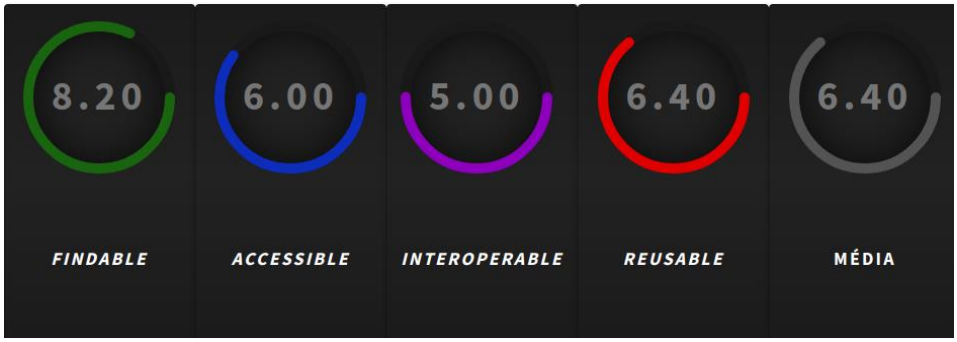
The responses related to the Accessible principle of the FAIR Principles indicate a satisfactory level of data accessibility on the CAPES portal, with some areas requiring improvement. Data availability to the General Public (A1) is a strong point, ensuring the data are accessible to a broad audience. The "Download file online" option (A2) eases direct access to the data, allowing users to obtain the information practically. The fact that the datasets can be accessed via their identifiers (A3) is a positive aspect that enhances the ease of locating and accessing the data. However, metadata unavailability when the data are not accessible (A4) and a missing open and universal protocol (A5) limit the long-term accessibility and interoperability of the data. The ability to download the metadata sets (A7) reinforces accessibility. However, it is crucial to implement more open protocols and maintain the metadata to ensure the continuous and effective accessibility of the data.

The analysis of the answers relating to the Interoperable principle reveals essential aspects of integrating and communicating data on the CAPES portal. Confirming that datasets are available in preferred formats (I1) is a positive indication, suggesting that data can be easily used and integrated into different systems and contexts. Using an "Explicit Schema" (I2) to structure the data demonstrates a conscious effort to maintain organization and clarity, making it easier for stakeholders to understand and use the data. However, the absence of "Internal/External Links" (I3) to connect the datasets to other metadata limits interoperability, as linking data is fundamental for effective integration into a more comprehensive data ecosystem. In addition, the lack of terminology control resources or vocabularies (I4) represents a significant gap because these tools

are crucial for ensuring the consistency and standardization of the terms used, essential elements for interoperability between different systems and contexts. Controlled vocabularies can restrict the type of value allowed when filling in data fields to restrict and standardize the values. Therefore, while there are positive points, there is a clear opportunity to improve data interoperability on the portal by implementing links and terminology control tools.

The answers relating to the Reusable principle of the FAIR Principles indicate that the CAPES portal has attributes that favor data reuse, but some areas need attention. The fact that the datasets are licensed (R1) and the license is open (R2) represents fundamental aspects that promote data reuse because they ensure that users are clear about the rights and conditions of use. A detailed provenance of the datasets (R4) is another positive point because it provides users with important information about the origin and context of the data, which is essential for responsible and informed reuse. On the other hand, the absence of information on the origin of the workflow that produced the metadata (R5) is a limitation because this information is crucial to understanding the context and quality of the data, which is essential for reuse. In addition, the need to align with the domain standards (R6) can restrict data reuse in specific contexts because the domain standards facilitate the understanding and integration of data in different areas. Therefore, although the portal reflects a commitment to data reuse, it would be beneficial to improve the provenance documentation and adopt the domain standards to maximize the usefulness and applicability of the data in various contexts.

Figura 1 – Faculdade de Filosofia e Ciências da UNESP, Campus de Marília



Source: Research data (2023).

The evaluation of the CAPES' open data portal, using the FairDataBR tool and the FAIR Principles, reveals a mixed picture of adherence to these essential principles for effective open data management. Figure 1 shows the highest score for the Findable principle, at 8.20, which indicates the portal's strong ability to ensure data are easily locatable and identifiable. This result suggests that the CAPES portal efficiently provides unique identifiers and adequate metadata to make it easier for users to access and discover data.

On the other hand, the scores in the Accessible, Interoperable, and Reusable principles, with respective values of 6.00, 5.00, and 6.40, indicate areas needing significant improvement. The median scores for Accessible and Interoperable suggest challenges in the ongoing accessibility of data and the ability to integrate it into different systems and contexts. The slightly higher Reusable score indicates an acceptable portal adequacy for data reuse. However, there are still opportunities for improvement, mainly according to domain standards and provenance documentation. The overall average of 6.40 represents a balance between strengths and areas for development, highlighting the need for continued efforts to improve the open data management on the CAPES portal, aligning more closely with international best practices in open science and data transparency.

3 Final considerations

The application of FairDataBR in assessing the quality of open data provided by CAPES marks a significant advancement in evaluating and improving government data management in Brazil. Open government data, as demonstrated by the CAPES portal, plays a pivotal role in supporting research and promoting transparency. Access to high-quality, openly available data, aligned with the FAIR principles, is crucial for fostering innovation, facilitating scientific collaboration, and enhancing accountability in public administration. Well-organized and accessible data serve as a valuable foundation for knowledge creation and act as a catalyst for transparency and innovation. This emphasis on data quality and transparency is a key driver for the sustainable development of open science practices in Brazil, underscoring the importance of efficient data management in the digital age.

Open government data, as exemplified by the Capes portal, is crucial for advancing research and promoting transparency in Brazil. Access to open, quality data, aligned with the FAIR Principles, is vital to boosting innovation, facilitating scientific collaboration, and promoting greater accountability and efficiency in public management. By ensuring that data are findable, accessible, interoperable, and reusable, the CAPES portal strengthens the research ecosystem. It contributes to democratizing knowledge and building a more informed and engaged society. This commitment to the quality and transparency of open government data is a crucial step towards sustainable development and the consolidation of open science practices in Brazil, reinforcing the vital role of data management in the digital age.

Acknowledgements

The authors gratefully acknowledge the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (second author: process 307765/2023-7), as well as the Sandwich Doctorate Scholarship (Call Nº 6/2024 PDSE) from the Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

Barker, Michelle, et al. "Introducing the FAIR Principles for research software". *Scientific Data*, vol. 9, n.º 1, out 2022, doi: https://doi.org/10.1038/s41597-022-01710-x.

Borgesius, Frederik Zuiderveen, et al. “Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework.” *Berkeley Technology Law Journal*, vol. 30, no. 3, 2015, pp. 2073–131. JSTOR, https://www.jstor.org/stable/26377585.

Breitman, Karin, et al. "Open government data in Brazil". I*EEE Intelligent Systems*, vol. 27, n.º 3, maio 2012, pp. 45-49, doi: https://doi.org/10.1109/mis.2012.25

Buckland, Michael K. Information as thing. Journal of the American Society for information science, v. 42, n. 5, p. 351-360, 1991

Corrêa, Andreiwid Sheffer, et al. "Transparency portals versus open government data". *the 15th Annual International Conference*, ACM Press, 2014, doi: https://doi.org/10.1145/2612733.2612760.

De Castro, Leonardo, et al. "A fair allocation approach to the ethics of scarce resources in the context of a pandemic: The need to prioritize the worst-off in the Philippines". *Developing World Bioethics*, set 2020, doi: https://doi.org/10.1111/dewb.12293.

De Figueiredo, Glaucia Botelho, et al. "Aligning DMBOK and Open Government with the FAIR Data Principles". *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 13-22, doi: https://doi.org/10.1007/978-3-030-34146-6_2.

Delfino, Samyr Santos. *Interoperabilidade de dados em saúde: OpenFairEHR-um modelo conceitual no contexto dos princípios FAIR*. 2023. Universidade Federal da Paraíba, João Pessoa, tese de Doutorado em Ciência da Informação. *UFPB*.

Ding, Li, et al. "Linked Open Government Data [Guest editors' introduction]". *IEEE Intelligent Systems*, vol. 27, n.º 3, maio 2012, pp. 11-15, doi: https://doi.org/10.1109/mis.2012.56.

Dymora, Paweł, et al. "Open data – an introduction to the issue". *ITM Web of Conferences*, vol. 21, 2018, p. 00017, doi: https://doi.org/10.1051/itmconf/20182100017.

Gonçalves, Vagner Mendonça, e José Braz de Araujo. "Critical Success Factors for Open Government Data Management Information Systems in a Public Body in the State of São Paulo". *SBSI '23: XIX Brazilian Symposium on Information Systems*, ACM, 2023, doi: https://doi.org/10.1145/3592813.3592816.

Götz, Andy. "The fair principles: Trusting in fair data repositories". *Open Access Government*, vol. 39, n.º 1, jul 2023, pp. 262-63, doi: https://doi.org/10.56367/oag-039-10749.

Hempelmann, Nils et al. *FAIR principles for climate services information systems*. Copernicus Meetings, 2021. doi: https://doi.org/10.5194/ems2021-488

Jacobsen, Annika, et al. "FAIR Principles: Interpretations and Implementation Considerations". Data Intelligence, vol. 2, n.º 1-2, jan 2020, pp. 10-29, doi:10.1162/dint_r_00024.

Lamprecht, Anna-Lena, et al. "Towards FAIR principles for research software". *Data Science*, vol. 3, n.º 1, jun 2020, pp. 37-59, doi: https://doi.org/10.3233/ds-190026.

Marques, Francis Bento; Maculan, Benildes Coura Moreira dos Santos; Souza, Renato Rocha. Levantamento e tratamento dos dados de pesquisa da Ciência da Informação à luz dos dados abertos da CAPES. In: Natália Bolfarini Tognoli; Ana Cristina de Albuquerque; Brígida Maria Nogueira Cervantes. (Orgs.). Organização e representação do conhecimento em diferentes contextos: desafios e perspectivas na era da datificação. – Londrina: ISKOBrasil, 2023, pp. 374-383. https://isko.org.br/wp-content/uploads/2023/06/livro-isko-Brasil_23.pdf.

Matheus, Ricardo, et al. "New perspectives for electronic government in Brazil". *the 6th International Conference*, ACM Press, 2012, doi: https://doi.org/10.1145/2463728.2463734.

Molinaro, Marco, et al. "Supporting fair principles in the astrophysics community: the european experience." *arXiv preprint arXiv:2111.14468.* 2021.

Nikiforova, Anastasija. "Smarter Open Government Data for Society 5.0: Are Your Open Data Smart Enough?" *Sensors*, vol. 21, n.º 15, jul 2021, p. 5204, doi: https://doi.org/10.3390/s21155204.

Oliveira, Marcelo Iury S., et al. "Open Government Data Portals Analysis". dg.o '*16: 17th International Digital Government Research Conference*, ACM, 2016, doi: https://doi.org/10.1145/2912160.2912163.

Penteado, Bruno Elias, et al. "Modelo de infraestrutura para publicação de dados abertos educacionais conectados de qualidade". *Workshops do Congresso Brasileiro de Informática na Educação, Sociedade Brasileira de Computação*, 2021, doi: https://doi.org/10.5753/wcbie.2021.217316.

Roche, Dominique G., et al. "Open government data and environmental science: a federal Canadian perspective". *FACETS*, vol. 5, n.º 1, jan 2020, pp. 942-62, doi: https://doi.org/10.1139/facets-2020-0008.

Ruijer, Erna, et al. "The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency". *The American Review of Public Administration*, vol. 50, n.º 3, dez 2019, pp. 260-74, doi: https://doi.org/10.1177/0275074019888065.

Ruijer, Erna, e Albert Meijer. "Open Government Data as an Innovation Process: Lessons from a Living Lab Experiment". *Public Performance & Management Review*, vol. 43, n.º 3, fev 2019, pp. 613-35, doi: https://doi.org/10.1080/15309576.2019.1568884.

Sales, Luana et al. GO FAIR Brazil: a challenge for brazilian data science. **Data Intelligence**, v. 2, n. 1-2, p. 238-245, 2020.

Silva, Fabiano Couto Corrêa da , et al. "Diagnóstico dos Repositórios de Dados no Brasil". *Brazilian Journal of Information Science: research trends*, vol. 17, ago 2023, p. e023031, doi: https://doi.org/10.36311/1981-1640.2023.v17.e023031.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific Data*, vol. 3, n.º 1, mar 2016, doi: https://doi.org/10.1038/sdata.2016.18.

Zuiderwijk, Anneke, e Marijn Janssen. "Open data policies, their implementation and impact: A framework for comparison". *Government Information Quarterly*, vol. 31, n.º 1, jan 2014, pp. 17–29, doi: https://doi.org/10.1016/j.giq.2013.04.003.

Zupancic, Klemen; Pavlek, Tea; Erjavec, Jana. *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*. Wiley & Sons, Incorporated, John, 2021.

Received: 21/08/2024 Accepted: 30/10/2024