Article

em questão
E-ISSN 1808-5245

Open government data: maturity diagnosis model for quality data published on the web

Dirceu Flavio Macedo [I]

[I] Postgraduate Program in Information Science at the Federal University of Espírito Santo (UFES), Vitória, ES, Brazil; dfmacedo@gmail.com; https://orcid.org/0000-0003-4837-3618

Daniela Lucas da Silva Lemos [II]

[II] Library Science Department of the Federal University of Espírito Santo (UFES), Vitória, ES, Brazil; daniela.l.silva@ufes.br; https://orcid.org/0000-0003-1565-7366

Abstract: Increased interest in open government data initiatives stems from long debates about state modernization. This study investigates the challenges imposed on public administration regarding data publishing from an open data perspective to propose a maturity diagnosis model for open data portals, aiming at the provision of in-depth, consistent, efficient, and transparent government information. The research comprised three phases. Firstly, we conducted a pre-test on open data portals from the United States, the United Kingdom, Brazil and Ireland, motivated by studies on the history of the open data movement as well as the current panorama of initiatives, evolution stages and challenges, which defined the fundamental analytical dimensions for the proposed maturity model. Secondly, we developed a maturity diagnosis model for open government data portals. Finally, the model was applied and validated on the pre-tested portals. Results from the diagnosis instrument can guide public administration in realizing a more efficient and responsible data governance, benefiting the government, the open data movement and civil society.

Keywords: open government data; information organization; maturity diagnosis model; quality data; data governance

1 Introduction

The popularization of information and communication technologies (ICTs) and the Internet enabled the emergence of a favourable environment for questioning and reflecting on the State's role towards its citizens. In this context, governments required felt pressed to strengthen their relationship with the community via new approaches, which inevitably lead to taking measures for a more efficient model. Following an increasingly digital government, from which emerged the open government data movement (Davies, 2010; Davies; Bawa,

2012; Gray, 2014; Juana-Espinosa; Luján-Mora, 2019; Luna-Reyes; Najafabadi, 2019; Wang; Shepherd, 2020).

The open data movement is a major effort towards knowledge sharing and democratic expansion, benefiting both society and governments in terms of data source accessibility (Janssen *et al*., 2017; Gascó-Hernández *et al*., 2018; Luna-Reyes; Najafabadi, 2019) which increases accountability, transparency, government efficiency, innovation, anti-corruption initiatives, and civic empowerment, as well as encourages information use and reuse for interest actions (Open Government Partnership, 2011; Attard *et al*., 2015; Juana-Espinosa; Luján-Mora, 2019; Wang; Shepherd, 2020).

Given this scenario, data open projects have been analysed by several interested parties (Kučera *et al*., 2015; Shepherd *et al*., 2019), including organizations, activists, governments, and researchers, to investigate whether the precepts established by the open data movement are in fact bringing the benefits and social transformations claimed and applying the best practices for data publishing.

Studies in the fields of information science (IS) and computer science (CS) have been studying information organization in digital environments aiming to increase the scope of interoperability and integration of online databases (Zeng; Qin, 2016; Wilkinson *et al*., 2016; Zeng, 2019; Guizzardi, 2020; Lemos; Souza, 2020; Martins *et al*., 2022).

Linked Open Data (LOD) principles, which design the connection between databases from heterogeneous sources to create knowledge networks for a given domain to be explored by a community, are also used (Bizer; Heath; Berners-Lee, 2009; Machado; Souza; Da Graça Simões, 2019). Research has highlighted the use of metadata standards, ontologies, and controlled vocabularies in structuring qualified databases and in representing semantic and conceptual relations to improve navigation and information search and retrieval processes in digital reality (Wilkinson *et al*., 2016; Guizzardi, 2020; Lemos; Souza, 2020).

In addition to the technical challenges associated with implementation, principles and formats to be considered when developing the information

infrastructure (Borgman, 2010) required for data opening, governments must address other barriers linked to making their data available (Kučera *et al*., 2015; Segundo, 2015; Pinto; Almeida, 2020), especially related to digital governance strategies, demanding a greater interdisciplinary effort to ensure government data availability online without imposing risks to their integrity and reliability (Shepherd *et al*., 2019).

Horvath (2017) argues that governance must use instruments such as indicators and evaluations to achieve management goals. Anne *et al*. (2017) suggest using a diagnosis model containing degrees of ability to assess the maturity of the element under investigation as an evaluation method. Within the scope of data governance, such maturity assessments should be performed periodically to measure progress and prioritize next steps (Federal Data Strategy, 2020). Agencies should therefore integrate governance models better suited to their reality (Horvath, 2017).

A maturity model analyses all aspects of the procedures and operations related to government data infrastructure, which includes data governance, available resources, systems and tools, data analytics, skills, team data capacity and culture, and compliance with data laws and policies (Federal Data Strategy, 2020).

In short, the open government data movement has data governance issues (Shepherd *et al*., 2019; Macedo; Lemos, 2021), lacking attention to qualification and training and data publishing by public legislators, managers and public servants (Anne *et al*., 2017; World Wide Web Foundation, 2018). As the World Wide Web Foundation (2018, p. 23) states, “the open data movement needs to go beyond experimentation and implement the fundamental policies and practices to support a sustainable culture of open data across governments.”

Clarifying the combined use of principles from IS (functioning of the available forms of information organization and representation), CS (how publishers implement aspects of data and metadata interoperability between institutions and their information systems), and public management (incorporation of governance concepts and processes) can help to develop a governance model to guide data management authorities towards a higher quality web data publishing.

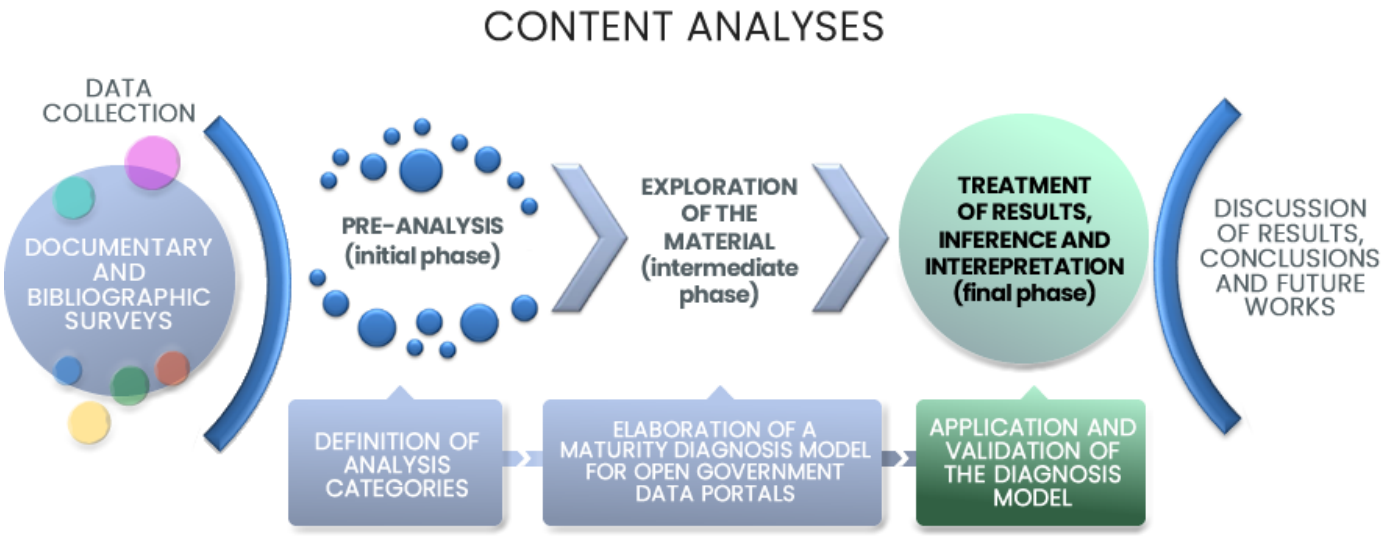
In this perspective, this study proposes a comprehensive maturity diagnosis model for open data portals capable of being applied by governments. We expect the model to create a more trustworthy and credible picture for the maturity of the available government data, pointing out possible gaps and opportunities to be explored by data governance in an ongoing effort to improve open data.

2 Methodology

Our research is interdisciplinary, borrowing from information science and computer science, fields with consolidated methodology relevant to open data practices, especially regarding organization and representation of information and knowledge in digital environments. Public administration studies were also analysed for further considerations on information and technology in data governance and policy.

Data were collected and analysed by means of content analysis (Bardin, 2016), a research tool that adopts a set of methods and techniques to create categories of analysis for comprehensive understanding of a given phenomenon. Figure 1 illustrates the entire methodological process, including the stages outlined in Bardin's method (2016).

Figure 1 - Methodological process



Source: Elaborated by the authors.

The research comprised three phases:

a) pre-analysis (initial phase), in which we defined the categories of analysis;
b) exploration of the material (intermediate phase), in which we developed the maturity diagnosis model for open government data portals, and;
c) treatment of results, inference and interpretation (final phase), in which we applied and validated the diagnosis model.

Around 100 publications met the inclusion criteria defined for the bibliographic and documentary surveys, of which 30 were selected after reading the abstract and keywords. Duplicates and studies not explicitly related to open government data initiatives and challenges were excluded.

Based on the literature review and analysis of the challenges faced by open government data initiatives (Shepherd *et al*., 2019; Pinto; Almeida, 2020), and to better understand the reality of data governance in this context, we conducted a pre-test throughout May 2021 on open government data portals from the US (United States, 2023), the UK (United Kingdom, 2023), Brazil (Brazil, 2023), and Ireland (Ireland, 2023). This pre-test sought to verify the maturation and subsequent standardization of the analysis categories to be used in the subsequent phases (Macedo; Lemos, 2021).

As for the portals to be investigated and validated by the proposed model, we established that they should:

a) have a centralized federal portal, and;
b) be OGP active members, an organization aligned with the Sebastopol principles.

Among the eligible countries, we selected:

a) The US and the UK for being pioneers both in joining the open data movement (Davies, 2010; Gray, 2014) and in creating open data portals and adhering to LOD principles (Bauer; Kaltenböck, 2011). Nonetheless, research shows that they are falling in maturity rankings (World Wide Web Foundation, 2018; Macedo; Lemos, 2021);
b) Brazil, because it is one of the OGP founding countries and to understand the maturity level of its initiative and;

c) Ireland, due to its great innovations in open government data (Blank, 2019).

The pre-test showed that all open government data initiatives analysed still fall short of a reasonable data governance scenario favourable to informational value and data transparency, thus corroborating the challenges highlighted and discussed by Macedo and Lemos (2021). More specifically, they lack visibility regarding the public data policies in force, their scope, social involvement, actions to be performed, and transparency in activity performance.

Analysis of the portals' technical documentation analysed showed that despite their efforts to disclose the metadata standards adopted for data semantic representation, the initiatives neither indicate the location of the annotation models used nor inform if the data publishing follows, for example, the LOD standard, mentioning only the 5-star ranking importance (Bizer; Heath; Berners-Lee, 2009).

Regarding the technological aspect, all the analysed portals use Comprehensive Knowledge Archive Network (CKAN) as a metadata modelling platform, suggesting a predilection on the part of international initiatives for this free tool. However, how the tool resources are used in each portal varies significantly from one initiative to another.

Issues related to training, qualification, legislation, and dissemination of data culture itself need further attention and maturation. For example, disclosure of courses and events held to discuss open data are outdated or non-existent.

Challenges were identified and organized based on data collection and analysis, which cast a critical look at certain open data portals and consolidated the pre-test phase. As an analysis tool, this pre-testing served to consolidate the challenges as analytical dimensions analysis categories according to Bardin's method (2016). Such dimensions were considered safe indicators to elaborate a maturity diagnosis model for open government data portals. Thus, data extraction, analysis, and interpretation (in the materials selected for this study) produced the following analytical categories:

a) IT infrastructure - adapt the organization's ICTs to support the applications and information systems responsible for providing open data capacity;
b) accessibility and usability - ensure that any individual can access and use the data by offering adequate and unobtrusive portal navigation resources and using accessible language for a better user experience;
c) political and social - identify and disclose datasets of great social impact, using the portal as a tool for showcasing data impact and monitoring the implementation of open government data initiatives;
d) economic - value initiatives that support the economic and sustainable viability of the portal, such as promoting public and private initiative participation and pointing out the economic benefits of data reuse for public service improvement;
e) data curation and publishing - define, within the organization, the authority responsible for the data repository and ensure that datasets are disclosed respecting best publishing practices and data quality;
f) organization and internal processes - create a government strategy for data management, maintenance and use, constituting this data into strategic assets for better decision-making and fostering a data culture within the administration;
g) legal and privacy aspects - define, in a clear manner, the privacy and security aspects that support the data and under which type of license to include dataset use;
h) qualification and training - foster a data culture within society and enhance data skills among students and government officials who work directly with open data.

On the subsequent phase (exploration of the material) we moved to developing the maturity diagnosis model for open government data portals. Governance model elaboration requires formalizing the capacity levels explored by a maturity diagnosis instrument, thus resulting in a more accurate maturity assessment of the organization investigated (Anne *et al*., 2017). The assessment levels, presented below (Table 1), vary between zero (0) and three (3), in which

zero indicates lack of the assessed element, and three the maximum level of maturity.

Table 1 - Maturity levels

Maturity level	Rationale	Score
Level 0	Most primitive assessment stage. It demonstrates inertia related to technical aspects, compliance with good practices, sustainability, governance, and transparency in the portal actions.	0
Level 1	Initial stage of compliance with the assessed element. It demonstrates an emerging concern with technologies, data access and use, good publishing practices, but there is still lack of a better approach to data governance, sustainability, transparency, and engagement with social sectors.	1
Level 2	Assessment proper. It demonstrates formal concern with good infrastructure and publishing practices, a greater concern with portal transparency and sustainability, initial greater engagement with other social actors; however, the guidelines and systematization required for ideal data governance still lack maturity and transparency.	2
Level 3	High capacity assessment stage. At this level, the government can provide an open data portal with the best technologies, the most modern publishing practices, modern and transparent data governance that serves internal management and trains its employees, as well as a consolidated engagement with various important sectors of society.	3

Source: Elaborated by the authors.

The proposed instrument, developed based on the analysis categories (indicators) defined in the pre-analysis, was then applied to open government data portals to assess their maturity regarding these indicators. Chart 1 presents the indicators, their respective requirement variables, and their associated maturity levels.

Chart 1 – Maturity diagnosis model for open government data portals

Indicator 1: IT infrastructure				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**

1.1. Does the open data portal use any data management system (DMS) to enable data description, publishing, sharing, reuse, and content management?	The portal uses static HTML pages without a DMS.	The portal uses a proprietary DMS[1], generating costs to the institution.	The portal uses a government-developed DMS. Despite the lower costs, it can cause maintenance and interoperability issues.	The portal uses an open-source, modern, and free DMS, which enables cost reduction, interoperability handling and access to innovations developed by active practice communities.
1.2. Does the open data portal inform the original publishing date, date of the last update, and the publisher sharing the data?	No information available.	At least one piece of information is available.	At least two pieces of information are available.	Original publishing date, data of the last update, and dataset publisher identification is provided.
1.3. Does the open data portal inform the delay between modifying the dataset in the original source and the portal updating?	No information available.	No information is available but the portal offers a link to the original source for data comparison.	The information is available, and changes take a few days to appear on the portal.	The information is available, and changes take only a few hours or minutes to appear on the portal.
1.4. Does the open data portal inform how long the server was available in the month?	No information available.	The information is provided, with the server available on average below 95% of the time within a year.	The information is provided, with the server available on average between 95% and 98% of the time within a year.	The information is provided, with the server available 100% or on average 99% of the time within a year.
Indicator 2: Accessibility and usability				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
2.1. Portal navigation and data presentation	Portal navigation without highlights or structure.	Portal with highlighted navigation and auxiliary navigation on deep link pages.	Portal with highlighted navigation and auxiliary navigation on deep link pages. Pages lack graphic pattern and visual consistency.	Portal with highlighted navigation, auxiliary navigation on deep link pages, and pages following an information architecture pattern that reinforces visual consistency during navigation.
2.2 Dataset search	The portal does not provide a search field.	The portal offers a search field for locating	The portal offers a search field for locating datasets and refinement	The portal offers a search field for locating datasets, refinement filters, and a thematic organization by

		datasets.	filters.	datasets.
2.3. Data presentation	Data shared with no descriptive elements.	Data shared with a few descriptive elements or using a complex language.	Data shared with descriptive elements in an accessible language and may contain a preview.	Data shared with descriptive elements in an accessible language and may contain a preview accompanied by graphics or relevant analysis.
2.4. Resources for dataset download	No download option is offered.	Only a dataset download link is provided.	Dataset download link is provided with options for data exporting in different formats (HTML, PDF, CSV, JSON, etc.).	Dataset download link is provided with options for data exporting in different formats, as well as an API (Application Programming Interface) access.
2.5. Does the open data portal offers features that allow users to report errors in datasets?	No features are offered.	The data publisher's email or a generic contact form is provided.	An email or form that specifically addresses error situations is provided.	Error reporting resource provides a protocol number to follow the process until its resolution.
2.6. Does the open data portal offers interaction features that allow users to classify, comment, and evaluate datasets by means of a survey?	No interaction features are provided.	At least one interaction feature is provided.	At least two interaction features are provided.	The portal content can be rated, commented, and evaluated by means of a survey. This feedback is considered in future improvements to the portal.
Indicator 3: Data curation and publishing				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
3.1. What is the scope of data publishing defined by the legislation or open data policy?	No information available.	Scope is restricted to the national level.	Scope is restricted to national and regional levels.	Scope defined at the national, regional, and local levels.
3.2. Does the open data policy provide a strategy for ensuring quality data publishing?	No strategy identified.	It considers making the data available in a non-proprietary, machine-readable format.	It considers making the data available in a machine-readable format with common identifiers (URIs).	It considers making the data available at the 5-star LOD level, foreseeing associations to other data from external sources with well-defined semantics for contextualization (e.g., use and reuse of domain-specific ontologies for organization and semantic interconnection of dataset content).

3.3. Does the government publish on its open data portal any technical documentation on the metadata models used?	No documentation available.	Documentation is published, but lacks a data dictionary.	Documentation is published with data dictionary.	Documentation is published with data dictionary, as well as information about external sources related to the metadata schema.
3.4. Which actors are authorized to publish data on the portal?	Government publishers only.	Government publishers and non-governmental organizations.	Government publishers, non-governmental organizations, and the private sector.	Government publishers, non-governmental organizations, the private sector, and any citizen.
Indicator 4: Political and social				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
4.1. Does the portal provide a public consultation field to diagnose which open government data should be disclosed?	No open public consultation identified.	Open public consultation with restricted participation.	Open public consultation with no participation limitations.	Open public consultation with no participation limitations. Results are demonstrably published on the portal following a schedule.
4.2. Does the government publish on its open data portal examples of data reuse services that have brought political, social, and environmental impacts or benefits to society?	No examples provided.	Examples are provided with basic information.	Examples are provided by detailed reports for consultation.	Examples of applications that reuse data to fight against inequality and corruption and preserve nature.
4.3. Does the government detail on its open data portal the number of agencies and bodies that have already disclosed their data?	No details identified.	Between 1% and 40% of government agencies and bodies have already disclosed their data.	Between 40% and 80% of government agencies and bodies have already disclosed their data.	Between 80% and 100% of government agencies and bodies have already disclosed their data.
4.4. Does the government foster data reuse initiatives on its open data portal aimed at creating applications for marginalized communities?	No initiative disclosed.	Public contests and hackathons held but not updated.	Public contests and hackathons held with public sector involvement only.	Public contests and hackathons held with involvement of the public sector and several members of society.
Indicator 5: Economic				

Variable	Maturity level			
	Level 0	**Level 1**	**Level 2**	**Level 3**
5.1. Does the open data policy foresee any business activation strategy by using open data?	No strategy mentioned.	Strategy mentioned without further details.	Strategy mentioned containing systematization.	Strategy mentioned containing systematization and guidelines for its adoption (actions, execution time and responsibilities).
5.2. Does the government foster data reuse on its open data portal initiatives aimed at creating innovative applications for public services?	No initiative disclosed.	Public contests and hackathons held but not updated.	Public contests and hackathons held with public sector involvement only.	Public contests and hackathons held with involvement of the public sector and several members of society.
5.3. Does the government publish on its open data portal examples of data reuse that have brought economic impacts or benefits to society?	No examples provided.	Examples are provided with basic information.	Examples are provided accompanied by detailed reports for consultation.	Examples are provided through applications that reuse data for economic development, employment and income generation, and stimulus to innovation.
Indicator 6: Organization and internal processes				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
6.1. Does the open data policy include initiatives to produce an openness culture within government agencies and bodies?	No initiatives mentioned in.	The initiative is mentioned in but with no further details.	The initiative is mentioned containing its systematization.	The initiative is mentioned containing its systematization and guidelines to encourage leadership, management, supervision, and internal communication policies.
6.2 Does the open data policy contemplate a plan for data management, maintenance and use as strategic assets within public administration?	No plans mentioned.	The plan is mentioned but with no further details.	The plan is mentioned containing its systematization.	The plan is mentioned containing its systematization and guidelines to ensure the existence of authorities, functions, organizational structures, and resources to support data management, maintenance, and use as strategic assets.

6.3. Does the government share on its open data portal examples of data reuse within public administration that enabled better policy and decision-making processes?	No examples provided.	Examples are provided with basic information.	Examples are provided with detailed reports for consultation.	Examples are provided with detailed reports for consultation and accompanied by infographics for analysis and better understanding of the impacts generated within public administration and the benefits for decision-making.
Indicator 7: Legal and privacy aspects				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
7.1. Is legislation on open data regularization published on the portal?	Legislation is not available.	Mentions only a few parts of the legislation.	Legislation is available, but is of difficult access or hidden.	Legislation is available in a transparent, visible, and easily accessible manner.
7.2. Is open data published with its use license?	Data are published without any mention of their use license.	License displayed without details regarding its use extent.	License displayed with details on its use extent, but it is not readily found on the page.	License displayed with details on its use extent and easily found on the page.
7.3. Does the government offer any resource on its open data portal to report a dataset that fails to comply with legal and privacy aspects?	No resources offered.	The data publisher's email or a generic contact form is provided.	An email or form that specifically addresses complaints regarding legal and privacy aspects is provided.	Error reporting resource provides a protocol number to follow the process until its resolution.
Indicator 8: Qualification and training				
Variable	**Maturity level**			
	Level 0	**Level 1**	**Level 2**	**Level 3**
8.1. Does the government make events to discuss the open data available on its open data portal?	No events available.	Events are available but are not updated.	Recent events, with public sector involvement only, are made available.	Recent events with involvement of the public sector and several members of society are made available.

8.2. Does the government disclose actions to develop data ecosystem skills and competencies in educational curricula on its open data portal or data policy?	No actions disclosed.	Manifestation of action only.	Developed actions, which involved the public sector only, are disclosed.	Developed actions, which involved the public sector and organizations that represent society, are disclosed.
8.3. Does the government publish content on its open data portal related to open data skills training?	No open data skills training initiatives disclosed.	Contents (guides and documents) are disclosed on the portal to support employee qualification.	Contents (guides and documents) and training programs to support employee qualification are disclosed on the portal.	Content (guides and documents), training programs, as well as data tools or software to support employee qualification are disclosed on the portal.

Source: Elaborated by the authors.

Our maturity diagnosis model comprises, therefore, eight indicators, totalling 30 requirement variables. Table 2 presents the number of variables and maximum score for each indicator and the total score an open government data initiative can achieve.

Table 2 - Maturity diagnosis model for open government data portals

Indicator	**Number of variables**	**Maximum score**
IT infrastructure	4	12
Accessibility and usability	6	18
Data curation and publishing	4	12
Political and social	4	12
Economic	3	9
Organization and internal processes	3	9
Legal and privacy aspects	3	9
Qualification and training	3	9
Total number of variables and score of the maturity diagnosis for open data portals	**30**	**90**

Source: Elaborated by the authors.

Once finalized, we applied the maturity diagnosis model to open data portals from the United States, the United Kingdom, Brazil, and Ireland for validation (third phase). Application was conducted between July 11 and August 17, 2021.

First, we identified the investigated portals' respective datasets and then submitted them to the model. Other information about the datasets was extracted from portals' content, explored from their main navigations. Not-easily identifiable information was accessed by secondary navigation, which led to repositories with technical and data policy documentation.

Together with the result of the diagnosis, the presentation of the maturity level reached by each evaluated initiative is foreseen by means of an average calculation. The result of all this investigation is presented in the next section.

3 Results

Table 3 shows the results for each initiative, the maturity level achieved by each portal, as well as the compliance percentage of each initiative with the maturity diagnosis. Maturity levels was estimated by calculating the mean of the final score of each initiative in relation to the number of variables.

Table 3 – Ranking of the portals evaluated

Position	Country	Final score	Maturity level	Conformity with the diagnosis
1st	Ireland	66	2.20	73.3%
2nd	USA	54	1.80	60.0%
3rd	Brazil	53	1.76	58.9%
4th	United Kingdom	50	1.66	55.5%

Source: Elaborated by the authors.

Ireland came first, followed by the USA and Brazil with similar technical results, and finally the United Kingdom in last place.

Regarding the level of maturity, Ireland has shown concern about its data policy, paying attention to portal sustainability and modernity. The initiative is

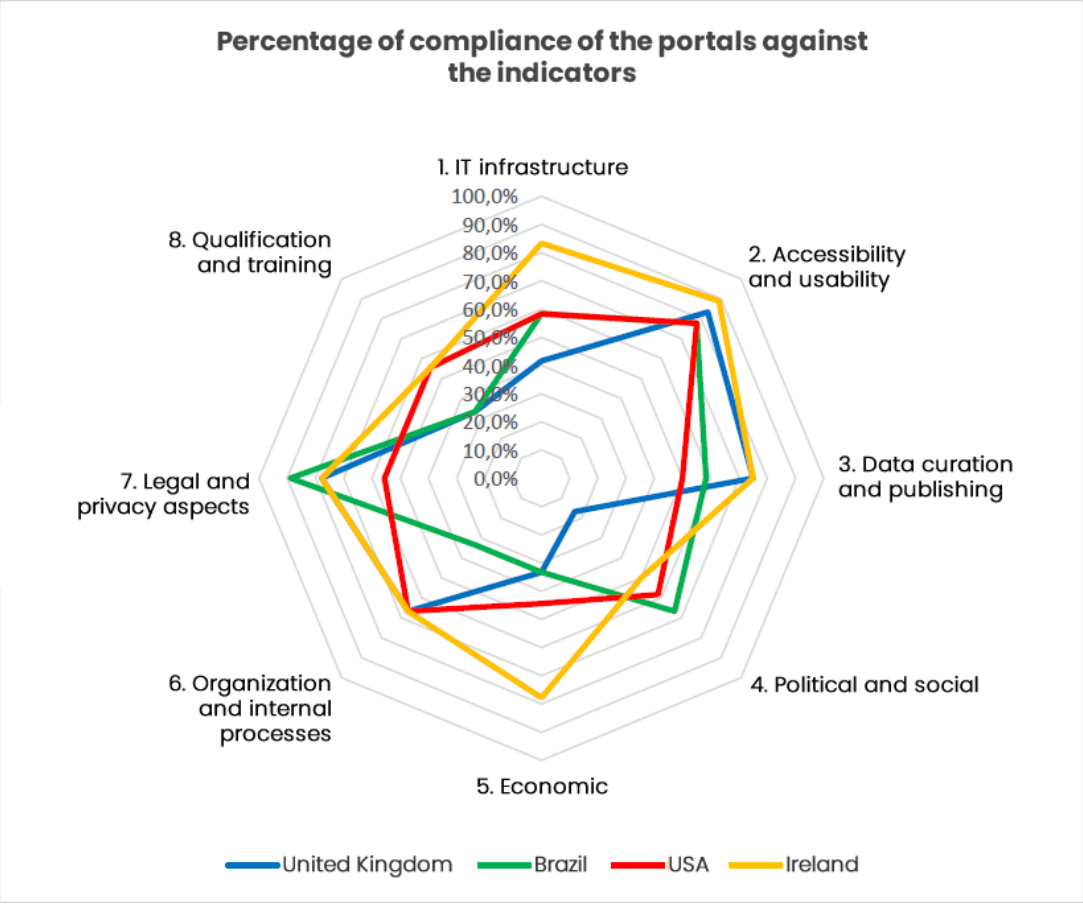
at an established stage of maturity, remaining at level 2, and should improve governance actions in its social aspect and performance.

The US has a resource-rich portal that needs more social engagement and updating. Its policies are not comprehensive and have not been applied satisfactorily, casting doubt on the government's commitment to good data management, resulting in a maturity level 1.

Brazil's portal also had a maturity level 1, with little concern for its sustainability and monitoring of actions. The country's data policy needs to be more comprehensive and committed to implementing the actions established.

The United Kingdom also stagnated at stage 1, with the portal showing little social concern, not data impact and little transparency in the actions conducted by its governance. Figure 2 presents the data compiled and summarized in relation to the eight indicators determined in the research and shows their relation with each country evaluated.

Figure 2 – Portal performance in relation to indicators



Source: Elaborated by the authors.

We explored the following sections on Ireland's portal: datasets, guide for publishers, resources & publications and open data license. The graph for Ireland overlaps the other initiatives in *IT infrastructure* and *Economic* indicators. The fund of investments for open government data reuse actions carried out by the government suggests that activities to generate economic value from data are successful given the number of applications cited on the portal with which new business has been generated, which makes economy more vibrant.

The return of foreign exchange to the public coffers offers opportunities for reinvestment in the government data sector, guaranteeing resources for data governance to improve the technology sector, which may justify the good IT capacity.

The indicators *Data curation and publishing*, *Legal and privacy aspects*, and *Accessibility and usability* have, on average, a good maturity capacity for publishing quality open government data on the web. On *Data curation and publishing*, however, one aspect to be improved is the greater engagement of society in data publishing on the portal, a recurrent issue in the evaluated portals.

Evaluation of the datasets on the Irish portal revealed the absence of user-oriented resources for reporting both technical errors and flaws in its legal and privacy aspects.

Ireland's last three indicators performed the worst in terms of graphic design: *Political and social*, *Organization and internal processes,* and *Qualification and training*. Ireland fails to provide a transparency dashboard designed to monitor agencies' data publishing and does not report on initiatives to reuse data for the benefit of disadvantaged communities.

Successful examples, such as the data server availability monitor and the economic investment fund, could be replicated to fill these gaps. In the last two indicators, the key to increasing compliance would be increasing the government's transparency on the portal, showing the benefits that data reuse brings to the public administration, in addition to continued investments in developing data skills in its educational system.

On the U.S. portal, we explored the data and resources (external link) sections. The best performances were in *Accessibility and usability* and *Organization and internal processes*, with compliance between 77.8% and 66.7%, respectively. The graph shows a certain stability, with no major peaks and the remaining indicators at 60% and 50% compliance, except for *Economic*, with 44.4%. Performance in *Organization and internal processes* led the US to top the indicator, which showed some points to be learned, despite 66% compliance.

The US data governance strategies are well formulated and present extensive technical details, especially its Action Plan. The scope, activities, actions, responsibility assignments and goals are usually published and well documented, but, as the literature review pointed out, the US portal's score has declined and worsened in relation to open government data.

We notice that the policies are not yet being applied satisfactorily, which may explain why the indicators *IT infrastructure*, *Data curation and publishing*, *Political and social*, *Economic*, *Legal and privacy aspects*, and *Qualification and training* are at medium maturity levels, varying between an initial or established capacity.

Unlike the other portals evaluated, the US strategy did not include LOD, which casts doubt on the current government's commitment to adhering to good data governance practices.

There is also a lack of greater civic engagement and participation in the portal, with few points of interactive contact between members of society and the digital tool. The events and impacts areas reflect the lack of engagement, being mostly out of date and with little participation from sectors of society in their organization.

Dataset and content were the sections explored on the Brazilian portal. The country's graph showed a great imbalance, with a peak in the *Legal and privacy aspects* indicator, in contrast to low performances in the *Organization and internal processes*, *Economic*, and *Qualification and training* indicators. The pre-test had already indicated great concern with the legislation aspect, which was reflected in the score for this indicator. The remaining indicators had their variables with initial or established average capacity.

In the *Data curation and publishing* indicator, Brazil, like the US, has a limited scope for executing its data policy at the federal level. Although some publications from other entities are visible on its portal, the standardization of data policy between all spheres for a higher publishing quality is a point that needs further discussion and consensus, especially for sharing interoperability standards, as is the case with ePING[2].

Similar to most of the initiatives analysed, Brazil references LOD for data quality in its plan, as well as good technical documentation on the portal about publishing issues. In practice, however, what we observe is low adherence to the LOD, with governments accepting the three-star level (European Union, 2021; Macedo; Lemos, 2021). This reveals the government's low commitment to making the precepts formalized in its policy instruments a reality. There is no clear action plan with responsibilities, goals and deadlines for putting the full potential of LOD into practice. Mentioning the adoption of good practices can be misinterpreted by people as a real action by the government to apply these measures systematically.

Brazil scored the worst in the *Organization and internal processes* indicator among the portals evaluated. Despite considering the creation of a data culture in its policy, the document did not include data reuse strategies within the public administration. As with the other initiatives analysed, the portal needs to clarify the government actions that are being taken in this regard and disclose the results achieved in an appropriate area of the portal for public appreciation.

The *Qualification and training* indicator also performed poorly. Although recent events have been held to stimulate discussion about open government data, we failed to identify either in the policy and action plan documents or on the portal the actions that the government has taken in the national curricula in favour of data education. Even the content related to employee training in open government data skills needs to be improved. These aspects should be discussed together with bodies representing civil society and academia, so that not only the country's data culture is fostered, but also so that a higher level of quality is achieved in the data published.

The *Economic* indicator was affected by unclear or non-existent disclosure of governance actions to activate new businesses based on data use.

Brazil's government and its governance committees need to be more attentive to generating strategic value from open government data reuse. Generating value activities can leverage the portal, helping to make it sustainable, which would be very welcome for constant improvement and reinvestment of resources in the platform.

Our analysis of the Brazilian portal identified a great deal of social concern, which led the Brazilian initiative to top the *Political and social* indicator. Also noteworthy is the ombudsperson's office, which concentrates in a single page important possibilities of interaction with the portal and the government, such as making suggestions for new datasets, reporting technical errors and reporting legal and privacy issues.

In addition to the dataset search on the homepage, documentation and footer content were explored on the UK portal. Three indicators stand out in the graph analysis: *Accessibility and usability*, *Data curation and publishing* and *Legal and privacy aspects*. The portal topped the *Data curation and publishing* indicator, standing out with a data policy that contains very detailed data governance strategies and good technical documentation. But these results conflict with the literature review, which indicated the UK as one of the worst in terms of data publication, with the pre-test suggesting datasets in poor conformity with the best formats.

This can be explained from the findings in the literature review on the government's organizational reality, and confirmed in this research by the maturity diagnosis, that is, a low maturity level in the indicators *Organization and processes* and *Qualification and training*, and a low maturity level in the political and social issue. These indicate that the strategies and action plans designed in the data policy have not been properly applied, casting doubt on the effectiveness of national data governance.

UK's portal needs to offer greater transparency regarding the economic, social and political results and impacts of data reuse. As highlighted by the study, it is not enough to define actions and methods to achieve publication maturity at LOD level if a large number of datasets continue to be made available in inadequate formats, as demonstrated in the literature and in the pre-tests conducted.

4 Discusion

The indicator scores and maturity levels achieved indicate that the open government data initiatives analysed lack broader and more effective data policies, failing to address important data quality dimensions.

We also observe a lack of political commitment to putting into practice the action plans formalized, with little to no transparency regarding action performance. In addition to a still very timid social interaction in the portal activities, social engagement in data policy development is scarce.

A scenario reflected in the average results of the open government data maturity diagnosis — a disappointing reality seen as they are mostly pioneering initiatives in the open government data movement. A government's open data portal is the main gateway to data access at the national level, disclosing the main benefits and impacts of data reuse and promoting greater social engagement with these data.

Current government handling of data reuse is hindering possible improvements to their internal processes, which hampers the transformative capacity of data and compromises government efficiency. This institutional scenario reflects the operational difficulty government publishers face when publishing data of greater semantic value (using metadata standards, controlled vocabularies, ontological models, and LOD adherence), illustrating well the challenges yet to be resolved.

IS and CS theoretical and methodological practices point to possible intelligent solutions for describing web information resources (Machado; Souza; Da Graça Simões, 2019; Lemos; Souza, 2020; Martins *et al*., 2022). Issues related to standardization, quality and the exchange of descriptions that to some extent could take advantage of the potentialities of the Semantic Web and LOD (Machado; Souza; Da Graça Simões, 2019). Appreciation of the use of standards for the description of digital resources at the structure, value, content and data communication levels (Gilliland, 2016) reduces. In a way, quality data from a digital curation process are those described by those principles.

Martins *et al*. (2022) argue that use of knowledge organization systems (e.g., taxonomies, thesaurus) and more sophisticated representational artifacts

(e.g., ontologies) should be considered as a prerequisite in information organization actions, as to provide semantic enrichment to metadata, a more significant database organization and, consequently, to improve information retrieval systems aligned with the 5-star LOD principles. Such practice would help combat a major issue regarding the fifth star for open government data: the use of vocabularies that enable relating data from different bases.

Public administration should consider data as an essential asset for good government management; therefore, the actions taken to achieve it should be clear and monitored for real effectiveness (Horvath, 2017; Anne *et al*., 2017). By making active data strategic for decision-making, organizations foster data culture with an efficiency gain underpinned by evidence-based informational resources. A more capable management leads to more proactive actions for generating new businesses based on data reuse, thus benefiting the entire national innovation dynamic.

Ultimately, it would generate the capacity to provide economic sustainability to data initiatives by means of systematic reinvestments, thus benefiting sectors that require high quality open government data.

5 Conclusion

The maturity diagnosis model proposed in this study was developed based on findings and inferences consolidated through content analysis techniques.

By analysing the specialized literature and the history of the open government data movement in several countries, we identified local efforts, challenges, and opportunities for making government data available to contemporary society. Based on this state of the art, we systematized a set of indicators mapped in the actions taken by some countries to make government data available, which was considered as input in the model.

We conducted a pre-test on government open data portals from the US, the UK, Brazil, and Ireland, to gain a better understanding of how data governance actually occurs in open data initiatives. All the material collected by the bibliographic and documentary surveys were analysed considering the challenges identified in the state-of-the-art analysis.

We then defined the categories of analysis, a stage especially important to unveil the reality of how government open data provision occurs in practice, the findings of which were used to develop and present a maturity diagnosis model for evaluating government open data portals.

Validation was conducted by applying the model to the four government open data portals previously investigated. Data collected by the maturity assessment underwent content analysis and their semantic conclusions were discussed considering the opportunities and challenges faced by important nations such as Ireland, the US, Brazil and the UK in implementing open data initiatives. We expect this maturity diagnosis model to serve as a reference for quality data publishing to other countries engaged in opening their government data. Such quality, in turn, should be based on a governance policy guided by the fundamental indicators recommended here.

From our findings, future research should propose a governance model aimed at better quality data publishing, management guidance for proposing more comprehensive and effective policies, and mitigation of barriers imposed to data availability.

This maturity diagnosis model could therefore contribute to proposals of a governance model capable of qualifying the open data published on government portals, thus meeting the benefits expected from government data availability.

References

ANNE, Kirk M. *et al*. Building capacity for digital humanities: A framework for institutional planning. **Educause Center for Analysis and Research**. Louisville, 2017.

ATTARD, Judie *el al*. A systematic review of open government data initiatives. **Government Information Quarterly**, [*S. l.*], v. 32, n. 4, p. 399-418, 2015. Available at: https://doi.org/10.1016/j.giq.2015.07.006. Access: 13 nov. 2023.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

BAUER, Florian; KALTENBÖCK, Martin. **Linked open data**: the essentials. Vienna: Edition mono/monochrome, 2011.

BRAZIL. Controladoria-Geral da União. **Portal Brasileiro de Dados Abertos**, [*S. l.*], 2023. Available at: https://dados.gov.br/home. Access: 11 oct. 2023.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: The story so far. *In*: **Semantic services, interoperability and web applications: emerging concepts**. [*S. l.*]: IGI global, 2009, p. 205-227.

BLANK, Marit. **Open Data Maturity Report 2019**. [*S. l.*]: European Data Portal 2019, 2019. Available at: https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf. Access: 17 nov. 2022.

BORGMAN, Christine L. **Scholarship in the digital age: Information, infrastructure, and the Internet**. Cambridge: MIT press, 2010.

DAVIES, Tim. **Open data, democracy and public sector reform**, [*S. l.*], 2010. Available at: https://www.timdavies.org.uk/2010/08/26/online-version-open-data-democracy-and-public-sector-reform/. Access: 10 oct. 2023.

DAVIES, Tim; BAWA, Zainab Ashraf. The promises and perils of open government data (OGD). **The Journal of Community Informatics**, [*S.l.*], v. 8, n. 2, p. 1–6, 2012.

EUROPEAN UNION. **European data portal**, [*S.l.*], 2021. Available at: https://data.europa.eu/pt/trening/what-open-data. Access: 3 mar. 2021.

FEDERAL DATA STRATEGY. **Data Governance Playbook**, [*S.l.*], 2020. Available at: https://resources.data.gov/assets/documents/fds-data-governance-playbook.pdf. Access: 22 nov. 2022.

GASCÓ-HERNÁNDEZ, Mila *et al*. Promoting the use of open government data: Cases of training and engagement. **Government Information Quarterly**, [*S.l.*], v. 35, n. 2, p. 233-242, 2018. Available at: https://doi.org/10.1016/j.giq.2018.01.003. Access: 13 nov. 2023.

GILLILAND, Anne J. Setting the stage. *In*: M. Baca, & G. R. Institute (Ed.). **Introduction to metadata**, 3 ed. Los Angeles: Getty Research Institute, p. 1-19, 2016.

GRAY, Jonathan. Towards a genealogy of open data. *In*: The paper was given at the General Conference of the European Consortium for Political Research in Glasgow., **SSRN**, Glasgow, 2014. Available at: http://doi.org/10.2139/ssrn.2605828. Access: 13 nov. 2023.

GUIZZARDI, Giancarlo. Ontology, ontologies and the "I" of FAIR. **Data Intelligence,** [*S.l.*], v. 2, n. 1-2, p. 181-191, 2020. Available at: https://doi.org/10.1162/dint_a_00040. Access: 13 nov. 2023.

HORVATH, Aniko. 'Governance'–in crisis? A cross-disciplinary critical review of three decades of 'governance'scholarship. **Centre for Global Higher Education working paper series**, [*S.l.*], n. 20, p. 1-181, 2017.

IRELAND. DEPARTMENT OF PUBLIC EXPENDITURE AND REFORM. **Ireland's Open Data Portal**, [*S.l.*], 2023. Available at: https://data.gov.ie/. Access: 11 oct. 2023.

JANSSEN, Marijn *et al*. Transparency-by-design as a foundation for open government. **Transforming Government: People, Process and Policy**, [*S.l.*], v. 11, n. 1, p. 2-8, 2017. Available at: https://doi.org/10.1108/TG-02-2017-0015. Access: 13 nov. 2023.

JUANA-ESPINOSA, Susana de; LUJÁN-MORA, Sergio. Open government data portals in the European Union: Considerations, development, and expectations. **Technological Forecasting and Social Change**, [*S.l.*], v. 149, p. 119769, 2019. Available at: https://doi.org/10.1016/j.techfore.2019.119769. Access: 13 nov. 2023.

KUČERA, Jan *et al*. Methodologies and Best Practices for Open Data Publication. DATESO. [*S.l.*], p. 52-64, 2015.

LEMOS, Daniela L. da S.; SOUZA, Renato Rocha. Knowledge organization systems for the representation of multimedia resources on the web: A comparative analysis. **Knowledge Organization**, v. 47, n. 4, p. 300-319, 2020. Available at: https://doi.org/10.5771/0943-7444-2020-4-300. Access: 13 nov. 2023.

LUNA-REYES, Luis Felipe; NAJAFABADI, Mahdi M. The US open data initiative: The road ahead. **Information Polity**, Amsterdam, v. 24, n. 2, p. 163–182, 2019. Available at: https://doi.org/10.3233/IP-180106. Access: 13 nov. 2023.

MACEDO, Dirceu Flavio.; LEMOS, Daniela L. da S. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 10, n. 2, p. 14 - 26, abr. 2021. Available at: http://doi.org/10.5380/atoz.v10i2.77737. Access: 13 nov. 2023.

MACHADO, Luís Miguel O.; SOUZA, Renato Rocha; DA GRAÇA SIMÕES, Maria. Semantic web or web of data? a diachronic study (1999 to 2017) of the publications of tim berners-lee and the world wide web consortium. **Journal of the Association for Information Science and Technology**, [*S.l.*], v. 70, n. 7, p. 701-714, 2019. Available at: https://doi.org/10.1002/asi.24111. Access: 13 nov. 2023.

MARTINS, Dalton Lopes *et al*. Information organization and representation in digital cultural heritage in Brazil: Systematic mapping of information infrastructure in digital collections for data science applications. **Journal of the**

Association for Information Science and Technology, [*S.l.*], v. 74, n. 6, p. 707-726**,** 2022. Available at: https://doi.org/10.1002/asi.24650. Access: 13 nov. 2023.

OPEN GOVERNMENT PARTNERSHIP. **Open Government Declaration**, 2011. Available at: https://www.opengovpartnership.org/process/joining-ogp/open-government-declaration/. Access: 9 jan. 2022.

PINTO, Jaime Andrade; ALMEIDA, Maurício Barcellos. Ontologias públicas sobre governo eletrônico: Uma revisão sistemática da literatura. **Brazilian Journal of Information Science**: Research Trends, Marília, v. 14, n. 3, p. e020003, 2020. Available at: https://doi.org/10.36311/1940-1640.2020.v14n3.10105. Access: 13 nov. 2023.

SEGUNDO, Jose Eduardo Santarem. 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**, [*S.l.*], v. 8, n. 2, 2015.

SHEPHERD, Elizabeth *et al*. Open government data: critical information management perspectives. **Records Management Journal**, [*S.l.*], v. 29, n. 1/2, p. 152–167, 2019. Available at: https://doi.org/10.1108/RMJ-08-2018-0023. Access: 13 nov. 2023.

UNITED KINGDOM. GOVERNMENT DIGITAL SERVICE. **U.K Government's Open Data Portal**, 2023. Available at: https://data.gov.uk/. Access: 11 oct. 2023.

UNITED STATES. U.S. GENERAL SERVICES ADMINISTRATION. **U.S. Government's Open Data Portal**, 2023. Available at: https://www.data.gov/. Access: 11 oct. 2023.

WANG, Victoria; SHEPHERD, David. Exploring the extent of openness of open government data–A critique of open government datasets in the UK. **Government Information Quarterly**, [*S.l.*], v. 37, n. 1, p. 101405, 2020. Available at: https://doi.org/10.1016/j.giq.2019.101405. Access: 13 nov. 2023

WILKINSON, Mark D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific data**, London, v. 3, n. 1, p. 1-9, 2016. Available at: https://doi.org/10.1038/sdata.2016.18. Access: 13 nov. 2023.

WORLD WIDE WEB FOUNDATION. **Open Data Barometer - Leaders Edition**. Washington DC: World Wide Web Foundation. 2018. Available at: https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf. Access: 13 jan. 2022.

ZENG, Marcia Lei. Interoperability. **KO Knowledge Organization**, Baden-Baden, v. 46, n. 2, p. 122-146, 2019. Available at: doi.org/10.5771/0943-7444-2019-2. Access: 13 nov. 2023.

ZENG, Marcia Lei; QIN, Jian. **Metadata**. Chicago: ALA Neal-Schuman, 2016.

Dados abertos governamentais: Modelo de diagnóstico de maturidade voltado a qualidade de dados para publicação na web

Resumo: A escalada na abertura de dados governamentais é um fenômeno originado a partir de longos debates sobre a modernização do Estado. A pesquisa é motivada em compreender os desafios impostos à administração pública na publicação de dados de forma alinhada ao movimento aberto sob a ótica dos portais de dados abertos. O objetivo deste artigo consiste em propor um modelo de diagnóstico para a maturidade de portais de dados abertos, com foco na disponibilização dos recursos informacionais do governo com abrangência, consistência, eficácia e transparência. A metodologia foi dividida em três fases. A primeira fase foi um pré-teste realizado em portais de dados abertos nos Estados Unidos, Reino Unido, Brasil e Irlanda, motivado por pesquisas que evidenciaram o histórico do movimento de dados abertos, conjuntamente a um panorama atual de iniciativas, estágios de evolução e desafios, que serviu para definir as dimensões analíticas fundamentais para a proposição do modelo de maturidade. Na segunda fase, foi desenvolvido um modelo de diagnóstico de maturidade para portais de dados governamentais abertos. Na terceira fase, o modelo foi aplicado e validado nos mesmos portais do pré-teste. Os resultados da aplicação do instrumento de diagnóstico podem orientar a administração pública na condução de uma governança de dados mais eficiente e responsável, beneficiando o governo, o movimento de dados abertos e a sociedade civil.

Palavras-chave: dados governamentais abertos; organização da informação; modelo de diagnóstico de maturidade; qualidade de dados; governança de dados

Recebido: 19/05/2023
Aceito: 03/11/2023

Declaração de autoria

Concepção e elaboração do estudo: Dirceu Flavio Macedo, Daniela Lucas da Silva Lemos.

Coleta de dados: Dirceu Flavio Macedo, Daniela Lucas da Silva Lemos.

Análise e interpretação de dados: Dirceu Flavio Macedo, Daniela Lucas da Silva Lemos.

Redação: Dirceu Flavio Macedo, Daniela Lucas da Silva Lemos.

Revisão crítica do manuscrito: Dirceu Flavio Macedo, Daniela Lucas da Silva Lemos

Como citar

MACEDO, Dirceu; LEMOS, Daniela Lucas da Silva. Open government data: maturity diagnosis model for quality data published on the web. **Em Questão**, Porto Alegre, v. 30, e-132617, 2024. DOI: https://doi.org/10.1590/1808-5245.30.132617

1 When using some type of proprietary solution, the institution ends up being dependent on the software vendor.

2 E-Government Interoperability Standards