Open data from the Federal Institutes of Education: analysis of themes suggested by the Office of the Comptroller General

Dados abertos dos Institutos Federais de Educação: análise de temas sugeridos pela Controladoria-Geral da União

Yuri Bento Marques[1], Marcello Peixoto Bax[2], Geraldo Lopes Júnior[3], Gustavo Henrique Silva de Souza[4], Eduardo Cardoso Melo[5]

[1] Federal Institute of the North of Minas Gerais (IFNMG), Belo Horizonte, MG, Brazil. ORCID: https://orcid.org/0000-0003-3414-2325

[2] Federal University of Minas Gerais (UFMG), Belo Horizonte, MG, Brazil. ORCID: https://orcid.org/0000-0003-0503-3031

[3] Federal Institute of the North of Minas Gerais (IFNMG), Belo Horizonte, MG, Brazil. ORCID: https://orcid.org/0000-0002-2243-2336

[4] Federal Institute of the North of Minas Gerais (IFNMG), Belo Horizonte, MG, Brazil. ORCID: https://orcid.org/0000-0002-4046-9669

[5] Federal Institute of the North of Minas Gerais (IFMG), Belo Horizonte, MG, Brazil. ORCID: https://orcid.org/0000-0002-1323-5859

Mail to/Autor para correspondência/Correo a: Yuri Bento Marques, yuri.marques@ifnmg.edu.br

Abstract

Introduction: Opening government databases can positively contribute to democracy, transparency and social control, as it provides opportunities for citizens to access, reuse, modify and redistribute non-confidential information from public bodies. However, not all data of interest to society are currently published in open formats. This study verifies whether the data referring to the Federal Institutes of Education, Science and Technology, which are commonly requested from the Office of the Comptroller General by civil society, are being cataloged in the Brazilian Open Data Portal. **Method:** Quantitative research was conducted on the Brazilian Open Data Portal with the support of the documentary research technique, which allowed us to analyze the datasets published by the Federal Institutes. An index was created to analyze each of the 14 collected themes, which originated a quantitative indicator used to compare the results of each Federal Institute involved in the research. **Results:** It was identified that 73.70% of the Federal Institutes cataloged at least one dataset on the Brazilian Open Data Portal. However, there were strikingly several gaps in the availability of open data by these institutions, in addition to the low adherence to opening data in accordance with the recommendations of the Office of the Comptroller General. **Conclusions:** Studies such as the one carried out in this paper contribute to the Public Administration, especially considering the context of Federal Institutes, to have subsidies that indicate points for improvement with regard to society's demands on transparency and governance.

Keywords: Access to Public Information; Open Government Data; Transparency; Federal Institutes of Education.

Resumo

Introdução: A abertura das bases de dados governamentais pode contribuir positivamente com a democracia, transparência e controle social, pois oportuniza ao cidadão o acesso, reutilização, modificação e redistribuição livre das informações não sigilosas dos órgãos públicos. Porém, nem todos os dados de interesse da sociedade se encontram publicados em formatos abertos no presente momento. Esse estudo verifica se os dados referentes aos Institutos Federais de Educação, Ciência e Tecnologia, que são comumente solicitados à Controladoria-Geral da União pela sociedade civil, estão sendo catalogados no Portal Brasileiro de Dados Abertos. **Método:** Foi conduzida uma pesquisa quantitativa no Portal Brasileiro de Dados Abertos com suporte da técnica de pesquisa documental, o que permitiu analisar os conjuntos de dados publicados pelos Institutos Federais. Criou-se um índice para analisar cada um dos 14 temas coletados, o qual originou um indicador quantitativo utilizado na comparação dos resultados de cada Instituto Federal envolvido na pesquisa. **Resultados:** Identificou-se que 73,70% dos Institutos Federais catalogaram pelo menos um conjunto de dados no Portal Brasileiro de Dados Abertos. No entanto, verificaram-se contundentemente diversas lacunas na disponibilização de dados abertos por parte dessas instituições, além da baixa aderência à abertura de dados em conformidade às recomendações da Controladoria-Geral da União. **Conclusão:** Estudos como o realizado neste artigo contribuem para que a Administração Pública, especialmente considerando o contexto dos Institutos Federais, tenham subsídios que indiquem possíveis pontos de melhorias no que tange às demandas da sociedade sobre transparência e governança.

Palavras-chave: Ética; Pesquisa; Ciência da Informação; Comitê de Ética; Registro de consentimento.

INTRODUCTION

In 1946, the United Nations General Assembly recognized access to information as a universal human right, so everyone has the fundamental right to collect, transmit and publish information. In order to make their actions transparent and open, governments must provide access to their information. In the 1980s, in the midst of the Free Software movement, the Open Culture or Open Movement emerged to support Transparency and Access to Information (Cardoso, 2019, p. 32).

In both the public and private spheres, it is no news that organizations need to manage their information resources in order to improve their productivity, competitiveness, and performance. In order to encourage the adoption and implementation of resources that help with this management, institutions need to study and recognize which organizational and cultural factors contribute to this process (Bergeron, 1996). An organization that manages to integrate knowledge creation, senses, and decision-making is at an advantage, as it can carry out actions based on a shared understanding of the context, with intelligence, creativity, and even cunning (Choo, 2006).

Specifically in the public sphere, the application of the concepts of open government data (OGD) means that organizations make their databases available on the web in an open and structured format. This openness contributes positively to strengthening democracy, transparency and social control, as it allows citizens to freely access, reuse, modify and redistribute non-sensitive information from public bodies (Possamai, 2016). In addition, OGDs help public institutions to maximize the integration and interoperability of their data, obtain new information to support decision-making, and promote government-society collaboration in the creation of new services and applications based on open data (Possamai, 2016). To support interoperability and data reuse, the W3C (World Wide Web Consortium [W3C], 2017) recommends reusing consolidated Ontologies/Vocabularies, with the aim of increasing consensus on concepts and terms among users and reducing redundancy and ambiguity.

However, the lack of standards for publishing open government data hinders transparency, analysis, and understanding by citizens. Furthermore, little data of interest to society is published in open and easily accessible formats. As a result, new requests for access to information are generated via the Access to Information Act (Gonçalves, 2017), eventually overloading those involved, as they need to reorganize their work routine to meet the demand for information received. We must also consider an important aspect addressed by Monteiro and Lucas (2019) regarding the lifespan of the data made available, since the offer in a shared environment needs to be accompanied by planning so that it can be constantly updated in order to continuously meet the demands of society.

Lack of structuring, low semantics and unavailability of information of interest to society in OGD is common to Federal Institutes of Education, Science, and Technology. As Klein (2018, p. 77) points out, no Federal Institute complies with all the transparency criteria defined by good practice legislation in its open data plans. According to Nunes (2018), the Federal Network of Professional, Scientific and Technological Education (RFEPCT) is made up of thirty-eight Federal Institutes of Education, Science, and Technology, created by Law No. 11,892/2008: the Federal Technological University of Paraná, the Celso Suckow da Fonseca and Minas Gerais Federal Centers of Technological Education, twenty-two Technical Schools linked to Federal Universities and the Pedro II College. In addition, the fact that the Federal Institutes are made up of a multi-campus structure allows them to expand the scope of their actions, offer courses and strengthen productive, social and cultural arrangements.

There has been an increase in studies focused on the systematization of information through semantic data, such as the studies by Britto, Ruy, and Azevedo (2020) on the semantic integration of public expenditure, Chen et al. (2020) on the ontological representation of proteins and Zaidan and Bax (2013) on adding value to data by applying the semantic technology of Linked Open Data (LOD). The relevance of this anchors the latent need to investigate the contributions of semantic technologies applied in the context of OGD at Federal Institutes.

Aleixo (2020) states that although research into open government data is currently in evidence, it is not analyzing how technologies impact on data access and availability. The author also adds that Brazilian legislation itself is unclear about the procedures that must be carried out to guarantee access to and publication of open data. That said, the complex structure of the RFEPCT, with its 661 units, demands open management based on transparency and accountability. The aim of this work is to analyze the databases opened by Federal Institutes in relation to the topics suggested by the Comptroller General of the Union (CGU), such as the publication of information on students, courses, disciplines, among others, in order to identify whether this data is properly cataloged on the Brazilian Open Data Portal. We hope to help the Public Administration make the necessary adjustments and create an accessible environment with greater transparency.

The remaining of the article is organized as follows: Section 2 presents the theoretical concepts involving open government data that support this article, as well as studies carried out by other researchers in the field. The methodology used to construct this work is explained in Section 3, as are the techniques used to obtain the results. Section 4 presents and discusses the results obtained in the article. Finally, Section 5 concludes the study.

THEORETICAL BASIS

Semantic Web

The Semantic Web is considered an extension of the modern World Wide Web, in which documents are explained to machines through semantically standardized metadata and techniques. In this way, both human users and machines enjoy more open, structured and connected data (Alencar, Xavier, Chaves, & Souza, 2018).

Hypertext Markup Language (HTML) was developed for humans to consume the content of Internet sites, but it doesn't serve this purpose for machines. For more sophisticated processing of websites by machines, various models are used to structure the data. The formal data and syntax model RDF (Resource Description Framework) is among the most popular. It makes it possible to create metadata that enables Semantic Web applications to consume information from websites autonomously (Breitman, 2005).

The lack of a systematic process for creating and reusing ontologies that describe the information contained in open data is one of the main challenges to the interoperability of systems (Santarem Segundo, 2015).

Open Government Data

According to Silva (2018), although the Constitution of the Republic, in articles 5, 37 and 216, already introduces access to public information, information policy as a government policy for government documents has only been in evidence since 2011, due to Brazil's adherence to the Open Government Partnership (OGP) and the approval of Law No. 12,527 of 2011, better known as the Access to Information Law (ATI). The OGP aims to promote government budget transparency practices at a global level, as well as access to information and participation by society. The purpose of the ATI is to guarantee the fundamental right of access to information about public administration data.

In order to ensure and facilitate society's access to data generated or held by the federal executive branch, the National Open Data Infrastructure was established through Normative Instruction No. 4 of April 12, 2014. It is responsible for disseminating materials with guidelines for opening up data and drawing up the Open Data Plan (ODP) for public information (Silva, 2018).

With the exception of sensitive data, public bodies must make all their databases available in an open format. However, this requires structuring, hosting and integrating the data, which can be a complex and costly process. In view of this, it is necessary to involve society and apply a priority matrix to define which databases need to be made available. Furthermore, in order for the day-to-day actions carried out by the government body to be aimed at opening up their respective databases, a culture of open data needs to be cultivated within the body, as well as activities aimed at engaging those involved. Finally, the continuous support of the Information Technology sector in the process of preparing the ODP helps to incorporate the opening up of databases, even during the development stage of the institution's software (Controladoria-Geral da União [CGU], 2020).

Several studies have been found in the literature investigating the availability of open government data. Therefore, only those studies that are closest to the research proposed in this article are presented below.

Klein (2018) investigated the levels of active and passive transparency, good practices and open data plans of universities and federal institutes. To achieve this goal, the researcher created a transparency indicator based on Brazilian legislation and transparency indicators from the Office of the Comptroller General and the Federal Public Prosecutor's Office. The researcher drew the following conclusions after applying the indicator: (1) no Federal University or Federal Institute is 100% transparent; (2) the higher the budget per enrolment, the less transparent it tends to be; (3) the higher the level of transparency, the fewer the complaints and requests for information and appeals.

Carossi, Pedan, and Teixeira Filho (2020) investigated the most frequently asked questions in the Electronic System of the Citizen Information Service for public Federal Higher Education Institutions in Brazil. This research used Grounded Theory as its methodology and resulted in the identification of categories and themes that would be a priority for subsidizing the Open Data Plans of these institutions.

Lima, Cunha, and Maior Filho (2017) carried out a case study at the Federal Institute of Education, Science and Technology of Paraíba (IFPB) in order to identify the weaknesses and potential of Open Government at the institution. As a result, they found that although the IFPB makes various budget data available to society, on the other hand, some of this data is difficult to locate on the website and, due to inadequate formats, difficult to reuse.

Lemos (2017) identified a lack of maturity and standardization in the process of opening and maintaining open data in Federal Institutes. As a result, he proposed a process for publishing OGD for the RFEPCT based on Business Process Management (BPM), using the BPMN (Business Process Model and Notation).

Silva (2018) developed a metric for the reuse of open government data made available by Brazilian Federal Public Institutions. As a result of applying the metric to a federal public administration body, the author concluded that the reuse of open data was incipient in relation to the amount of data made available.

As for the availability of open data by Federal Institutes of Education, Science and Technology, according to Nunes (2018), the main motivation for these institutions to draw up an Open Data Plan (ODP) was to prevent control bodies from charging for the availability of such data. In addition, the lack of Information Technology systems and the sensitization of the institution's civil servants are factors that hinder the preparation and implementation of the ODP.

[Macedo and Lemos](#) (2021) state that the literature lacks scientific studies, whether theoretical or practical, that broaden the understanding of the current scenario of open government data, whether in the Brazilian context or worldwide. Therefore, while progress has been made in investigating the availability of open government data to society, there is a lack of evaluation of the themes, i.e., what information is made available through open data at the Federal Institutes.

METHOD

This section presents a contextualization of the analysis tool designed as a data source for the article, as well as the techniques used to carry out the proposed activities and the characteristics of the analysis model created. It is important to note that the bibliographic survey of the entire theoretical basis used in this study was carried out on the Google Scholar platform with research related to the central theme in question, analyzing scientific articles, books, theses, and dissertations. Legislative artifacts were obtained by accessing government repositories.

Contextualizing the analysis tool

To draw up a ODP, public institutions must hold discussions with the agency's departments, define those responsible for its creation and implementation, list all databases, allow society to inform which databases will be opened and, finally, draw up a prioritization matrix informing which databases will be made available (Controladoria-Geral da União [CGU], 2020).

The National Open Data Infrastructure Steering Committee (CGINDA) is made up of invited members from civil society, the academic sector, Ministries, the General Secretariat of the Presidency of the Republic, the Civil House of the Republic and the Brazilian Institute of Geography and Statistics (IBGE) (Aleixo, 2020). The prioritization of the opening of databases demanded by society is evidenced in the following CGINDA resolution:

> Art. 1 In order to promote a culture of public transparency, according to item II of § 2 of art. 5 of Decree no. 8.777, of May 11, 2016, the databases to be made available must be prioritized and justified, in the Open Data Plans - ODP, according to their potential in terms of public interest, considering what is applicable:
>
> I - the degree of relevance to the citizen;
>
> II - stimulating social control;
>
> III - the legal obligation or commitment to make the data available;
>
> IV - the data refers to strategic government projects;
>
> V - the data demonstrates direct and effective results of public services made available to citizens by the state;
>
> VI - its capacity to foster sustainable development;
>
> VII - the possibility of promoting business in society;
>
> VIII - the most requested data in passive transparency since the beginning of Law No. 12,527, of November 18, 2011 - Access to Information Law, (Resolução nº 3, de 13 de outubro de 2017, 2017, s/p)(Resolution No. 3, 2017, s/p).

In order to assist the process of publishing databases, the CGU suggests a prioritization matrix for datasets from universities and federal institutes based on data that is commonly requested by civil society, making up a total of 77 pieces of information (items) in 14 themes. All these items and themes are shown in Table 1, which provides the information contained in Controladoria-Geral da União [CGU] (2020) in a graphically adapted form.

Nº	Tema	Informações
1	Students	Number of students, Number of students per course, Number of dropouts, Enrolments closed, Number of students graduated
2	Degrees	Degree name, Department head, E-mail address, Phone number, Modality (in-person or distance learning), Degree type (Bachelor's degree or Licentiate's degree)
3	Classes	Subject, Workload, Number of openings, Prerequisites, Related Courses
4	Professors per Department	Name of professor, Name of department, Workload, Line of research
5	Faculty weekly working hours	Name of professor, Name of department, Schedules, Statistics
6	Statistics	Average IRA (Academic Performance Index), Failure rate by class, Percentage of Quota Filling, Patents (filing date, application number, title, inventors, subject), attendance by type of service, extension activity, library collection, list of service points, public facilities
7	Student aid	Type of aid, Expenditure on the program, Number of beneficiaries
8	Database of theses, monographs, and articles	Title, Abstract, Authors, Keywords, Document date
9	Scientific production	Production title, Abstract, Authors, Keywords, Document date, URL, Journal, Conference
10	Projects involving external funding	Technical Cooperation Agreements (National and International), Decentralized Execution Terms, Exchange, Agreements outside SICONV
11	Incubated businesses	Number of incubated businesses, Purposes, areas of activity, List of junior businesses, List of outsourced employees
12	Construction work	Ongoing work, Project completion, Cost per project
13	Public procurement	Positions, Number of openings, Class, SIAPE Code, Position code
14	List of properties for rent	Property type, Location, Destination, Number of rooms, Rent, Condominium fee, Property value, Status (rented/vacant)

Table 1. Database prioritization matrix suggested by the CGU.
Source: Adapted from Controladoria-Geral da União [CGU] (2020).

Data collection and analysis model

This study carried out a quantitative survey using the documentary research technique on the Brazilian Open Data Portal (PBDA), analyzing the datasets made available by the Federal Institutes in relation to the themes suggested by the CGU for opening databases. Although it was found that six Federal Institutes made more datasets available on their own institutional website, only the datasets catalogued on the PBDA were analyzed, since only "those that are catalogued on the aforementioned Portal will be considered as effectively published databases" (Controladoria-Geral da União [CGU], 2020, p. 10).

Between April 15 and May 1, 2021, 543 sets of open data from 38 Federal Institutes available on PBDA were analyzed. In order to verify whether the themes and information suggested by the CGU are being made available through open data by each of the Federal Institutes, the title, description, data dictionary and content of each of the sets were analyzed, which are available at https://tinyurl.com/ogdif2021.

In order to analyze the 77 items (information) of each of the 14 themes, an index was established in the form of a 3-point scale to measure the degree of transparency of the information requested, with "0" (zero) being assigned to the item with non-existent data availability, "1" (one) to the item with partial data availability and "2" (two) to the item with complete data availability. Thus, an indicator was created to compare the 14 themes, considering the score obtained from the sum of the values of their respective items. The interpretation of the items by means of indices therefore allowed for a parameterized analysis of each of the themes. To calculate the indicator for each theme, the average score per item was adjusted to a score from "0" to "10", where "0" indicates that none of the items in the respective theme had their data made available, even partially, and "10" indicates that all the items in the respective theme had their data made available in full.

RESULTS

After the data collection process, it was found that 73.70% of the Federal Institutes had cataloged at least one database on the Brazilian Open Data Portal. Of these, 60.71% did not describe their datasets using data

dictionaries. In addition, some dataset links had a "404 error", i.e., the page that should contain the file could not be found. As a result, these datasets were not analyzed. A considerable number of datasets had no descriptions, so the analysis of the theme was limited to the title and content of the data in their respective files. No Federal Institute made ontologies available or referenced them in their open datasets. Finally, several datasets were out of date or did not contain any date relating to the information made available; despite this, analysis of the theme was carried out, since temporal analysis is not part of the scope of this research. Table 2 shows the results of the 14 themes consolidated by Federal Institute.

Institute	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IFAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFAL	0	5	0	5	0	2,22	10	0	0	5	0	0	8	0
IFAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFAP	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFB	10	0	0	8,75	0	3,33	0	5	5	7,5	0	0	0	0
IFBA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFBAIANO	2	0	6	5	0	1,11	0	0	0	0	0	0	0	0
IFC	4	0	0	7,5	0	2,2	10	0	0	5	0	0	0	0
IFCE	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFES	2	3,64	0	7,5	0	1,11	6,67	0	0	0	0	0	4	0
IFF	0	3,18	0	7,5	0	0	0	0	0	0	0	0	8	0
IFFARROUPILHA	4	4,09	4	5	0	1,11	0	0	0	0	0	0	0	0
IFG	8	6,82	5	5	0	1,67	0	0	0	0	0	1,67	6	0
IFGOIANO	1	3,18	0	8,75	0	0	0	0	0	0	0	0	0	0
IFMA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFMG	4	4,09	0	5	0	2,22	0	0	0	2,5	0	10	4	0
IFMT	10	3,18	0	8,75	0	2,22	0	0	0	0	0	1,67	0	0
IFMS	8	4,09	0	8,75	0	1,11	0	0	0	0	8,75	10	6	0
IFNMG	8	6,82	0	2,5	0	1,11	10	0	0	2,5	0	0	4	0
IFPA	4	0	0	1,25	0	3,33	5	0	0	5	0	0	5	0
IFPB	10	4,55	2	7,5	0	1,11	6,67	0	0	0	0	0	0	0
IFPE	0	0	0	0	0	0	0	0	1,25	1,25	0	0	0	0
IFPI	0	0	0	5	0	0	0	0	0	0	0	0	0	0
IFPR	10	5	3	5	0	2,22	10	0	0	5	0	0	2	0
IFRJ	2	5,91	0	0	0	1,11	0	0	0	0	0	0	0	0
IFRN	10	0,91	0	7,5	8,33	2,22	3,33	0	0	0	0	0	0	0
IFRO	0	5,91	3	5	0	1,11	0	0	0	2,5	0	1,67	0	0
IFRR	0	0	0	7,5	0	0	0	0	0	0	0	0	0	0
IFRS	0	0	0	0	0	1,11	0	0	0	0	0	0	4	0
IFS	8	5	4	7,5	0	2,22	0	0	0	2,5	0	6,67	4	0
IFSC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFSP	8	5	0	7,5	0	2,22	0	0	6,25	2,5	0	0	0	0
IFSPE	10	6,82	0	7,5	0	1,11	0	0	0	0	0	0	0	0
IFSUDESTEMG	10	5	6	7,5	0	1,67	6,67	0	0	0	0	0	0	0
IFSUL	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFSULDEMINAS	4	5,45	4	7,5	0	1,11	6,67	0	0	0	0	10	0	0
IFTM	8	4,09	4	7,5	0	3,33	6,67	0	6,25	0	0	6	0	0
IFTO	0	0	0	0	0	0	0	0	0	0	0	0	0	0

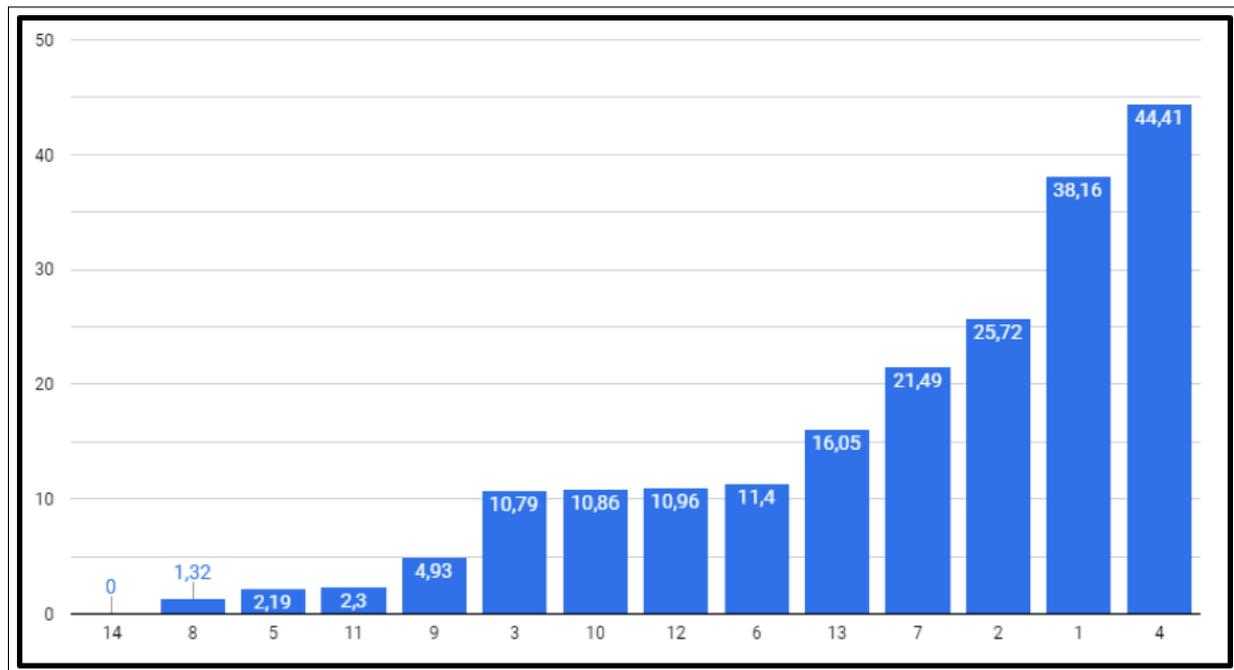**Table 2.** Availability score by themes suggested by the CGU
Note: 1 - Students; 2 - Degrees; 3 - Classes; 4 - Professors by Department; 5 - Faculty weekly working hours; 6 - Statistics; 7 - Student aid; 8 - Database of theses, monographs, and articles; 9 - Scientific production; 10 - Projects involving external funding; 11 - Incubated businesses; 12 - Construction work; 13 - Public procurement; 14 - List of properties for rent.

The average scores (within the range of 0 to 10) for the topics by region were as follows: Midwest region 2.27 (standard deviation: 0.93); Southeast region 2.16 (standard deviation: 0.80); Northeast region 1.21 (standard deviation: 1.15); South region 1.12 (standard deviation: 1.23); North region 0.51 (standard deviation: 0.73).

None of the items were provided by all 38 Federal Institutes. Specifically with regard to the items pertaining to the topic of "properties for rent", all the Federal Institutes neglected to present the data. In addition, the items "e-mail" and "phone number" under the heading "degrees", the items "number of places" under the heading "subjects", the item "library collection" under the heading "statistics" and the item "keywords" under the heading "scientific production" were not provided by any Federal Institute. The items "Name of professor" and "Name of department" in the "Teachers by department" theme, the items "Name" and "Degree (Bachelor's or

Licentiate's)" in the "Degrees" theme and the item "extension activity" in the "Statistics" theme were made available by at least half of the Federal Institutes.

As can be seen in Figure 1, themes "4 (Professors by Department)" and "1 (Students)" showed the best results in terms of data availability, but still had margins of less than 50% of all Federal Institutes.

Figure 1. Percentage of availability of topics suggested by the CGU
Note: 1 - Students; 2 - Degrees; 3 - Classes; 4 - Professors by Department; 5 - Faculty weekly working hours; 6 - Statistics; 7 - Student aid; 8 - Database of theses, monographs, and articles; 9 - Scientific production; 10 - Projects involving external funding; 11 - Incubated businesses; 12 - Construction work; 13 - Public procurement; 14 - List of properties for rent.

Discussion

The results of this study indicate low availability of data on the themes and items suggested by the CGU. The lack of any catalogued information by some Federal Institutes, as well as the absence of data dictionaries and descriptions of the datasets, made it difficult to analyze the data in relation to the themes and items. These results are in line with Klein (2018) assertion that no Federal Institute is 100% transparent, considering the availability of its data as recommended by the CGU. In addition, the low publication of data and lack of organization of information have a negative impact on the consumption and reuse of open government data, as evidenced in other related studies (e.g. Lima et al., 2017; Silva, 2018).

In this sense, the absence of ontologies in the Federal Institutes' open data can hinder its interoperability and retrieval in a semantic environment, as shown by Santarem Segundo (2015). Furthermore, one of the causes of the low availability of certain open data is probably related to the lack of Information Technology systems that facilitate its extraction and provision on the Web (Nunes, 2018).

The high variability of the indicators shows the discrepancy between institutions, even within the same region. In other words, while one federal institute made information available on various items, another did not publish any information at all. There was also a lack of synchronization between the publication of open data on the institutional website and the PBDA. This reality of the lack of maturity and standardization of Federal Institutes in the process of opening and maintaining their open data was observed by Lemos (2017).

It is possible to group the three most available topics ("Professors by Department", "Degrees" and "Students") as the "Teaching" category, and some of the least available topics ("Scientific Production, "Bank of theses, monographs, and articles") as the "Research" category, signifying the absence of integrated Information Technology systems at the Federal Institutes. In other words, one system is used to manage teaching data, which has functionalities for opening up data, and another system – without this capability – is used to control research data.

Lastly, the items under the heading "Buildings for rent" were not made available by any Federal Institute. Considering that the CGU suggested the same topics for universities and federal institutes, we believe that data on this topic is more frequent for federal universities, due to the size and structure of these institutions.

CONCLUSION

This article analyzed the open data sets of the Federal Institutes of Education, Science and Technology, based on the suggestions contained in the CGU's Manual for Preparing Open Data Plans (ODPs) (2020), about database themes that could be opened up to society. These suggestions corroborate the public interest in better transparency and governance of public institutions.

Considering the objective proposed by the study, several gaps were found in the provision of open data by the Federal Institutes, as well as low adherence to opening up data in accordance with the CGU's recommendations. There was also considerable variation in the publication of data by institution, as well as a lack of synchronization between what is presented in the PBDA and in institutional repositories. In functional terms, it is important to understand the relevance of providing the information dealt with here for public management, analyzing, as a priority, the administrative and technological means of complying with these recommendations.

Since this study indicates that data on the topics suggested by the CGU is not widely available, it is worth noting, as a limitation of the research, that it was not checked whether the Federal Institutes have this data to be published in another way, or whether the specific information did not apply to the context of each institution. For example, the topic "incubated companies" was made available by only 2.3% of the Federal Institutes, however, if an institution has not yet started any business incubation process, either due to lack of demand or unavailability of infrastructure or legal framework, it means that it does not have the data on this topic to make available to society.

In order to continue the investigation into the opening up of open data by Federal Institutes, further research is needed to: (1) understand the reasons for the low availability of open data by some Federal Institutes; (2) analyze and categorize all the topics that the Federal Institutes identified through public hearings and reported in their respective PDAs; and (3) analyze and categorize which topics and information are available through open data from the Federal Institutes, i.e., unlike the scope of this research, consider topics that were not suggested by the CGU.

AGRADECIMENTOS

REFERENCES

Aleixo, D. V. B. S. (2020). O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil [Tese de doutorado]. *Instituto Universitário de Lisboa*. Retrieved from https://repositorio.unesp.br/handle/11449/191686

Alencar, A., Xavier, D., Chaves, L., & Souza, D. (2018). Publicação e consumo de dados abertos conectados acadêmicos. *Principia*, *42*(2018), 136-145. Retrieved from https://periodicos.ifpb.edu.br/index.php/principia/article/view/1988

Bergeron, P. (1996). Information resources management. *Annual Review of Information Science and Technology*, *31*(1996), 263-300.

Breitman, K. (2005). *Web semântica: a internet do futuro.* Curitiba: LTC.

Britto, G. C., Ruy, F. B., & Azevedo, C. L. B. (2020). Um ambiente para integração de dados abertos relativos à despesa pública. *Ontobras*, *13*(2020). Retrieved from http://ceur-ws.org/Vol-2728/paper13.pdf

Cardoso, P. H. (2019). Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação [Dissertação de mestrado]. *Instituto Universitário de Lisboa*. Retrieved from https://ridi.ibict.br/handle/123456789/1039

Carossi, D. F., Pedan, J. M., & Teixeira Filho, J. G. (2020). Utilização da metodologia "Grounded Theory" na identificação de informações para compor um plano de dados abertos de instituições federais de ensino superior. *iSys-Revista Brasileira de Sistemas de Informação*, *10*(2), 171-192. Retrieved from http://seer.unirio.br/index.php/isys/article/view/6152

Chen, C., Huang, H., Ross, K. E., Cowart, J. E., Arighi, C. N., Wu, C. H., & Natale, D. A. (2020). Protein ontology on the semantic web for knowledge discovery. *Scientific Data*, *7*(337), 1–12. Retrieved from https://www.nature.com/articles/s41597-020-00679-9

Choo, C. W. (2006). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International journal of information management*, *16*(5), 329–340. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/0268401296000205

Controladoria-Geral da União [CGU]. (2020, jun). *Manual de elaboração de planos de dados abertos (PDAs).* Brasília, DF: STPC/CGU. Retrieved from https://repositorio.cgu.gov.br/handle/1/46702

Gonçalves, V. (2017). *PSO: uma ontologia de domínio sobre servidores públicos federais* (Dissertação de mestrado, Universidade Federal de Pernambuco, Brasil). Retrieved from https://repositorio.ufpe.br/handle/123456789/29774?mode=full

Klein, C. R. (2018). *Transparência da gestão pública das autarquias federais de ensino vinculadas ao Ministério da Educação - MEC* (Dissertação de mestrado, Universidade de Brasília, Brasil). Retrieved from https://repositorio.unb.br/handle/10482/32187

Lemos, J. M. M. (2017). *Um processo para publicação de dados abertos em institutos federais baseado em BPM* (Dissertação de mestrado, Universidade Federal de Pernambuco, Brasil). Retrieved from https://repositorio.ufpe.br/handle/123456789/25351

Lima, J. P., Cunha, S. F., & Maior Filho, U. G. S. (2017). *Governo Aberto: um estudo de caso sobre o Instituto Federal da Paraíba.* João Pessoa, PB, Brasil: Encontro Brasileiro de Administração Pública. Retrieved from http://plone.ufpb.br/ebap/contents/documentos/0073-088-governo-aberto-um-estudo-de-caso-sobre-o-ifpb.pdf

Macedo, D. F., & Lemos, D. L. d. S. (2021, Apr.). Dados abertos governamentais: iniciativas e desafios na abertura de dados no brasil e outras esferas internacionais. *AtoZ: novas práticas em informação e conhecimento*, *10*(2), 14. Retrieved from https://doi.org/10.5380/atoz.v10i2.77737 doi: 10.5380/atoz.v10i2.77737

Monteiro, G., & Lucas, E. R. d. O. (2019, June). Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. *AtoZ: novas práticas em informação e conhecimento*, *8*(1), 13. Retrieved from https://doi.org/10.5380/atoz.v8i1.67253 doi: 10.5380/atoz.v8i1.67253

Nunes, V. K. A. (2018). *Plano de dados abertos nos institutos federais de Educação, Ciência e Tecnologia* (Dissertação de mestrado, Universidade Federal de Viçosa, Brasil). Retrieved from https://www.locus.ufv.br/handle/123456789/22591

Possamai, A. J. (2016). *Dados abertos no governo federal brasileiro: desafios de transparência e interoperabilidade* (Tese de doutorado, Universidade Federal do Rio Grande do Sul, Brasil). Retrieved from https://www.lume.ufrgs.br/handle/10183/156363

Resolução nº 3, de 13 de outubro de 2017. (2017). *Aprova as normas sobre elaboração e publicação de Planos de Dados Abertos, conforme disposto no Decreto nº 8.777, de 11 de maio de 2016.* Brasília, DF. Retrieved from https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes/resolucao-n-o-3-de-13-de-outubro-de-2017

Santarem Segundo, J. E. (2015). Web semântica, dados ligados e dados abertos: uma visão dos desafios do brasil frente às iniciativas internacionais. *Tendências da Pesquisa Brasileira em Ciência da Informação*, *8*(2), 219–239. Retrieved from https://revistas.ancib.org/index.php/tpbci/article/view/359

Silva, P. N. (2018). *Dados Governamentais Abertos: Métricas e indicadores de Reuso* (Tese de doutorado, Universidade Federal de Minas Gerais, Brasil). Retrieved from https://repositorio.ufmg.br/handle/1843/BUBD-AYNG4U

World Wide Web Consortium [W3C]. (2017). *Data on the web best practices: W3C recommendation 31 January 2017.* Retrieved from https://www.w3.org/TR/dwbp/

Zaidan, F. H., & Bax, M. P. (2013, Aug.). Linked open data como forma de agregar valor às informações clínicas. *AtoZ: novas práticas em informação e conhecimento*, *2*(1), 44. Retrieved from https://doi.org/10.5380/atoz.v2i1.41319 doi: 10.5380/atoz.v2i1.41319

NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	Yuri Bento Marques	Marcello Peixoto Bax	Geraldo Lopes Júnior	Gustavo Henrique Silva de Souza	Eduardo Cardoso Melo
Concepção do manuscrito	X	X	X	X	X
Escrita do manuscrito	X	X	X	X	X
Metodologia	X	X	X	X	
Curadoria dos dados	X	X			
Discussão dos resultados	X	X	X	X	
Análise dos dados	X	X			X

Disponibilidade de Dados Científicos da Pesquisa

Os conteúdos subjacentes ao texto da pesquisa estão disponíveis sob demanda dos pareceristas.