MGI

Master Degree Program in
Information Management

PUBLIC POLICY: TURNING OPEN DATA INTO DEMOCRATIC DATA

Portal Quality Assessment – Comparative Analysis

André Drumond Mateus

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

PUBLIC POLICY: TURNING OPEN DATA INTO DEMOCRATIC DATA

By

André Drumond Mateus

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence

Supervisor: Prof. Flávio L. Pinheiro, Ph.D

February 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, February 27, 2023

ABSTRACT

As data, information, and their respective provisioning gets more and more ubiquitous, people start to look for - and even demand - transparency and data to support the policies in effect right now that, either directly or indirectly, affects them. There are many expectations related with Open Government Data initiatives, such as improving policymaking, increase in transparency of government spending, advance citizen engagement with the institutions, etc. This master's proposal aims to offer research that pertains to this theme; including an in depth look into one of the most reputed OGD maturity report (EU Open Data Maturity Report), a systematic literature review of Open Data's main objectives and goals, the influence on publication of scientific literature as well as the potential socio-economic and transparency impact they may have, in order to proceed to an assessment of the portal quality in Portugal by evaluating the usage of its data on scientific papers and articles; through the usage of bibliometrics and PRISMA methodology. My thesis research drills down on these topics: What are the most used Portuguese OGD portals in academic literature? What are the authors that make the most use of Portuguese OGD portals? What characterizes the authors and the publications?

KEYWORDS

Open Data; Transparency; Socio-economic impact; Open Data Quality; Open Data Portal Benchmark; Bibliometrics.

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

OGD	Open Government Data
OKFN	The Open Knowledge Foundation
OGP	Open Government Partnership
AMA	Agência para a Modernização Administrativa, IP
RCAAP	Repositório Científico de Acesso Aberto de Portugal
GODI	Global Open Data Index
ODB	Open Data Barometer
ODIN	Open Data Inventory
ODMR	Open Data Maturity Report
OGDI	Open Government Development Index
OKF	Open Knowledge Foundation
W3F	World Wide Web Foundation
OECD	Organization for Economic Co-operation and Development
EU	European Union
UN	United Nations
ODRA	Open Data Readiness Assessment
PSI	Public Sector Information
WJP OGI	World Justice Project Open Government Index
ITU	International Telecommunication Union
ICT	Information and Communication Technologies
OCD	Open Corporate Data
IS	Information Systems
ADO	Antecedents, Decisions, Outcomes
TAM	Technology Acceptance Model
UTAUT	Unified Theory of Acceptance and Use of Technology
DKAN	Drupal-based Knowledge Archive Network
IDC	International Data Corporation

INTRODUCTION

Open Government Data (OGD) is a philosophy that aims at promoting transparency, accountability, and value creation by making government data available to all (OECD, 2013). It translates into using digital media to make data available with the technical and legal characteristics necessary to be freely used, reused, and redistributed by anyone, anytime, and anywhere (Open Data Charter, 2015).

Public organizations have progressively embraced OGD principles and made more datasets available in the form of Open Government Data (Attard et al., 2015) through web-based portals. As such, OGD is open data that anyone can freely access, meaning that it does not have any privacy-related restrictions, nor is it abridged by confidentiality concerns (Kawashita et al., 2022). It is made available without restrictions on their usage, modification, or distribution (Janssen et al., 2012).

There are many expectations related to OGD: it is expected to strengthen transparency and democratic processes, stimulate economic growth and innovation, and lead to more effective public services and programs (Attard et al., 2015; Garcia Saez, 2022). It holds great business potential, with a global economic value estimated between 3.2 to 5.4 trillion EUR annually and forecasted cost savings of 1.7 billion EUR for the EU28+ countries (Krasikov et al., 2021).

Some of the perceived benefits and expectations for opening government data are (Janssen et al., 2012):

- **Improve policy-making** – with country-specific data analysis, open data provides policymakers and organizations with useful insights on various societal themes and issues.
- **Increase transparency –** overview of detailed spending and fundraising by the governments and public agents provides transparency in public office activities.
- **Citizen engagement –** citizen participation thrives when public administration activities are open for scrutiny.
- **Social and commercial value –** Data can be used for many different purposes than were originally intended. The government encourages stakeholders to innovate by publishing such data, possibly creating new services.
- **Participatory Governance –** By publishing government data, the citizens are empowered and given the opportunity to participate in the decision-making process, effectively transforming the citizen into a stakeholder of the governance process instead of just voting sporadically (Attard et al., 2015).

Moreover, the growing demand for more OGD comes not only from society itself, with its increasing digitalization (IDC - *The Digitization of the World,* 2018), but there is also regulatory pressure: The Open Data Directive entered into force on 16 July 2019, with Member States having until 16 July 2021 to implement it. The Directive 2019/1024/EU drives:

- Creation of products and services based on public information.
- Free circulation of information and communication.
- Publication of dynamic data and accepting Application Programming Interfaces (APIs).

In recent years risks regarding data rights, privacy, transparency, and trust have become widely discussed. It is important to realize that despite the demand for OGD and the benefits associated with it are a frequent topic in the literature, as of now, evidence of such transformation is scarce, and there

are still several barriers that hinder, or even impede, its use (Janssen et al., 2012; Kawashita et al., 2022).

The Local Open Data Census, developed by the Open Knowledge Foundation in 2014, is an essential tool for civil society to monitor open data publishing at the local, regional, and municipal levels across countries. The census offers a comprehensive set of indicators to evaluate the level of openness of local governments (de Castro Neto et al., 2017). However, the assessment of data collected from the census indicates that most of the defined indicators, approximately 13 out of 15, are provided by the central government, whereas local government or private companies provide the rest. Furthermore, it is noteworthy that most municipalities only release data that they are legally obligated to provide, and data sharing strategies are still lacking in most of them. Instead of perceiving data sharing as a means to increase transparency and create business opportunities, municipalities view it as a means of fulfilling their legal obligations. Hence, there is an urgent need to improve and support data sharing strategies and platforms by municipalities or inter-municipal communities (de Castro Neto et al., 2017). Given this context, it is crucial to develop a standard and automated benchmarking method for open data portals to encourage more effective and meaningful data sharing practices among municipalities.

Portugal is only a Beginner in Open Data maturity programs (Publications Office of the European Union., 2022). There are several OGDs available through the Portuguese national open data platform dados.gov.pt; however, the number of unique visitors per month is one of the lowest in the EU. Interestingly, the share of visits from abroad is one of the highest recorded (44% of all unique visitors are foreign) (Publications Office of the European Union., 2022). In contrast, according to the OECD Digital Government Index, which measures the progress of the journey towards the digital government, Portugal is above average, occupying 9th place in all the OECD countries. It becomes important to understand who uses the Portuguese Open Governmental Data portals.

Researchers are increasingly using open data to conduct studies and analyze trends, as it provides a wealth of information that would otherwise be difficult to obtain. This study aims to assess the usage of data from open data portals in academic literature, specifically focusing on the case of Portugal. The study will examine the extent to which open data portals in Portugal are being utilized in academic research, as well as the impact of open data on the quality and relevance of research produced. By doing so, this study seeks to contribute to the understanding of the role of open data in scientific research and the potential benefits that can arise from its use.

In that sense, this thesis aims to assess the usage of data from such portals in academic literature. Using bibliometrics data and the PRISMA methodology, we look to answer the following research questions:

- What are the most used Portuguese OGD portals in academic literature?
- What are the authors that make the most use of Portuguese OGD portals?
- What characterizes the authors and the publications?

LITERATURE REVIEW

In this chapter, we provide an account of the historical goals of Open Data, an overview of studies and previously used methods for evaluating the quality of Open Data portals across Europe, and their potential socio-economic impact.

2.1. OPEN GOVERNMENT DATA IN EUROPE

Open Data is one of two major developments, along with big data, reshaping relationships between customers, citizens, businesses, governments, and society with information. Open data is not a binary choice between closed and open data. The degree of data openness is a continuum. Open Data in government refers to data sharing and release to the public, citizens, accredited third parties, or other government organizations (IDC - *The Digitization of the World,* 2018).

Using Open Data by governments, companies, and individuals can generate economic, social, and/or environmental benefits. In this paradigm of Open Data, anyone can redistribute data, only subject to the requirement of reference to who originated/aggregated them. Open Data becomes useful when available online and in a commonly readable format that ensures interoperability (Nogueras-Iso et al., 2021).

"Data and AI are the ingredients for innovation that can help us to find solutions to societal challenges, from health to farming, from security to manufacturing. In order to release that potential, we have to find our European way, balancing the flow and wide use of data while preserving high privacy, security, safety, and ethical standards."

European Commission President
Ursula von der Leyen in "A Union that strives for more - My agenda for Europe" (2019)

The economic impact of Open Data is transversal to several sectors, influencing employment, efficiency gains, and corresponding cost/resource savings (IDC - *The Digitization of the World,* 2018). Indeed, according to a survey conducted by IDC's ongoing Global DataSphere research, where 2400 enterprise decision makers were surveyed – in addition to in-depth interviews with senior IT executives from several industries - 49% of the data used by organizations surveyed is Open Data, and 77% of organizations plan to use even more open data (IDC - *The Digitization of the World,* 2018).

The rise of new technologies has made it easier for governments to publish data, and there is growing pressure for them to do so in order to increase accountability and improve decision-making. The need for open data platforms is therefore increasing globally (Nogueras-Iso et al., 2021)**.**

To promote open government data, there isn't a single, effective policy or approach that works in every case. Instead, various strategies must be implemented to realize the full potential of open government (Schmidthuber & Hilgers, 2021). It is also important to understand that the requirements and policies evolve along with technology. In that sense, the EU put forth the Public Sector Information directive in 2003 (Directive 2003/98/EC), later revised in Directive 2013/37/EU to account for the increasing transition from analog to digital services. Directive 2003/98/EC encourages the re-use of public sector information from EU Member States by leveraging several key dimensions:

- **Open Format** – It introduces the need to adopt the principle of open format and machine readability with the respective metadata. Both the format and the metadata should, as far as possible, respect open formal standards.
- **Cultural Institutions** – The inclusion of certain cultural institutions as public sector bodies such as libraries, museums, and archives.
- **Transparency** – Increased transparency in the calculation of fees, where the basis for calculating these fees should be pre-established and published, if possible, electronically
- **Fees to be charged** – Fees for accessing public sector information based on the marginal costs incurred for its reproduction, availability, and dissemination.
- **Practical modalities for re-use** – Establish practical modalities that facilitate document search for reuse and encourage multilingual document search.
- **Extension of the Directive** – The Directive covers texts, databases, audio files and movie fragments.

The directive was once more revised in 2019 in order to promote the creation of products and services based on public information, facilitating the free circulation of information and communication (European Union, 2019). The new Directive 2019/1024 main highlights are:

- **Implementation** — The Open Data Directive entered into force on July 16, 2019, and Member States have until July 16, 2021, to implement it.
- **APIs** — Encourage the publication of dynamic data and the acceptance of Application Programming Interfaces (APIs).
- **Public-Private Agreements** — Strengthen transparency requirements for public-private agreements involving public sector information, avoiding exclusive agreements.
- **Access Policy** — Develop open access policies for publicly funded research data that will facilitate the reuse of research data that is already contained in open repositories.
- **High Value Data** — Requires the adoption by the Commission of a list of high value datasets to be made freely available.
- **Limitation of Exceptions** — Reduce the exceptions that allow public bodies to charge more than the marginal costs of disclosing and reusing their data.

The European Data Portal recently published a report on the economic impact of open data (Publications Office of the European Union., 2022), showing the industries with the most growth potential from OGD: agriculture, wholesale and retail trade, transportation, information and communication, banking and insurance, and real estate. Each show 15% or more growth potential. The evaluation took into account the industry's level of digitalization and data need, the possibility for open data supply to match the demand, and the possible economic impact (Hassine & Clément, 2020).

Regarding entrepreneurship, the volume of OGD publishing and a nation's level of entrepreneurship is positively and significantly correlated. A few nations excel exceptionally at realizing the entrepreneurial benefits of OGD, which is why there is a positive correlation between OGD and the entrepreneurship (Huber et al., 2022). The findings demonstrate that simply publishing OGD does not guarantee entrepreneurial success. It is demonstrated in the research paper by (Huber et al., 2022) that published OGD and country-level entrepreneurship have a positive relationship that is positively moderated by the quality of national institutions. In nations with excellent institutional quality, there

is a particularly strong correlation between OGD publication and the entrepreneurship (Huber et al., 2022).

A systematic literature review of 169 empirical OGD studies develops a framework to "(…) *unify and grasp the accumulating isolated evidence on OGD in the context of the digital economy and provide a theory-informed research agenda to tap the potential of information systems (IS) research for OGD*" (Wirtz et al., 2022). The authors argue that there are six key topics of OGD, based on a theoretical review framework of Antecedents, Decisions, and Outcomes (ADO), that need to be addressed to tap the potential of IS research (Wirtz et al., 2022). Future work in benchmarking the Portuguese open government data portals against these criteria, can add to the understanding of the strengths and weaknesses of each portal and how they can be improved to better support all the interested stakeholders.

2.2. OPEN DATA BARRIERS

There are various barriers to open government data (OGD), such as government culture, organizational reluctance, and concerns about data misuse. In addition, the heterogeneity of data formats used by public administrations poses technical challenges for data producers and users. The abundance of different data structures and lack of comparability across government data portals are also issues that need to be addressed (Attard et al., 2015).

Several myths associated with OGD have been identified, such as the belief that publicizing data will automatically yield benefits, that all information should be unrestrictedly publicized, and that open data will lead to open government. These misconceptions fail to recognize the complexity and diversity of OGD, as well as the challenges in using and analyzing open data (Janssen et al., 2012).

Moreover, E. Ruijer et al. (2020) have exposed the politics of OGD and identified "strategic opaque transparency" as a tactic used by governments to selectively release OGD in "easy" and "harmless" domains while withholding information in other domains. Organizations tend to follow OGD practices when they improve efficiency or legitimacy and are supported by formal and informal norms and organizational objectives, but they may compromise in uncertain electoral contexts or with conflicting stakeholders.

Open Government Data (OGD) has the potential to offer a wide range of benefits, but also presents numerous challenges such as difficulties in work, use, law, information quality, and participation (Janssen et al., 2012).

Despite the potential advantages of OGD, it may not always be released in all policy domains due to political and strategic considerations. Strategic opaque transparency, or the selective release of OGD in "easy" and "harmless" domains, can be a tactic used by organizations to limit the data that is available or distribute it among various datasets. This study suggests that organizational responses to OGD may differ depending on the policy domain and that low contextual uncertainty may increase data release receptivity, while high contextual uncertainty and high legitimacy may increase reluctance (E. Ruijer et al., 2020).

These findings are relevant to the benchmarking of open data portals, as they highlight the importance of understanding the politics and strategic considerations that influence the release of OGD in different policy domains.

The benefits of open data portals identified by Kawashita et al. (2022) include increased transparency, social control, civic participation, public engagement, accountability, administrative efficiency, reduced operating costs, access to external capacity and resources, improved processes, products, and services, improved data management, improved access to public services, informed decision-making, and increased problem-solving capacity.

The barriers to using open data portals include political and legal factors such as inadequate policies and laws, cultural factors such as a lack of organizational culture favorable to open data, economic and financial factors such as a lack of budget and high costs to hire people with the necessary skills, and operational and technical factors such as poor data quality and accessibility issues.

The enablers for using open data portals include political leadership commitment, awareness and promotion of open data by management and public managers, legal compliance pressure, organizational culture that facilitates open data use, availability of financial resources, promotion and internal disclosure of open data use, digitization capacity, motivated managers, public-private partnerships, and availability of high-quality data, tools, and mechanisms for working with open data (Kawashita et al., 2022).

Open government data platforms have proliferated worldwide, but their impact has been limited. To fully realize the benefits of open data, there needs to be a complex process of learning, interaction, and networking within and between government organizations and the community. This process goes beyond simply removing barriers. Living lab experiments have shown a positive correlation between the provision and usage of open data, but organizations are reluctant to scale up these experiments. The innovation challenge is to find a method to move beyond experimentation and recognize the extensive organizational effort required to make open data successful. Changing the larger macro context by developing legal and informal standards that encourage discussion about open data policies may also be necessary to scale up the use of open data (E. Ruijer & Meijer, 2020).

In fact most of the research that studies the progress of OGD after its adoption shows that "(…) the implementation of OGD was not nearly as successful as some individuals had hoped it would be" (Haini et al., 2020). Due to a lack of qualified personnel, IT infrastructure, political commitment, and outside influence, several government agencies found it difficult to demonstrate high performance in OGD initiatives (Haini et al., 2020).

To understand what factors contribute to OGD implementation after its adoption (to guarantee its long-term viability in the eyes of the data suppliers), Mustapa (2022) conducted a study to analyze the post-adoption phase: the findings show that the post-adoption phase of OGD is influenced by corporate culture, relative advantage, complexity, and top management support. As a result, it is expected of the data providers to help maintaining OGD while considering the points raised by the stakeholders. This will enable the OGD implementation strategy to be established and remain relevant in the years to come (Mustapa et al., 2022).

2.3. OPEN DATA BENCHMARKING & METRICS

Benchmarking is becoming more crucial for developing countries' OGD portals (Lnenicka et al., 2022). An overview of the most widely used maturity benchmarks can be found in Table 2.1.

Table 2.1 - Overview of open data indices and rankings published by international organizations
(Lnenicka et al., 2022)

Title	Publisher	First Report	Last Report	Nr. of Reports
GODI	OKF	2013	2016	4
ODB	W3F	2013	2017	5
OURdata Index	OECD	2015	2019	3
ODIN	ODW	2015	2020	5
ODMR	EU	2015	2021	7
OGDI	UN	2020	2020	1

The Global Open Data Index (GODI), Open Data Barometer, OURdata Index, and Open Data Inventory (ODIN) are all benchmarks that assess the openness and coverage of open data in different countries. These indices evaluate various aspects of open data such as legal and technical openness, readiness, implementation, impact, sustainability, and data coverage. The methodologies of these indices differ, and some rely on expert surveys, dataset analysis, and government self-assessments, while others use statistical indicators and data types to evaluate data coverage. The Online Government Development Index (OGDI) is a less transparent benchmark developed by the United Nations, which assesses policy and institutional frameworks, national portals, and data availability in specific areas.

The Open Data Maturity Report (ODMR) is a benchmarking tool developed by the European Union that assesses the maturity of open data systems in European countries based on open data policies, portals, impact, and quality (Lnenicka et al., 2022).

They measure in an aggregated manner four dimensions, and within those four dimensions there are business metrics (Figure 2.1):

Dimension	Metrics
Open Data Policy	Policy framework
	Governance of open data
	Open data implementation
Open Data Impact	Strategic awareness
	Political impact
	Social impact
	Environmental impact
	Economic impact
Open Data Portal	Portal features
	Portal usage
	Data provision
	Portal sustainability
Open Data Quality	Currency
	Monitoring and measures
	DCAT-AP compliance
	Deployment quality and linked data

Figure 2.1 – Metrics used by dimensional grouping (Publications Office of the European Union., 2022)

One of the conclusions of this year's report was that the scores of several countries are very high and "(…) gradually approaching 100%" (Publications Office of the European Union., 2022). That being so, it becomes necessary to continually improve the revision system and methodology in order to "(…) stimulate the Member States to continue to improve and grow beyond the current assessment, while ensuring consistency and comparability of the results" (Publications Office of the European Union., 2022).

As shown in Figure 2.2, Portugal still has some work to do to approach the European average and to leave the current "Beginner" clustering (Figure 2.2). In the Annexes of this thesis some guidelines from the ODMR can be found to tackle this issue.

Even though the trend of the last years shows the growth in overall Open Data maturity in Portugal, all the countries are making a similar effort, meaning that a higher score is needed to belong to a certain cluster (Publications Office of the European Union., 2022):
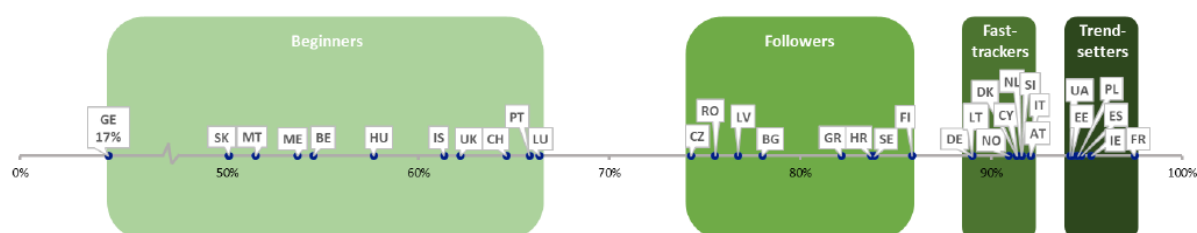

Figure 2.2 - Clustering of the 2021 maturity score of the participating European countries
(Publications Office of the European Union., 2022)

The results, including the cut-off points of the clusters, are listed below. Countries marked with an asterisk (*) are not part of the EU27 (Publications Office of the European Union., 2022):

- **Trend-setters (94%-98%):** France, Ireland, Spain, Poland, Estonia, and Ukraine*

- **Fast-trackers (89%-92%):** Austria, Italy, Slovenia, Netherlands, Cyprus, Denmark, Norway*, Lithuania, and Germany

- **Followers (74%-86%):** Finland, Sweden, Croatia, Greece, Bulgaria, Latvia, Romania, Czech Republic

- **Beginners (17%-66%):** Luxembourg, **Portugal**, Switzerland*, the UK*, Iceland*, Hungary, Belgium, Montenegro*, Malta, Slovakia, and Georgia*

Portugal showed significant improvement, increasing its score from 48% to 66%. The structure and scoring of points throughout the years is in the Annex, as well as additional information regarding other indices and rankings being used in the scientific literature that were not selected for the criteria and weights analysis.

The analysis of the main contributing factors to a good score on open government data benchmarks reveals that while many benchmarks claim to measure the impact of open data, they often only consider the presence of conditions for impact to occur, rather than the actual impact made. Some benchmarks focus on data publication, others on data use, and a few adopt a comprehensive approach to evaluating the concept of open data. The lack of consideration for established partnerships between data producers and users, and actual participation by citizens and actors in the use of open data, is a common shortcoming of these benchmarks. Although few focus on value generation, several models in the literature highlight the importance of participation and user involvement in more advanced open government data programs. The methods used for gathering data on the development of open data vary, including the number of countries covered, information sources, frequency of benchmarking, and validity checks. These methods may change over time for practical reasons, rather than based on new research from the open data literature. (Zuiderwijk et al., 2021).

It is important to understand the metrics used to assess the quality of open data across the world. Since the appearance of the first OGD portals, several scholars have created assessment models for them, concentrating on various facets, including the portal's maturity, the usability of the data, or the degree of stakeholder involvement—all of which relate to public effectiveness and openness. Some scholars concentrated on the openness and accountability of the data in the portals, which they thought would lead to efficiency and openness in government. Data content and metadata quality are frequently used as benchmarking tools (de Juana-Espinosa & Luján-Mora, 2019).

de Juana-Espinosa & Luján-Mora (2019) considered the national OGD portals of the EU-27 plus the UK over a period of three years, using cluster analysis to conclude that the countries are gradually harmonizing their OGD practices, which is consistent with the EU harmonization process and the creation of the single market, albeit individual nations are still free spirits and can maintain their unique cultural traditions.

Studies show that a significant percentage of these free datasets are of poor quality, which limits their usability even as they swarm and the platforms that publish them multiply. Data is only as helpful as

its quality, even while gathering, opening, sharing, and publishing such vast volumes of data is undoubtedly a positive move (Hassine & Clément, 2020).

The approach taken by the two most popular Open Data platforms in France (ETALAB and OpenDataSoft) differs according to the type of data provisioned. For ETALAB (the Data Office of the French government), with regards to Reference Datasets, "(…) the publishers are required to ensure the quality of the metadata of the dataset they publish, and the re-users are expected to define their optimum data quality level suited for their use case". Indeed, various people may utilize the same data collection in various ways, and their standards for the quality of the data may differ substantially. For some re-users, incomplete data may be preferable to no data. Therefore, it is challenging to identify a single set of data quality metrics outside of any context related to the reusability (Hassine & Clément, 2020); nevertheless, we can see in Table 2.2 three quality indicators that ETALAB tracks for reference datasets.

Table 2.2 - Quality indicators used for reference datasets (Hassine & Clément, 2020).

Quality Indicator	Definition
Number of Downloads	Number of downloads of the file since its release
Number of users feedbacks	Number of comments
Completeness of metadata description	Percentage of filled out metadata values

There are also metadata requirements that must be provided with the dataset (Table 2.3):

Table 2.3 - Mandatory metadata as per ETALAB (Hassine & Clément, 2020).

Metadata	Definition
Title	The title of your dataset should be specific and the most precise possible. It needs to be searchable through search strings by re-users
Description	The description of your dataset allows potential re-users to get information on the content and the structure of the published data, the context related to its production, the contact references of the publishers, etc. The dataset description is typically the first thing that potential re-users read when they discover your dataset
Update Frequency	Update frequency of the file

The assessment of open data quality is crucial for enabling informed and effective use of data by various stakeholders, including researchers, policymakers, and citizens. Open data aggregators such as OpenDataSoft play a vital role in making data available and accessible to the public. However, assessing the quality of the published datasets is a complex and multidimensional task, as it involves evaluating not only the content but also the metadata and other contextual indicators such as credibility, auditability, and interlinking.

To address this challenge, a comprehensive and standardized method for assessing the quality of open data portals is needed. Such a method should incorporate relevant quality dimensions and related metrics that capture different aspects of data quality, including auditability, trustworthiness, relevance, accessibility, understandability, visibility, timeliness, coverage, uniqueness, and interlinking.

Moreover, the method should be designed to facilitate automatic benchmarking of different open data portals, allowing for efficient and scalable evaluation of data quality across multiple platforms.

The development of a standardized method for assessing open data quality is essential for promoting transparency, accountability, and trust in the open data ecosystem. By providing a systematic and objective approach to evaluating data quality, such a method can help users identify high-quality datasets that meet their specific needs and ensure that open data portals meet minimum quality standards. Therefore, it is critical to invest in the development of effective tools and technologies that enable automatic and standardized benchmarking of open data portals, thus promoting open and accountable governance and empowering citizens to participate more actively in decision-making processes (Hassine & Clément, 2020).

There can also found in the literature other activity level analysis on the OGD portals, such as (Aarshi et al., 2018):

- Open data visitor's activity
- Portal supplier's activity
- Applications development related activities
- Activity related to the generation of new knowledge
- Overall resource usage available on the portal activities.

After the COVID-19 pandemic broke out, several open data portals have emerged that have set out to disseminate regional, national, and even international information about Covid-19 information (Sampaio et al., 2022). An attempt to benchmark (based on a bibliographic review of the most diverse methodologies proposed to evaluate OGD portals) some of the most widely used portals focused on measuring (from 0 to 5, where 0 is "Strongly Disagree" and 5 is "Strongly Agree") twenty-eight aggregated metrics (Máchová & Lnénicka, 2017).

A recent study on the extent of metadata compliance in terms of completeness and consistency, of three OGD leading country portals (Canada, USA, New Zealand) shows that even for required fields, their presence is not 100% true. The results also show a need to further improve the consistency of both required and optional fields. It further concludes that In terms of enforcing the consistency and completeness of their published metadata, OGD portal administrators are not involved in a quality assurance process with their own OGD publication regulations (and it is worrying because the study was conducted with leading OGD countries) (Šlibar & Mu, 2022).

Only two metadata attributes—completeness and consistency—were examined in this work because they are thought to be the most significant and prevalent aspects of metadata quality and because creating automated enforcement methods for them is very simple. The results of the inadequate compliance raise questions about how well the other quality dimensions are conforming (Šlibar & Mu, 2022).

2.4. OPEN DATA INITIATIVES IN PORTUGAL

There is a proliferation of Open Data portals in Portugal, which are summarized in the landscape page below (Table 2.4), which will serve as objects of our benchmark when analyzing if they are referenced in scientific publications. Specific initiatives related solely to a limited academic scope were purposefully left out (Biodata, Integrall, Yeastract, CorkOakDB, Mitobreak, KiMoSys and Haeckaliens)

because the benchmark focuses on broader open government data (OGD) initiatives, not just those in the academic domain. The benchmark is intended to evaluate the overall OGD initiatives of the Portuguese government and not just the academic ones. Another omission is the PorData portal, which was left out for not qualifying as an Open Data portal, since it does not provision free datasets for public use; what it does is aggregate, compile, analyze and disseminate data from multiple sources. There may be other significant omissions in the benchmark, making the Portuguese OGD portals list incomplete; the criteria for selection of the portals were a comprehensive search, using search engines, databases, and relevant websites.

Furthermore, Portuguese open data portals may not be easily identifiable as such, as they may be hosted on different platforms or websites (a centralized list, made at a European level could be helpful). When assessing the quality and effectiveness of open government data (OGD) portals in Portugal, it is important to consider the number of users of each portal. By focusing on the most used portals, we can ensure that our analysis is based on data that is most relevant to the majority of users.

Moreover, it is worth noting that the OGD portals with the most users are often the most well-maintained and frequently updated. These portals also tend to have better quality data, as they are used by a larger audience and are therefore subject to more scrutiny and feedback.

Therefore, it is reasonable to include only the most used OGD portals in Portugal in our analysis, as doing so ensures that our findings are based on data that is most relevant, high-quality, and manageable.

The Portuguese government sponsors several initiatives that produce positive results, such as the Open Government Partnership (OGP). The OGP promotes transparency, public participation, and the use of technology to strengthen democracy. The OGP supports initiatives that aim to combat corruption and make concrete commitments to achieve these goals (Portugal Action Plan 2021-2023, 2021). Some initiatives supported by the OGP are:

- Arquivo.pt, besides safeguarding information for the future also organizes an annual contest for innovative works based on open data.
- *Portal Único da Educação*, which aims at aggregating data portals from public education entities.

However, Portugal still requires some development to centralize open portals' data publication and catch up with the EU standards. In that sense, some action points are being developed to address the current gap are:

- Identify and list which data providers are not yet publishing.
- Provide an instruction manual on how to provision data to the national portal.
- Tailored workshops and training, technical integration.
- Events that promote open data and its reuse.
- Implementing an 'Open Data National Strategy' (embodied by the open data initiative promoted by the Agency for Administrative Modernization, IP - AMA).
- The creation of data catalogs detailing which data cannot be opened and which data can be opened but is not yet published.

- Moreover, through 'Portugal Digital' (https://portugaldigital.gov.pt) there is the ability to survey data/data maturity and certify it in your company, set up webinars to raise awareness on the concept of open data concept and its benefits, define methodologies to monitor and evaluate the impact of open data, and design a user satisfaction survey.

Another important and urgent aspect that should be considered is access to real-time and dynamic data. This can include for example air quality data, weather data, traffic information, or even data on newly born babies. It is important to note though that most countries offer less than 10% of their data in real-time. Only Denmark, Latvia, and the Netherlands report having more than 30% of the metadata linked to real-time or dynamic data (Publications Office of the European Union., 2022).

The Local Open Data Census, created in 2014 by the OKFN, is a free tool to help civil society monitor the publishing of open data at a local, regional, and municipal level for each country. It proposes a set of indicators that assess the degree of openness of a given local government (de Castro Neto et al., 2017). The assessment of data collected from the census shows that most of the defined indicators (13 of 15) are provided by the central government, while the rest are provided by local government or private companies. Most of the data considered open by municipalities is legally obliged to be made available. The majority of municipalities still lack a strategy for sharing data and view it as fulfilling legal obligations rather than increasing transparency and creating business opportunities. There is room for improvement and support in building data sharing strategies and platforms by municipalities or inter-municipal communities (de Castro Neto et al., 2017).

Table 2.4 – Portuguese Open Data portals

Portal Name	Year of Creation	Mission	Governance	Datasets available	API Available
dados.gov **dados.gov.pt**	2012	Aggregate, reference and host open data from different bodies and sectors of the Portuguese Public Administration. Besides functioning as a shared service for hosting and publishing data, which can be used by any public body, it also functions as an indexing portal for content hosted on other open data portals/catalogs, whether sectoral (e.g., Health, Justice, Environment) or local (e.g., Lisbon City Hall, Águeda City Hall).	Agency for Administrative Modernization, IP (AMA)	5 684	Yes
Portal base: **www.base.gov.pt/base4**	2013	Inform about public procurement in Portugal.	Institute of Public Markets for Real Estate and Construction	1 585 983	No
Lisboa Aberta **lisboaaberta.cm-lisboa.pt** **lxi.cm-lisboa.pt**	2018	The strategy for open data followed by the municipality of Lisbon assumes that the information produced by municipal services, municipal companies and the various actors operating in the city, should be made available openly to all who seek it, contributing to a policy of transparency, citizen engagement, and the provision of better services.	Lisbon Municipality	371	No
Cascais Data **data.cascais.pt**	2017	Strengthen and support the development of a culture of dissemination of data and public management information, to build an ecosystem of reuse and aggregation of data value for the development of innovative solutions that qualify the life of the municipality	Cascais Municipality	191	No
Rural Dados.pt **www.ruraldados.pt**	-	The specific needs of the National Observatory for Biological Production (ONPB) and the National Desertification Observatory (OND) led to the adoption of a strategy to develop a common data infrastructure to support the construction of online monitoring tools for the evolution of indicators in each of the Observatories.	Ministry of Agriculture, Rural Development and Forestry	56	Yes
INE **www.ine.pt**	-	The Mission of Statistics Portugal is to produce, in an independent manner, high-quality official statistical information, relevant for the society, while promoting the coordination, the analysis, the innovation and the dissemination of the national statistical activity and ensuring integrated data storage.	INE, IP	9 850	Yes
DG Territorio **snig.dgterritorio.gov.pt/**	2018	The SNIG is the National Infrastructure for Geographic Information, which allows the registration and search of data and geographic data services produced by public and private entities in Portugal.	Directorate-General of the Territory (DGT)	4 581	Yes

Central de Dados centraldedados.pt	2015	The Central de Dados is an independent portal where various raw datasets are made available, relating to public information of all kinds, with a national and/or local scale. The data come from various public sources, are converted to open and standard formats, and are then made available following the principles of open data: open formats and free access, free of charge and without restrictions on reuse and redistribution.	Open Knowledge Portugal	13	Yes
BNP opendata.bnportugal.gov.pt/index.htm	2003	The National Library of Portugal (BNP) provides, on these pages, information and access to the data sets it makes freely available: the entire BNP bibliographic catalog, the National Bibliographic Database - PORBASE, the Portuguese National Bibliography (since 1931) and the Digital National Library, among others.	National Library of Portugal (BNP)	15	No
RCAAP dados.rcaap.pt	2008	RCAAP portal aims to collect, aggregate and index Open Access scientific contents from Portuguese institutional repositories. RCAAP constitutes a single-entry point for searching, discovery and recall of thousands of scientific and scholarly publications, namely journal articles, conference papers, thesis, and dissertations, distributed by several Portuguese repositories.	RCAAP, FCT/FCCN	2075	No
SNIRH snirh.apambiente.pt	1997	Make available, via Internet, static and dynamic pages with direct access to the SNIRH Database. The national water resources monitoring system is supported by a database prepared to store and publicly disseminate hydro-meteorological and water quality data (surface and underground), collected from the water resources monitoring network of the Ministry of Environment (through the system's portal https://snirh.apambiente.pt)	SNIRH	7210	No
SNS transparencia.sns.gov.pt	2016	The Transparency Area is an Open Data initiative carried out by the Ministry of Health, to make available and fully accessible the vast set of data underlying the operations and transactions that take place within the scope of the activities of the National Health System (SNS), in the various entities. The idea is to centralize, on an online platform with accessible and intuitive use, the data produced by the systems within the SNS, so that this information can be observed, analyzed, and reused, without any restriction or difficulty, by the general population, without requiring any authorization.	SNS	157	Yes

METHODOLOGY

In this section we discuss the methods used to benchmark the Portuguese open data portals; namely how the data collection method was perpetrated and also mention any limitations or assumptions that may impact the validity of the study's results.

3.1. DATA COLLECTION – SEARCH STRATEGY

We use the PRISMA (Transparent Reporting of Systematic Reviews and Meta-analysis) methodology to construct the bibliometric dataset that will be the focus of analysis. The PRISMA approach was designed for clinical medicine and is a well-established guideline for reporting systematic reviews and meta-analyses. Still, it also works well for a systematic literature review on any general topic and has the advantages of being transparent and replicable (E. H. J. M. Ruijer & Martinius, 2017). The PRISMA approach was chosen because it provides a standardized and transparent framework for reporting the methods and results of the study, enabling replication and evaluation by others.

The steps conducted to perform the extraction and evaluation of the literature using the PRISMA methodology are detailed in Figure 3.1.

First we performed a search string in Scopus using the following input **REF (dados.gov.pt OR www.ine.pt OR rcaap OR snig.dgterritorio.gov.pt OR data.cascais.pt OR base.gov.pt OR www.ruraldados.pt OR lisboaaberta.cm-lisboa.pt OR lxi.cm-lisboa.pt OR snirh.apambiente.pt OR transparencia.sns.gov.pt OR opendata.bnportugal.gov.pt)** in the References section of the articles only.
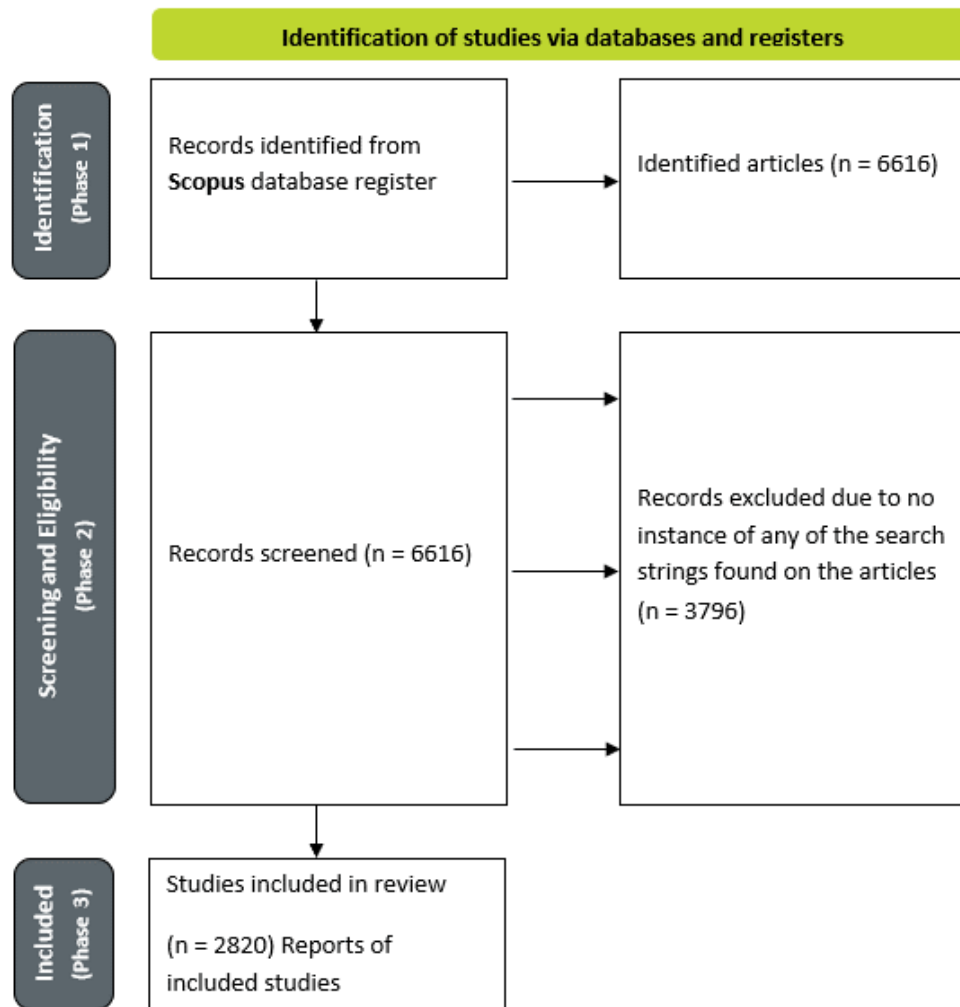
Figure 3.1 - PRISMA (adapted) flow chart: approach

The resulting dataset consisted of 6616 documents, but after screening and curation due to no actual Portuguese OGD portal used, the number reduced to 2820. The method used to adjust the dataset (creating a specific column in the dataset to enable the analysis of the Portuguese OGD portals) to make the different visualization possible through the Vos Viewer software is described as follows:

1. Export of the entire 6616 documents from the Scopus site in .CSV format (selecting only the Document title, the Year and the References),
2. Import the data into Excel using the 'Get Data from Text/CSV' function,
3. Add a column for each of the Portuguese OGD portals that figured in the search query,
4. Use an Excel formula to search within the References cell the desired keyword (example of formula used: **=IF(ISNUMBER(SEARCH("www.ine.pt";'scopus(5)'!$D2))=TRUE;"INE I.P.";"")**),
5. After searching for each of the portals, use a TEXTJOIN formula to concatenate the results (**=TEXTJOIN (", "; TRUE; E2:M2)**),
6. Filter out the joined portals column that do not contain any of the Portuguese OGD portals in study,
7. Export the range of the dataset with a comma delimiter, encapsulating the text within double quotes (so that the data is read properly in Vos Viewer tool).

BIBLIOMETRIC ANALYSIS

This study – Bibliometric Analysis – has an initial primary focus: to perform an adapted co-reference analysis on the dataset made up of all pertinent open data academic literature using the enrichment techniques of clustering and visualization. Despite having a fundamentally distinct difference, clustering and visualization approach frequently work together: for the dataset analyzed in this work, they are used simultaneously. The VOS viewer software tool was used to create the network maps and explore the insights from the data it generated.

To be able to perform the bibliometric analysis that we set out to do, given the limitations of the current way that data sources are cited, we had to 'fool' the VOS Viewer software by replacing the data under the 'Authors' dataset attribute with the names of the Portuguese OGD portals that are sourced within each document.

To start the analysis of the dataset that was imported to VOS Viewer software we chose to create a map based on bibliographic data. Next, we chose the 'co-authorship' (even though we are not analyzing the authorship of the items, this is our best choice to study their relatedness, given that we adjusted the dataset) type of analysis and chose one as the minimum number of documents for an OGD portal (this way all the nine OGD portals appear in the visualizations/ diagrams).

In Table 4.1 an aggregation of the results obtained from the VOS Viewer software is shown. The first line of the table shows the number of manuscripts in which the OGD portal was referenced, the second line indicates the total link strength (the number shows how many times the OGD portal was referenced in combination with another) of the portal and the third shows the number of links (number of other OGD portals commonly referenced). The fourth line indicates the cluster which was attributed to the specific OGD portal in the Network visualization (Figure 4.1) and the fifth line shows the average publication year of articles that use the specific OGD portal as reference (detailed in Figure 4.2).

Table 4.1 - Vos Viewer results aggregation table

	Number of documents	Total link strength	Links	Cluster	Average publication year
INE I.P.	2193	117	8	Red	2017,24
AMA I.P.	15	5	3	Red	2019,93
RCAAP	429	38	5	Green	2019,28
SNIG	37	15	4	Red	2020,32
Cascais Data	1	1	1	Purple	2022
Portal Base	16	7	3	Yellow	2017,88
Lisboa Aberta	44	32	4	Blue	2018,61
APA I.P.	165	24	17	Red	2019,11
SNS Data	61	53	4	Green	2020,41

In Figure 4.1, it is depicted that four clusters emerged through the use of VOS viewer which partitions the terminology into clusters based on their relevance to the co-referenced OGD portals in the articles.

The normalization method used was "association strength," which was the default option. The closer proximity of the nodes indicates a stronger relationship between them.

The popularity of the OGD portals referenced in the articles is depicted by the size of the nodes in Figure 4.1. The largest node represents the most frequently cited OGD portal, which was determined
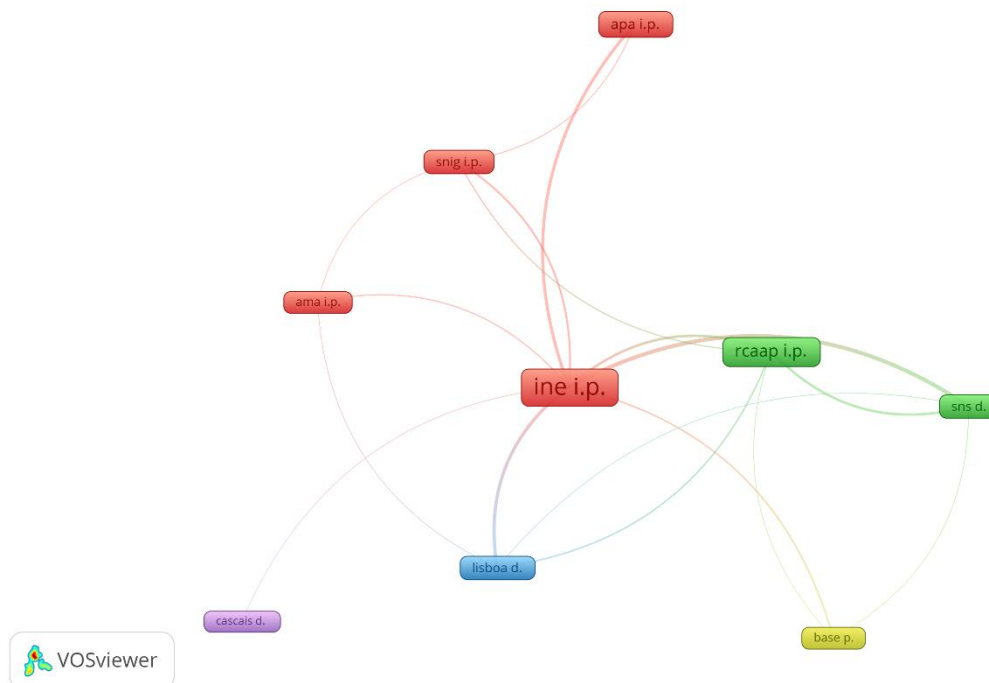
Figure 4.1 – Network visualization of Portuguese OGD portals: Weight – Number of Documents

by counting the number of references in the articles. The visual representation of the INE portal, including its size and its connections to all other portals, confirms its dominant position as the source that was cited the most frequently in the documents.

The width of the links in Figure 4.1 reveals a clear pattern of relationship between the portals. The thicker the link between two portals, the more frequently they are cited together in the data set. This can be observed in the link between INE and SNS, which is the thickest among all the links, with 37 documents sourcing both OGD portals.

The average publication year of the articles can be observed in Figure 4.2 using overlay visualization. The lowest average is seen in INE and Portal BASE, possibly due to their long existence. The highest average (2020,41) is seen in SNS, due to its recent establishment as an OGD portal.
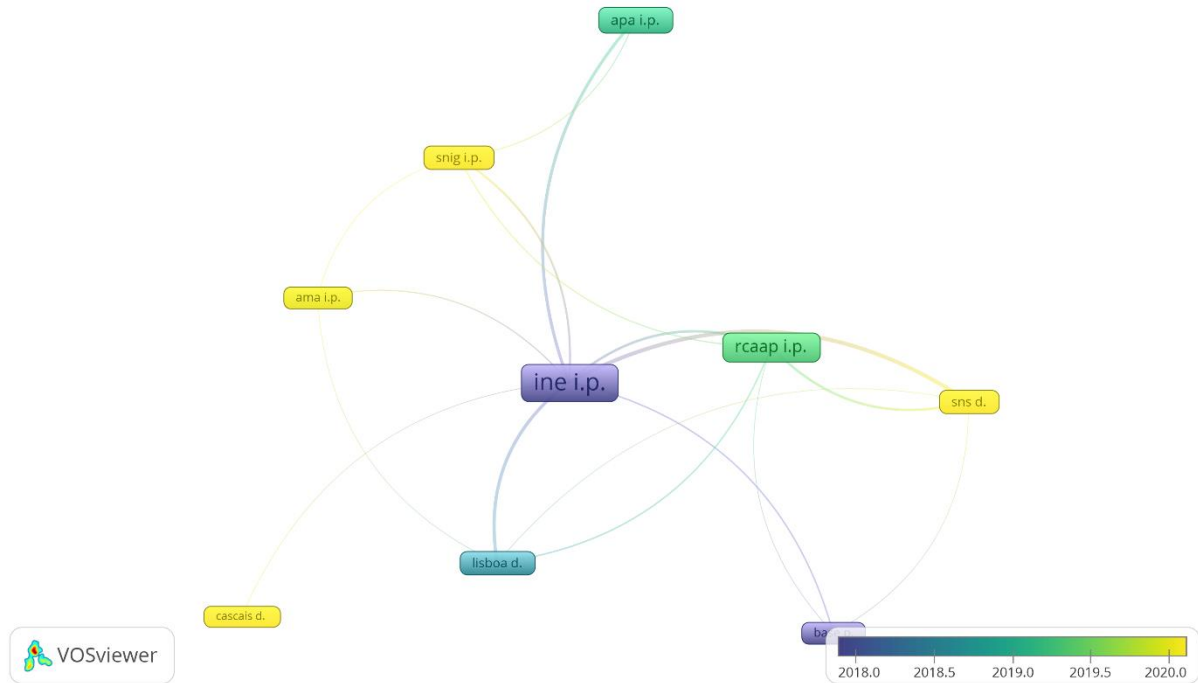
Figure 4.2 - Overlay visualization of Portuguese OGD portals: Curated dataset

A decision was made to extend the analysis of the Portuguese OGD portals' dataset, based on the positive results obtained using the VOS Viewer software. The objective of the supplementary analysis is to gain deeper insights into the scientific publications that cite the Portuguese OGD portals, focusing on the following aspects:

- Co-authorship (the relatedness of items is determined based on their number of co-authored documents):
    - Authors
    - Organizations
    - Countries
    - Journals
- Co-occurrence (the relatedness of items is determined based on the number of documents in which they occur together):
    - Author keywords
- Co-citation (the relatedness of items is determined based on the number of times they are cited together):
    - Cited authors

Co-citation analysis examines the inter-connected structure of cited articles by analyzing the frequency of joint citations. Articles that are frequently cited together form clusters of research, providing insights into significant themes in the field of investigation (Rani et al., 2022).

Figure 4.3 displays the co-authorship network among authors in the dataset, with 7270 authors being considered. 59 authors met the threshold criteria of having a minimum of 10 documents and 10 citations. Table 4.2 presents the results of number of citations, number of documents produced by authors, and the link strength of authors.
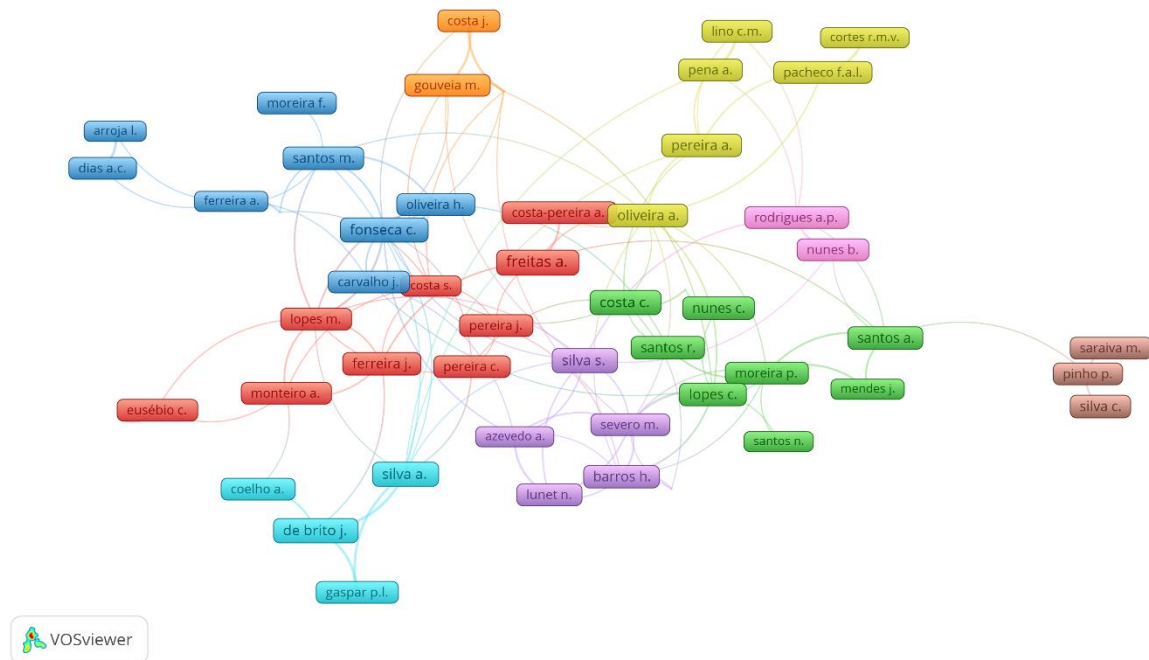
Figure 4.3 - Network visualization of co-authorship of curated dataset: Weight – Number of Documents

Table 4.2 – Top 20 of authors sorted by the Number of Documents produced

Author	Documents	Citations	Total link strength
freitas a.	37	242	12
silva s.	26	98	14
silva a.	25	280	31
costa c.	23	138	8
fonseca c.	23	145	26
de brito j.	22	822	30
barros h.	20	455	19
nunes c.	20	330	2
silva c.	20	128	2
ferreira j.	19	286	10
santos a.	19	216	12
lopes c.	18	119	15
santos r.	18	320	6
oliveira a.	17	155	14
santos m.	17	151	15
pereira a.	16	97	9

carvalho j.	14	129	6
nunes b.	14	118	9
pena a.	14	303	15
pereira c.	14	35	6

Figure 4.4 represents the co-authorship network based on the organizations that contributed to creating the 2820 documents in the analyzed dataset. Out of the 7285 organizations, 58 met the criteria of having at least 4 documents and 4 citations. The VOS Viewer software utilized "association strength" normalization method and identified 4 clusters. It should be noted that most organizations have a total link strength of zero, appearing as scattered on the graph. Table 4.3 lists the top 20 organizations with the highest number of citations in the publication references.

Figure 4.4 - Network visualization of co-authorship by Organization: Weight – Number of Documents

Table 4.3 - Top 20 of Organizations considering the sort by number of Citations

Organization	Documents	Citations	Total link strength
faculty of medicine, university of coimbra, coimbra, portugal	8	410	1
cesam, department of environment and planning, university of aveiro, 3810-193 aveiro, portugal	4	238	0
departamento de ciências da saúde pública e forenses e educação médica, faculdade de medicina, universidade do porto, porto, portugal	12	152	8
cense – center for environmental and sustainability research, nova school of science and technology, nova university lisbon, portugal	4	139	0

chemistry research centre, university of trás-os-montes and alto douro, ap. 1013, vila real, 5001-801, portugal	7	132	6
centre for the research and technology of agro-environment and biological sciences, university of trás-os-montes and alto douro, ap. 1013, vila real, 5001-801, portugal	6	128	6
university of lisbon, portugal	6	111	4
department of physical therapy, school of health technology of porto, polytechnic institute of porto, vila nova de gaia, portugal	4	105	0
school of economics and management, university of minho, braga, portugal	4	95	0
icvs/3b's, pt government associate laboratory, braga/guimarães, portugal	5	82	0
universidade de trás-os-montes e alto douro, vila real, portugal	4	74	0
ceabn/inbio, centro de ecologia aplicada "professor baeta neves", instituto superior de agronomia, universidade de lisboa, tapada da ajuda, lisboa, 1349-017, portugal	4	72	0
escola nacional de saúde pública, universidade nova de lisboa, lisbon, portugal	5	72	0
university of minho, braga, portugal	7	67	3
school of psychology, university of minho, campus de gualtar, braga, 4710-057, portugal	4	62	0
instituto de ciências sociais, universidade de lisboa, lisboa, portugal	5	61	3
university of coimbra, portugal	10	59	5
laboratório de farmacologia clínica e terapêutica, faculdade de medicina, universidade de lisboa, lisboa, portugal	6	57	11
department of clinical epidemiology, predictive medicine and public health, university of porto medical school, porto, portugal	6	54	2
department of community medicine, information and health decision sciences (medcids), faculty of medicine, university of porto, porto, portugal	5	53	1

Figure 4.5 exhibits the network of co-authorship among countries in the dataset. Out of the 119 countries present, 42 meet the requirement of having at least 4 documents for representation on the graph. Table 4.4 provides a list of the top 10 countries with the most citations.
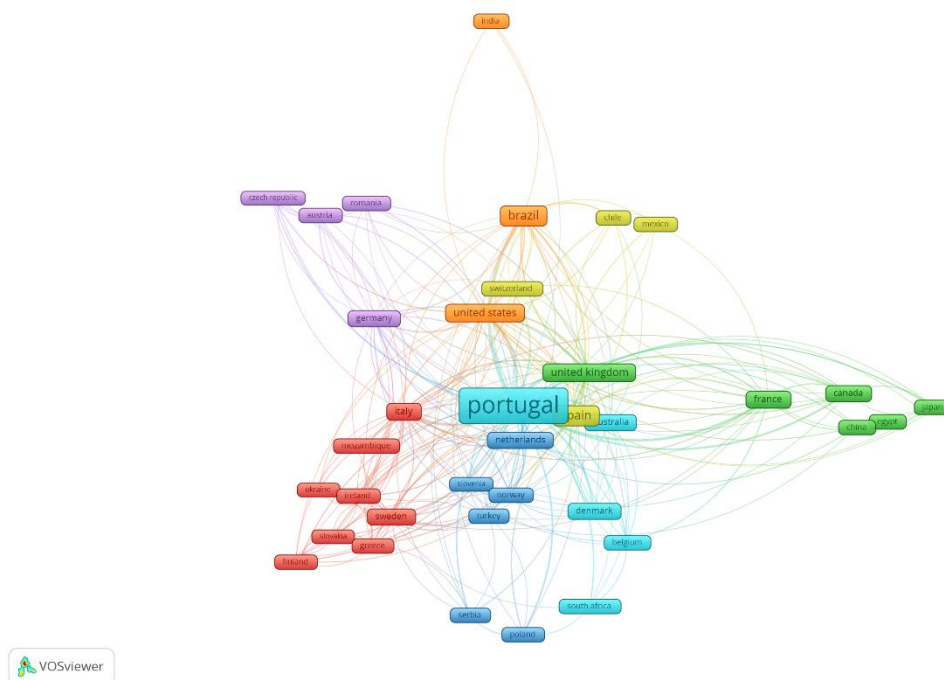
Figure 4.5 - Network visualization of co-authorship by Country: Weight – Number of Citations

Table 4.4 - Top 10 of countries considering the sort by number of Citations

Country	Documents	Citations	Total link strength
Portugal	2356	25387	834
Brazil	259	1599	162
Spain	254	2591	291
United States	139	2682	178
United Kingdom	135	1964	186
Italy	60	976	105
France	52	877	73
Netherlands	49	887	86
Germany	46	729	82
Australia	31	487	44

In Figure 4.6, a network of co-authorship between sources is displayed. Out of the 1505 sources present in the dataset, 46 sources met the requirement of having at least 7 documents and 7 citations. Table 4.5 presents a summary of the results by showcasing the top 20 journals with the most articles published in terms of the number of citations and documents produced by the authors.
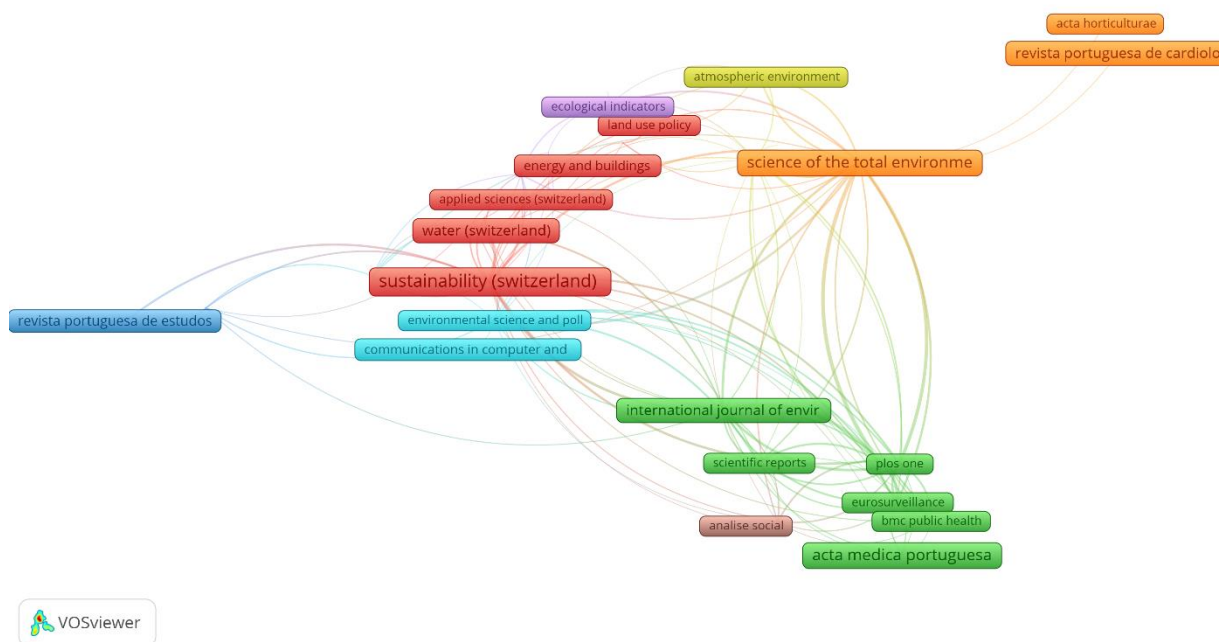
Figure 4.6 - Network visualization of Sources: Weight – Number of Documents

Table 4.5 - Top 20 journals considering the sort by number of Documents produced

Source	Documents	Citations	Total link strength
sustainability (switzerland)	69	411	255
acta medica portuguesa	50	322	12
science of the total environment	47	1359	381
international journal of environmental research and public health	32	246	236
journal of cleaner production	29	1270	60
revista portuguesa de cardiologia	26	203	1
water (switzerland)	26	199	66
iberian conference on information systems and technologies, cisti	25	26	1
revista portuguesa de estudos regionais	21	38	5
advances in intelligent systems and computing	20	31	2
communications in computer and information science	16	17	1
energies	15	68	22
revista portuguesa de saude publica	15	55	0
sociologia, problemas e praticas	15	33	0
revista de enfermagem referencia	15	17	0
energy and buildings	14	552	11
journal of tourism and development	12	15	26
plos one	11	265	335
ecological indicators	11	165	28

| environmental science and pollution research | 11 | 94 | 95 |

The network of co-occurrence of Author Keywords used in the articles in the dataset is depicted in Figure 4.7. Out of the 8661 keywords present, 259 met the criteria of a minimum of 5 occurrences to be displayed on the graph. The top 20 keywords with the highest occurrences in the dataset are presented in Table 4.6.

Figure 4.7 - Network visualization of co-occurrence: Weight – Number of Occurrences

Table 4.6 - Top 20 of Author keyword occurrences, sorted by number of occurrences

Keyword	Occurrences	Total link strength
portugal	360	424
covid-19	51	83
climate change	40	30
sustainability	39	44
tourism	36	46
elderly	31	38
mortality	26	38
public health	25	35
sustainable development	23	20
lisbon	22	29
children	21	26
epidemiology	21	40
quality of life	21	22
risk factors	21	41

aging	20	30
energy efficiency	20	19
nursing	19	26
older adults	19	11
ageing	18	18
depression	18	38

To visually display the author network of co-citations, Figure 4.8 presents the data where out of 71747 authors in the dataset, 108 authors have received at least 20 citations. The size of nodes reflects the level of frequently co-cited authors. The top 20 most co-cited authors are listed in Table 4.7 in order of the number of citations received.
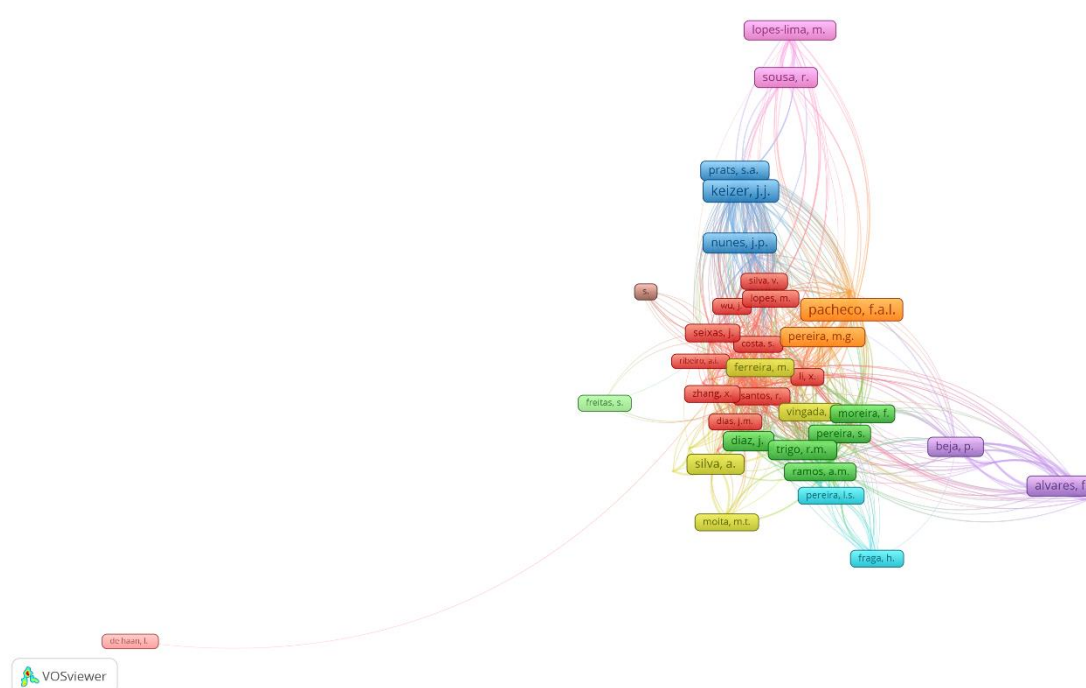
Figure 4.8 - Network visualization of co-citation between authors: Weight – Total Link Strength

Table 4.7 - Top 20 of co-cited authors considering the sort by number of Citations

Author	Citations	Total link strength
silva, a.	72	1098
keizer, j.j.	69	1881
pacheco, f.a.l.	58	1820
fonseca, c.	55	657
trigo, r.m.	54	889
seixas, j.	53	487
moreira, f.	51	571
de brito, j.	49	442
gouveia, j.p.	44	368

wang, y.	44	285
alvares, f.	43	1153
sanches fernandes, l.f.	41	1530
nunes, j.p.	39	961
liu, y.	39	257
zhang, y.	38	237
santos, m.	36	546
ferreira, m.	36	431
zezere, j.l.	35	452
fraga, h.	35	325

The assessment concludes with Table 4.8, which displays the 20 most frequently cited articles and their respective citation count, author information, publication year, and journal.

Table 4.8 – Top 20 most cited articles

Authors	Article Title	Citations	Journal	Year
lopes j.a.p.; soares f.j.; almeida p.m.r.	Integration of electric vehicles in the electric power system	1069	Proceedings of the IEEE	2011
sato t.; qadir m.; yamamoto s.; endo t.; zahoor a.	Global, regional, and country level need for data on wastewater generation, treatment, and use	357	Agricultural Water Management	2013
weiser t.g.; haynes a.b.; molina g.; lipsitz s.r.; esquivel m.m.; uribe-leitz t.; fu r.; azad t.; chao t.e.; berry w.r.; gawande a.a.	Size and distribution of the global volume of surgery in 2012	309	Bulletin of the World Health Organization	2016
hulteen r.m.; smith j.j.; morgan p.j.; barnett l.m.; hallal p.c.; colyvas k.; lubans d.r.	Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis	247	Preventive Medicine	2017
soares p.; rocha j.v.; moniz m.; gama a.; laires p.a.; pedro a.r.; dias s.; leite a.; nunes c.	The association between chronic disease and serious COVID-19 outcomes and its influence on risk perception: Survey study and database analysis	220	JMIR Public Health and Surveillance	2021
freitas s.; simões m.r.; alves l.; santana i.	Montreal Cognitive Assessment (MoCA): Normative study for the Portuguese population	217	Journal of Clinical and Experimental Neuropsychology	2011
loureiro s.m.c.; kastenholz e.	Corporate reputation, satisfaction, delight, and loyalty towards rural lodging units in Portugal	213	International Journal of Hospitality Management	2011
de oliveira e.m.; cyrino oliveira f.l.	Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods	205	Energy	2018

tizzoni m.; bajardi p.; decuyper a.; kon kam king g.; schneider c.m.; blondel v.; smoreda z.; gonzález m.c.; colizza v.	On the Use of Human Mobility Proxies for Modeling Epidemics	194	PLoS Computational Biology	2014
lisboa a.; skarmeas d.; lages c.	Entrepreneurial orientation, exploitative and explorative capabilities, and performance outcomes in export markets: A resource-based approach	184	Industrial Marketing Management	2011
coelho a.; de brito j.	Economic viability analysis of a construction and demolition waste recycling plant in Portugal - Part I: Location, materials, technology and economic analysis	162	Journal of Cleaner Production	2013
bing x.; bloemhof j.m.; ramos t.r.p.; barbosa-povoa a.p.; wong c.y.; van der vorst j.g.a.j.	Research challenges in municipal solid waste logistics management	149	Waste Management	2016
phithakkitnukoon s.; smoreda z.; olivier p.	Socio-geography of human mobility: A study using longitudinal mobile phone data	142	PLoS ONE	2012
loli? a.; paíga p.; santos l.h.m.l.; ramos s.; correia m.; delerue-matos c.	Assessment of non-steroidal anti-inflammatory and analgesic pharmaceuticals in seawaters of North of Portugal: Occurrence and environmental risk	140	Science of the Total Environment	2015
serpa d.; nunes j.p.; santos j.; sampaio e.; jacinto r.; veiga s.; lima j.c.; moreira m.; corte-real j.; keizer j.j.; abrantes n.	Impacts of climate and land use changes on the hydrological and erosion processes of two contrasting Mediterranean catchments	136	Science of the Total Environment	2015
wiesmann d.; lima azevedo i.; ferrão p.; fernández j.e.	Residential electricity consumption in Portugal: Findings from top-down and bottom-up models	130	Energy Policy	2011
ferreira l.; de brito j.; saikia n.	Influence of curing conditions on the mechanical performance of concrete containing recycled plastic aggregate	129	Construction and Building Materials	2012
cadima j.; leal t.; burchinal m.	The quality of teacher-student interactions: Associations with first graders' academic and behavioral outcomes	129	Journal of School Psychology	2010
pires a.; chang n.-b.; martinho g.	An AHP-based fuzzy interval TOPSIS assessment for sustainable expansion of the solid waste management system in Setúbal Peninsula, Portugal	123	Resources, Conservation and Recycling	2011
robaina-alves m.; moutinho v.; costa r.	Change in energy-related CO2 (carbon dioxide) emissions in Portuguese tourism: A decomposition analysis from 2000 to 2008	122	Journal of Cleaner Production	2016

RESULTS AND DISCUSSION

The purpose of this work is to understand and assess the impact of the contribution of Portuguese Open Government Data (OGD) portals on the world of scientific publications. The study begins by identifying the state-of-the-art in terms of OGD maturity assessments and how Portugal compares to other European nations. A literature review was conducted to gather information on the most commonly used OGD maturity assessments and metrics, as well as barriers and enablers of OGD adoption by different stakeholders. The study aims to produce clarity and objective gains in the use of OGD, with a focus on the Portuguese case. The study found that in most of the scientific studies analyzed, the reference to OGD sources is decentralized, and there is a lack of a standard way to reference the data sources, making it difficult to analyze automatically. The study also found that the Portuguese national OGD portal (dados.gov.pt) is not as commonly used as other portals, and there is a need to push for a standardized way to cite datasets.

The present study utilized a systematic PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach to survey and analyze scientific publications. Through this approach, a total of 6616 records were identified, which were then screened based on the predefined inclusion criteria. Following the screening process, 2820 records were deemed relevant and included in the final dataset for analysis. The use of PRISMA methodology also allowed us to maintain transparency and replicability in our search strategy, providing a strong foundation for the study's findings and conclusions.

Most of the documents in the dataset pertained to the subjects of Social Sciences, Medicine, Environmental Science, Mathematics, Engineering, and Computer Science. Most of the documents were Scientific Articles, followed by Conference Papers, primarily from journals, with a smaller proportion from Conference Proceedings. The country of origin for many of the documents was Portugal, followed by Brazil and Spain.

One of the main findings of this research is that in most of the scientific studies analyzed, there is a lack of centralization of OGD sources, concerning regional archives rather than the Portuguese national OGD portal (dados.gov.pt by AMA). Additionally, there is currently no standard way to reference data sources, which makes it challenging to analyze the data automatically. This is due to the scattered nature of the information within the publication and the confusion surrounding the correct credit for the data, whether it be to the national OGD portal or the original data provider.

The study results revealed a low usage of the Portuguese national OGD portal (dados.gov.pt by AMA) compared to other portals such as Instituto Nacional de Estatística and RCAAP - Repositório Científico de Acesso Aberto de Portugal. The usage of dados.gov.pt was found to be higher only than Cascais Data. This indicates that there is a need to increase the usage of the portal and to establish a standardized way of sourcing data from it.

It is crucial to emphasize the importance of standardizing how datasets are cited to measure metrics and correctly assign data provenance accurately. Our analysis of articles that cited the Portuguese national OGD portal (dados.gov.pt by AMA) found instances where the authors chose to credit the dados.gov website instead of the actual source of the data (SNIG – DG Territorio, in this case). This highlights the need for a standard method of citing data sources.

Notably, Cascais Data, SNIG, AMA, and SNS OGD portals were found to be commonly used as sources in publications from the year 2020, indicating that these portals are likely to be increasingly utilized in the future. Additionally, the study found that various portals had high 'total link strength' metrics, with INE having the highest number of links (117) followed by RCAAP, Lisboa Aberta, APA, and SNS with high scores. This suggests that many articles reference multiple Portuguese OGD portals as sources.

The results of the bibliometric analysis clearly show that INE is the most widely used Portuguese OGD portal. This is evident from the significant difference in the number of documents sourced from INE (2193) compared to the other portals (429 and 165) and its higher link strength. The only metric INE scored lower than the other portals is the average publication year (2017.2), which can be attributed to its long history compared to the other portals.

Our bibliometric analysis of the dataset revealed that INE is the most used Portuguese OGD portal, as demonstrated by its substantial lead in the number of documents (2193) and link strength compared to the other portals. In addition, further analysis was conducted to identify key contributors in the field, such as the most influential authors, prominent journals, frequently used keywords, notable universities/organizations, and dominant countries. These findings provide further insight into the composition and impact of the articles in the dataset.

The most prominent authors have been highlighted as shown in Table 4.2. The results of our analysis indicate that Alberto Freitas from the Faculty of Medicine at the University of Porto is the most frequently published author, with 37 articles to his name. Additionally, Jorge de Brito from the Instituto Superior Técnico was found to have the highest link strength score, indicating that his work is often referenced in combination with other authors.

Table 4.3 highlights the top 20 organizations the authors belong to, summarizing the number of documents produced and the number of citations. The Faculty of Medicine at Universidade de Coimbra had the most citations in the articles included in the dataset, with 8 documents and 410 citations. CESAM at Universidade de Aveiro followed with 4 documents and 238 citations, and the Departamento de Ciências da Saúde Pública e Forenses, e Educação Médica at Universidade do Porto had 12 documents and 152 citations.

Table 4.5 shows the top journals/sources in terms of published articles. The Sustainability journal from MDPI had the highest number of articles, followed by Acta Médica Portuguesa. Regarding citations, the Science of the Total Environment and the Journal of Cleaner Production had the most. PLOS One and Science of the Total Environment were the journals with the highest link strength.

The results of the co-citation analysis are presented in Figure 4.8 and Table 4.7, with Ana Silva from Instituto Superior Técnico, Jan Jacob Keizer from Universidade de Aveiro, and Fernando António Leal Pacheco from University of Trás-os-Montes and Alto Douro being the top authors. The co-occurrence of keywords is depicted in Figure 4.7 and Table 4.6, with terms such as "Portugal", "Covid-19", "climate change", "sustainability" and "tourism" being among the most frequently used in the dataset. Other notable keywords include "elderly", "mortality", "public health", "Lisbon", "epidemiology" and "risk factors". The network and clustering of countries referenced by the authors can be seen in Figure 4.5 and Table 4.4, with Portugal being the top country, followed by Brazil, Spain, the United States, and the United Kingdom. As the dataset comprises articles that use Portuguese OGD portals as a data

source, Portugal is expected to be the leading country in terms of number of documents published, citations, and link strength.

CONCLUSIONS AND FUTURE WORKS

The adoption of open government data (OGD) in Portugal is progressing positively, as evidenced by the emergence of multiple open data portals. However, the post-adoption phase of OGD is influenced by various stakeholders that can affect its success. Research suggests that factors such as corporate culture, relative advantage, complexity, and top management support positively impact OGD success in the post-adoption phase (Mustapa et al., 2022). Additionally, there is a positive correlation between OGD and entrepreneurship, with some nations excelling in realizing the entrepreneurial benefits of OGD (Huber et al., 2022).

To improve the success of OGD portals, it is recommended to use evidence from qualitative studies and needs assessments to inform future tools, and to include usability evaluations in the development stages of OGD visualizations. The usefulness of new visualization practices such as data storytelling and privacy-preserving visualizations should also be explored (Ansari et al., 2022).

One major barrier to better OGD usage is the quality of the open data. Therefore, developing a methodology for assessing flaws in public datasets and alerting government agencies to improve overall data quality and open processes is crucial. Some have also suggested a gamification approach to make OGD portals more accessible to lay citizens (Simonofski et al., 2022).

Assessing and building OGD capabilities at the local level is important for success, as it allows for engagement with the local community in the post-adoption phase to create new products and services based on OGD (Habib et al., 2022). A reference architecture that supports both publishing and data use is also necessary. Leveraging open data management platforms like DKAN can aid in the design of a full suite of cataloguing and visualizing user interactions (Luthfi & Janssen, 2022).

The limitations of this research include the use of a single database for data gathering and analysis, and the shortcomings of the VOS Viewer software. To mitigate these limitations in future research, it is important to standardize the format of references for automatic parsing tools and consider multiple databases. Additionally, it is important to note that the number of scientific papers published using a particular portal as a reference may not be the only or most important factor to consider when evaluating the performance of these portals. Other factors such as data quality, ease of use, and user satisfaction should also be considered.

As we conclude this thesis, it is important to note that while the benchmarking method was a valuable tool in assessing the state of open data portals in Portugal, it has its shortcomings. One of the main limitations of this method is that it leaves out many Portuguese open data portals, as there is no standardized and automatic way of citing the data sources used. In this study, the criteria for selection of the portals were a comprehensive search, using search engines, databases, and relevant websites.

Despite these limitations, the benchmarking method provided a valuable framework for evaluating the state of open data portals in Portugal. However, it is important to recognize that this method is not a one-size-fits-all solution and may need to be adapted and refined to suit different contexts' specific needs and characteristics. Moving forward, future research could explore alternative methods for assessing the state of open data portals that address the limitations of the benchmarking method.

In conclusion, this research demonstrates that it is possible to perform a bibliometric analysis to effectively and automatically benchmark OGD portals. This can provide a roadmap for future analysis and assessment of various subjects. However, it is important to consider a range of different benchmarks and metrics to get a comprehensive understanding of the performance of these portals. Moreover, a standard for citation in publications of OGD portals should be developed to give credit where credit is due, otherwise, the work developed so far goes unnoticed.

BIBLIOGRAPHICAL REFERENCES

Aarshi, S., Tariq, U., Hayat, B., Habib, F., Ashfaq, K., & Saleem, I. (2018). Dimensions of
Open Government Data Web Portals: A Case of Asian Countries. *International
Journal of Advanced Computer Science and Applications*, *9*(6).
https://doi.org/10.14569/IJACSA.2018.090663

Abdelaziz, A., Santos, V., & Dias, M. S. (2021). Machine Learning Techniques in the Energy
Consumption of Buildings: A Systematic Literature Review Using Text Mining and
Bibliometric Analysis. *Energies*, *14*(22), Article 22.
https://doi.org/10.3390/en14227810

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2018). Indicadores de
calidad de datos abiertos: El caso del portal de datos abiertos de Barcelona.
Profesional de la información, *27*(2), Article 2.
https://doi.org/10.3145/epi.2018.mar.16

Ahlin, K., & Crusoe, J. (2022). Why Should You Believe in Open Data? – A Document Study
Examining Persuasion Rhetoric of OGD Benefits. In M. Janssen, C. Csáki, I.
Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E.
Tambouris (Eds.), *Electronic Government* (pp. 274–287). Springer International
Publishing. https://doi.org/10.1007/978-3-031-15086-9_18

Alsmadi, A. A., Alrawashdeh, N., Al-Dweik, A. F., & Al-Assaf, M. (2022). Cryptocurrencies:
A bibliometric analysis. *International Journal of Data and Network Science*, *6*(3),
619–628. https://doi.org/10.5267/j.ijdns.2022.4.011

Ansari, B., Barati, M., & Martin, E. G. (2022). Enhancing the usability and usefulness of open
government data: A comprehensive review of the state of open government data
visualization research. *Government Information Quarterly*, *39*(1), 101657.
https://doi.org/10.1016/j.giq.2021.101657

Araluze, G. K. B. de, & Plaza, N. C. (2022). Open banking: A bibliometric analysis-driven definition. *PLOS ONE*, *17*(10), e0275496. https://doi.org/10.1371/journal.pone.0275496

Attard, J., Orlandi, F., & Auer, S. (2016). Data Driven Governments: Creating Value Through Open Government Data. In A. Hameurlain, J. Küng, R. Wagner, A. Anjomshoaa, P. C. K. Hung, D. Kalisch, & S. Sobolevsky (Eds.), *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII* (Vol. 9860, pp. 84–110). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-53416-8_6

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Ciancarini, P., Poggi, F., & Russo, D. (2016). Big Data Quality: A Roadmap for Open Data. *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, 210–215. https://doi.org/10.1109/BigDataService.2016.37

Cinnamon, J. (2020). Data inequalities and why they matter for development. *Information Technology for Development*, *26*(2), 214–233. https://doi.org/10.1080/02681102.2019.1650244

da Cruz, N. F., Tavares, A. F., Marques, R. C., Jorge, S., & de Sousa, L. (2016). Measuring Local Government Transparency. *Public Management Review*, *18*(6), 866–893. https://doi.org/10.1080/14719037.2015.1051572

de Castro Neto, M., Rego, J. S., Neves, F. T., & Cartaxo, T. M. (2017). Smart amp; open cities: Portuguese municipalities open data policies evaluation. *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/CISTI.2017.7975912

de Juana-Espinosa, S., & Luján-Mora, S. (2019). Open government data portals in the
European Union: Considerations, development, and expectations. *Technological
Forecasting and Social Change*, *149*, 119769.
https://doi.org/10.1016/j.techfore.2019.119769

de Juana-Espinosa, S., & Luján-Mora, S. (2020). Open government data portals in the
European Union: A dataset from 2015 to 2017. *Data in Brief*, *29*, 105156.
https://doi.org/10.1016/j.dib.2020.105156

*Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013
amending Directive 2003/98/EC on the re-use of public sector informationText with
EEA relevance*. (n.d.). 8.

*Gamification and e-learning for young learners: A systematic literature review, bibliometric
analysis, and future research agenda | Elsevier Enhanced Reader*. (n.d.).
https://doi.org/10.1016/j.techfore.2021.121445

Garcia Saez, C. (2022). Improving open data quality through citizen engagement and data
engineering. *Proceedings of the 18th International Symposium on Open
Collaboration*, 1–3. https://doi.org/10.1145/3555051.3555075

Gil-Garcia, J. R., Gasco-Hernandez, M., & Pardo, T. A. (2020). Beyond Transparency,
Participation, and Collaboration? A Reflection on the Dimensions of Open
Government. *Public Performance & Management Review*, *43*(3), 483–502.
https://doi.org/10.1080/15309576.2020.1734726

Habib, A., Prybutok, V. R., & Philpot, D. (2022). Assessing and building municipal open data
capability. *Information Systems and E-Business Management*, *20*(1), 1–25.
https://doi.org/10.1007/s10257-021-00539-y

Haini, S. I., Ab. Rahim, N. Z., Mohd. Zainuddin, N. M., & Ibrahim, R. (2020). Factors
Influencing the Adoption of Open Government Data in The Public Sector: A

Systematic Literature Review. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(2), 611. https://doi.org/10.18517/ijaseit.10.2.9488

Hassine, S. B., & Clément, D. (2020). Open Data Quality Dimensions and Metrics: State of the Art and Applied Use Cases. In W. Abramowicz & G. Klein (Eds.), *Business Information Systems Workshops* (pp. 311–323). Springer International Publishing. https://doi.org/10.1007/978-3-030-61146-0_25

Huber, F., Ponce, A., Rentocchini, F., & Wainwright, T. (2022). The wealth of (Open Data) nations? *Open government data, country-level institutions and entrepreneurial activity*. *Industry and Innovation*, *29*(8), 992–1023. https://doi.org/10.1080/13662716.2022.2109455

Jamieson, D., Wilson, R., & Martin, M. (2019). The (im)possibilities of open data? *Public Money & Management*, *39*(5), 364–368. https://doi.org/10.1080/09540962.2019.1611240

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Kawashita, I., Baptista, A. A., & Soares, D. (2022). *Open Government Data Use by the Public Sector—An Overview of its Benefits, Barriers, Drivers, and Enablers*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.315

Krasikov, P., Obrecht, T., Legner, C., & Eurich, M. (2021). Open Data in the Enterprise Context: Assessing Open Corporate Data's Readiness for Use. In S. Hammoudi, C. Quix, & J. Bernardino (Eds.), *Data Management Technologies and Applications* (pp.

80–100). Springer International Publishing. https://doi.org/10.1007/978-3-030-83014-4_4

Lnenicka, M., Luterek, M., & Nikiforova, A. (2022). Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. *Telematics and Informatics*, *66*, 101745. https://doi.org/10.1016/j.tele.2021.101745

Luthfi, A., & Janssen, M. (2022). Toward a Reference Architecture for User-Oriented Open Government Data Portals. In B. Shishkov (Ed.), *Business Modeling and Software Design* (pp. 259–267). Springer International Publishing. https://doi.org/10.1007/978-3-031-11510-3_17

Máchová, R., & Lnénicka, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1), 21–41. https://doi.org/10.4067/S0718-18762017000100003

Methodology—Global Open Data Index. (n.d.). Retrieved 21 November 2022, from http://index.okfn.org/methodology/

Mustapa, M. N., Hamid, S., & Nasaruddin, F. H. M. (2022). Factors influencing open government data post-adoption in the public sector: The perspective of data providers. *PLOS ONE*, *17*(11), e0276860. https://doi.org/10.1371/journal.pone.0276860

Mustapha, I., Khan, N., Qureshi, M. I., Harasis, A. A., & Van, N. T. (2021). Impact of Industry 4.0 on Healthcare: A Systematic Literature Review (SLR) from the Last Decade. *International Journal of Interactive Mobile Technologies (IJIM)*, *15*(18), Article 18. https://doi.org/10.3991/ijim.v15i18.25531

Nikiforova, A. (2020). Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment. *Baltic Journal of Modern Computing*, *8*(3). https://doi.org/10.22364/bjmc.2020.8.3.02

Nogueras-Iso, J., Lacasta, J., Ureña-Cámara, M. A., & Ariza-López, F. J. (2021). Quality of

    Metadata in Open Data Portals. *IEEE Access*, *9*, 60364–60382.

    https://doi.org/10.1109/ACCESS.2021.3073455

Open Data Charter. (n.d.). Retrieved 14 November 2022, from

    https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D.,

    Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

    Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson,

    E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated

    guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89.

    https://doi.org/10.1186/s13643-021-01626-4

Precinotto, A., Aquino, A., & Dias, L. (2022). Paradoxos na divulgação de salários de

    servidores em municípios. *Revista de Administração Pública*, *56*, 191–207.

    https://doi.org/10.1590/0034-761220200881

Publications Office of the European Union. (2020). *Open data maturity report 2019*.

    Publications Office. https://data.europa.eu/doi/10.2830/073835

Publications Office of the European Union. (2022). *Open data maturity report 2021*.

    Publications Office. https://data.europa.eu/doi/10.2830/394148

Raca, V., Velinov, G., Dzalev, S., & Kon-Popovska, M. (2022). A Framework for Evaluation

    and Improvement of Open Government Data Quality: Application to the Western

    Balkans National Open Data Portals. *SAGE Open*, *12*(2), 21582440221104812.

    https://doi.org/10.1177/21582440221104813

Rani, P., Yadav, A., Kumar, D., Pandey, J., Gull, M., Ansari, M. A., Ghosh, S., & Sahni, B.

    (2022). A Bibliometric Analysis of Literature on Covid-19 And Mental Health.

National Journal of Community Medicine, *13*(09), Article 09.

https://doi.org/10.55489/njcm.130920222131

Rashmi, K., & Kataria, A. (2021). Work–life balance: A systematic literature review and

bibliometric analysis. *International Journal of Sociology and Social Policy*, *ahead-of-*

print. https://doi.org/10.1108/IJSSP-06-2021-0145

Reinsel, D., Gantz, J., & Rydning, J. (2020). *The Digitization of the World*.

Ruijer, E., Détienne, F., Baker, M., Groff, J., & Meijer, A. J. (2020). The Politics of Open

Government Data: Understanding Organizational Responses to Pressure for More

Transparency. *The American Review of Public Administration*, *50*(3), 260–274.

https://doi.org/10.1177/0275074019888065

Ruijer, E. H. J. M., & Martinius, E. (2017). Researching the democratic impact of open

government data: A systematic literature review. *Information Polity*, *22*(4), 233–250.

https://doi.org/10.3233/IP-170413

Ruijer, E., & Meijer, A. (2020). Open Government Data as an Innovation Process: Lessons

from a Living Lab Experiment. *Public Performance & Management Review*, *43*(3),

613–635. https://doi.org/10.1080/15309576.2019.1568884

Sampaio, I. G. B., Andrade, E. de O., Bernardini, F., & Viterbo, J. (2022). Assessing

the Quality of Covid-19 Open Data Portals. In M. Janssen, C. Csáki, I. Lindgren, E.

Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E. Tambouris (Eds.),

Electronic Government (pp. 212–227). Springer International Publishing.

https://doi.org/10.1007/978-3-031-15086-9_14

Schmidthuber, L., & Hilgers, D. (2021). Trajectories of local open government: An empirical

investigation of managerial and political perceptions. *International Public*

Management Journal, *24*(4), 537–561.

https://doi.org/10.1080/10967494.2020.1853291

Simonofski, A., Zuiderwijk, A., Clarinval, A., & Hammedi, W. (2022). Tailoring open

  government data portals for lay citizens: A gamification theory approach.

  International Journal of Information Management, *65*, 102511.

  https://doi.org/10.1016/j.ijinfomgt.2022.102511

Šlibar, B., & Mu, E. (2022). OGD metadata country portal publishing guidelines compliance:

  A multi-case study search for completeness and consistency. *Government Information*

  Quarterly, *39*(4), 101756. https://doi.org/10.1016/j.giq.2022.101756

Šlibar, B., Oreški, D., & Begičević Ređep, N. (2021). Importance of the Open Data

  Assessment: An Insight Into the (Meta) Data Quality Dimensions. *SAGE Open*, *11*(2),

  21582440211023176. https://doi.org/10.1177/21582440211023178

Song, C., & Kim, H. (2022). Improvements of public data policy through data portal analysis

  of local governments. *Journal of Digital Contents Society*, *23*(4), 697–705.

  https://doi.org/10.9728/dcs.2022.23.4.697

Tejedo-Romero, F., & Araujo, J. F. F. E. (2021). Fatores influenciadores da divulgação de

  informação e do acesso à informação dos municípios. *Transinformação*, *33*.

  https://doi.org/10.1590/2318-0889202133e200038

Toots, M., McBride, K., Kalvet, T., & Krimmer, R. (2017). Open Data as Enabler of Public

  Service Co-creation: Exploring the Drivers and Barriers. *2017 Conference for E-*

  Democracy and Open Government (CeDEM), 102–112.

  https://doi.org/10.1109/CeDEM.2017.12

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016).

  Open data quality measurement framework: Definition and application to Open

  Government Data. *Government Information Quarterly*, *33*(2), 325–337.

  https://doi.org/10.1016/j.giq.2016.02.001

Wirtz, B. W., Weyerer, J. C., Becker, M., & Müller, W. M. (2022). Open government data: A

    systematic literature review of empirical research. *Electronic Markets*.

    https://doi.org/10.1007/s12525-022-00582-8

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2015). Quality

    assessment for Linked Data: A Survey: A systematic literature review and conceptual

    framework. *Semantic Web*, *7*(1), 63–93. https://doi.org/10.3233/SW-150175

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A

    framework for comparison. *Government Information Quarterly*, *31*(1), 17–29.

    https://doi.org/10.1016/j.giq.2013.04.003

Zuiderwijk, A., Pirannejad, A., & Susha, I. (2021). Comparing open data benchmarks: Which

    metrics and methodologies determine countries' positions in the ranking lists?

    Telematics and Informatics, *62*, 101634. https://doi.org/10.1016/j.tele.2021.101634

(N.d.).

Instituto Nacional de Estatística - Estudo sobre o Poder de Compra Concelhio : 2019 Lisboa : INE,

    2021. Disponível na www: <url:https://www.ine.pt/xurl/pub/6358988>. ISSN 0872-5977.

    ISBN 978-989-25-0578-7

ANNEXES

OPEN DATA MATURITY REPORT 2021: Recommendations for the Countries

Beginners: Think big, act small:

1. Develop a national strategy for open data and align it with broader strategies at national level (e.g. digital strategies, strategies for the modernization of public sector etc.

2. Rally support to the open data programme and political leadership from top level of government. Showcase international research around the value of open data, to emphasize economic benefits of data exploitation.

3. Set up a team at national level in charge of open data to ensure coordination of activities within the country and set up 'road-shows' to promote the team's scope and activities with the main public administrations. Include all levels of government in this process.

4. Organize a series of open data events at national level and focus on engaging both data publishers and re-users in your country. Prioritize the promotion of data publication best practices and reuse cases during such events.

5. Set up relevant communication channels and contact persons for data publication within public administrations (e.g. open data liaison officers). Maintain an active dialogue with the officers and enable regular exchange of knowledge amongst them, focusing on efficient online channels, in this time of pandemic (meetings, online forums etc.).

6. Identify the main data holders in the country and understand the main concerns and barriers to data publication. Take the first steps to overrun these barriers and unlock the publication of data.

7. Organize workshops and awareness-raising sessions with the main data holders. Use materials already developed in other countries and at European level for content and as source of inspiration.

8. Develop guidelines to enable publication of data, of its metadata and the take-up of suitable licensing conditions. If standard licenses are not suitable, as a last resort evaluate developing a custom national license. Learn from European best practices and reach out to colleagues in other countries when setting out to develop such guidelines. Raise awareness amongst main data publishers around the importance of metadata and promote the DCAT-AP standard, specifications, and existing guidelines developed at European level.

9. Make sure you run and maintain a modern portal that enables publication and discoverability of open data. Scout for European best practices and compare solutions to choose the most adequate to support your scope and mission. Set up dedicated news and blog sections to promote relevant developments and to showcase re-use. Ensure feedback channels are seamlessly integrated into the national portal. Be aware of users' rights and privacy as you perform web analytics, and choose your technology carefully, particularly following the invalidation of the EU-U.S. Privacy Shield.

10. Ensure that the national open data strategy guarantees scoping, management, and funding of the portal. Use action plans with actions and responsible entities or persons to ensure the strategy to be

carried out. Ensure that sufficient resources are allocated to open data awareness-raising activities with both publishers and potential re-users.

Table 0.1 - The method of scoring for the datasets in the GODI (Lnenicka et al., 2022)

Dimension	Criterion	OKF 2014	OKF 2015	OKF 2017
Technical Openness	Does the data exist?	5	5	Not Scored
	Is the data in digital form?	5	5	Not Scored
	Is the data available online?	5	5	15
	Is the data machine-readable?	15	15	20
	Available in bulk?	10	10	15
	Is the data provided on a timely and up to date basis?	10	10	15
	Total	**50**	**50**	**65**
Legal Openness	Openly licensed?	30	30	20
	Is the data available for free?	15	15	15
	Publicly available?	5	5	Not Scored
	Total	**50**	**50**	**35**

Table 0.2 - Overview of open data indices and rankings published by international organizations (Lnenicka et al., 2022)

Sub-Index	Component	2013	2014	2015	2016	2017
Readiness	Government policies	1/3	1/3	1/4	1/4	1/4
	Government action			1/4	1/4	1/4
	Entrepreneurs & business	1/3	1/3	1/4	1/4	1/4
	Citizens & civil society	1/3	1/3	1/4	1/4	1/4
	Weight of the sub-index	1/5	1/4	35%	35%	35%
Implementation	Accountability	1/3	1/3	1/3	1/3	1/3
	Innovation	1/3	1/3	1/3	1/3	1/3
	Social policy	1/3	1/3	1/3	1/3	1/3
	Weight of the sub-index	3/5	2/4	35%	35%	35%
Impacts	Political	1/3	1/3	1/3	1/3	1/3
	Economic	1/3	1/3	1/3	1/3	1/3
	Social	1/3	1/3	1/3	1/3	1/3
	Weight of the sub-index	1/5	1/4	30%	30%	30%

1. Open Data Policy

Focuses on the presence of specific policies and strategies to **foster open data at national level**. The dimension also analyses the existence of governance structures that allow the participation of private and third sector actors, as well **as implementation measures** that enable open data initiatives at national, regional, and local level. Furthermore, the dimension looks at **training schemes** that enhance the **data literacy skills of the civil servants working** with data and harvesting mechanisms that foster **the discoverability of all open data** available in the country.

2. Open Data Impact

Looks at the activities performed **to monitor and measure open data re-use** and the **impact derived** by such re-use. Beyond this first layer of "strategic awareness", the impact dimension focusses on four areas of sectoral impact: **political, social, environmental, and economic.** Within these areas, the questionnaire examines the extent to which monitoring is in place to document the re-use of open data published in these fields, the extent to which applications, products, and services have been developed to address challenges in these fields, as well as the extent to which **civil society initiatives exist** that are based on such open data and supported by government institutions. In addition, the dimension includes the efforts taken to commission and conduct studies **that measure the impact created through open data re-use** in each of the impact areas.

3. Open Data Portal

Focuses on **advanced portal functions** that enable both versed and less versed users to access open data via the national portal and features that enhance the interaction between publishers and re-users (via forum and discussion boards). Additionally, the dimension assesses the extent to which portal managers use web analytics tools to better **understand their users' needs and behaviour** and update the portals' features in line with the insights gained from these analyses. The dimension examines the **open data coverage** across different domains, as well as the approach and measures in place to ensure the portal's **sustainability**.

4. Open Data Quality

Focuses on the measures adopted by portal managers **to ensure the systematic harvesting of metadata** from sources across the country, as well as the currency of the available metadata and where possible the actual data, the monitoring of the **compliance with the DCAT-AP** metadata standard as well as the quality of deployment of the published data. The fourth dimension provides impulses for portal managers and policymakers to enable open data publication that is good quality all round: using open data formats, machine-readable, high-quality, and suitable to a linked data approach.

Figure 0.1 – Explanation of the four dimensions of assessment (Publications Office of the European Union., 2022)

Dimension	Indicator	2015	2016	2017	2018	2019	2020	2021
Policy	Open data policy	300	330	400	180	220	220	220
	Licensing norms	70	70	80	150			
	Governance of Open Data							220
	Coordination at national level	130	130	140	350	215	220	
	Open data implementation					210	210	210
	Weight				27%	25%	25%	25%
Impact	Political impact	120	120	120	130	130	130	120
	Strategic awareness				200	140	140	270
	Environmental impact				80	150	150	100
	Social impact	60	60	60	110	120	120	80
	Economic impact	120	120	120	130	110	110	80

		c1	c2	c3	c4	c5	c6	c7
	Weight				26%	25%	25%	25%
Portal	Usability of the portal	100	60	90	250	240	240	
	Re-usability of the portal	100	140	140				
	Portal Features							240
	Data Provision							100
	Spread of data across domains	50	50	50	160	110	110	
	Portal sustainability				120	150	150	150
	Portal usage	200	260	300	120	150	150	160
	Weight				26%	25%	25%	25%
Quality	Automation				100			
	Data and metadata currency				210	150	150	150
	DCAT-AP compliance				210	180	180	170
	Monitoring and measures					150	150	160
	Deployment quality and linked data					170	170	170
	Weight				21%	25%	25%	25%
SUM		1250	1340	1500	2500	2595	2600	2600

Table 0.3 - Scoring of dimensions and indicators in the ODMR (Lnenicka et al., 2022)

There were some other indices and rankings being used in the scientific literature that were not selected for the criteria and weights analysis:

- ODRA (Open Data Readiness Assessment) → Developed by the Open Government Data Working Group of the World Bank, this methodological tool aids in the planning of measures that the government authority may consider doing in order to build an Open Data program at various administrative levels. It offers eight dimensions, such as circumstances and questions to ask, to implement the Open Data program successfully. The primary types of datasets from the GODI from 2013 are reused by the ODRA. It also suggests considering other indexes, such as the ITU's ICT Development Index. However, even though this assessment tool may be helpful for nations in their open data initiatives, there is no output that would allow for benchmarking and ranking nations using this framework, as well as no other update of the methodology from 2015 (Lnenicka et al., 2022).

- PSI (Public Sector Information) Scoreboard → is a tool to assess the level of reuse of PSI and open data across the EU. Its objectives are based on EU regulations from 2003 and 2013, which offered a unified legal framework in this field. The most recent Scoreboard, from 2013, evaluated forms, costs, events, and activities. The ODMR has been the focus of evaluations of open data in the EU since 2015. (Lnenicka et al., 2022).

- WJP OGI (World Justice Project Open Government Index) → The WJP Rule of Law Index, which has been published annually since 2008, served as the conceptual foundation for the WJP OGI, which was only released once in 2015. The WJP OGI is focused on the experiences and perceptions of the public, and four factors, namely the right to information, civic engagement, publicized legislation and government data, and complaint channels. Based on responses to household surveys and in-country expert questionnaires, each is given a score between 0 and 1 (Lnenicka et al., 2022).

- Open Data 500 → Was introduced in 2014 under GovLab and focuses on 500 U.S. businesses who use OGD as a crucial resource for their operations. This program has made it possible to map and visualize the connections between businesses and publicly available statistics, as well as to analyse other industries. But as far as we are aware, this tool is no longer supported (Lnenicka et al., 2022).
- Open Data Impact → a repository that offers a thorough understanding of the numerous processes and elements that influence the demand, supply, release, usage, and impact of open data, as well as an assessment and supporting documentation of the possible positive social effects of open data. Additionally, they offer helpful advice to decision-makers, members of civil society, business owners, researchers, and anyone who wish to access or use open data, i.e., creating a manual (Lnenicka et al., 2022).
- Open Data Impact Map → is a public database created by the Open Data for Development Network that features businesses using the OGD from all over the world (OD4D). The creators of this database emphasize that neither do they offer a random or representative selection of use cases, nor do they seek to grade, assess, or quantify the economic or social value of open data. As data are gathered in three different ways—a web-based poll, a network of regional supporters representing 20 nations who provide examples and insights based on local expertise, and research—they cannot guarantee the correctness of any entry (Lnenicka et al., 2022).
- Open Data Census → is a platform run by Open Knowledge International that collaborates closely with the GODI to create what is known as a "Open Data Survey." It is used to assess how well various towns and localities are doing at sharing open data (Lnenicka et al., 2022).
- Open Data Monitor → is a framework that extracts datasets by gathering metadata from several open data sources. By spotting gaps that call for more open data, it may be able to evaluate and visualize metadata and so reveal the untapped potential of already-existing resources. However, only 2015's complete statistics are available, and neither the scores nor the fact that the framework has changed since then are mentioned. The Open Data Census and the Open Data Monitor are both categorized as data-oriented methodologies that concentrate on the publishing practices of the countries they investigate (Lnenicka et al., 2022).