



Enhancing Usability of Open Government Data Portals: Bridging the Gap with User-Centric Visualizations

Beatriz da Conceição Fernandes Thomaz

Thesis to obtain the Master of Science Degree in
Industrial Engineering and Management

Supervisor: Prof. Miguel Leitão Bignolas Mira da Silva

Examination Committee

Chairperson: Prof. Joana Serra da Luz Mendonça

Supervisor: Prof. Miguel Leitão Bignolas Mira da Silva

Member of the Committee: Prof. Cláudia Martins Antunes

July 2024

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

Acknowledgments

I would first like to thank my parents for affording me the incredible opportunity to pursue my studies at Técnico and for supporting me throughout this journey.

I would also like to thank my supervisor, Professor Miguel Mira da Silva, for guiding and motivating me throughout this master's thesis, having had a crucial role in keeping me on the right track.

Additionally, I would like to thank ANI – Agência Nacional de Inovação for providing me with the possibility to do an internship with them, which was crucial for the development of this thesis. A special thanks to Eng. Filipe Moreira and colleague João Ferreira, for welcoming me and helping me through the process of the creation of the Power BI visualizations.

Finally, I would like to thank my friends and family for standing by me during this journey.

Abstract

Open Government Data (OGD) has gained substantial interest in recent years due to its potential benefits in promoting transparency, accountability, and citizen engagement. However, various challenges and barriers hinder its effective implementation. It was found that one critical concern revolves around the usability of OGD portals, as governments often struggle to make data accessible and understandable to the public, especially to those with no experience with working with data. To address this problem, this Master thesis adopts a Design Science Research Methodology approach, with the goal of proposing a generic solution to enhance user engagement with data exploration using Business Intelligence (BI) tools. A demonstration was conducted to illustrate the proposed solution, involving the creation of visualizations in Power BI for three programs administered by the Portuguese governmental agency ANI – *Agência Nacional de Inovação*. Subsequently, an evaluation of these visualizations was undertaken through an interview process involving 10 participants. Notably, the participants were divided into three distinct groups: those with no experience in working with data, individuals with experience in data analysis but limited exposure to BI tools, and experts proficient in Power BI. The findings revealed a clear division in user interaction based on participants' experience levels. While participants succeeded in extracting the required information from the visualizations, challenges were observed regarding interactivity and intuitiveness, particularly for those with limited exposure to data analysis and BI tools.

Keywords: Open Data, Open Government Data, Open Data Portal, Innovation, Business Intelligence

Resumo

Open Government Data (OGD) tem despertado um interesse substancial nos últimos anos devido aos seus potenciais benefícios na promoção de transparência, responsabilidade e envolvimento do cidadão. No entanto, existem vários desafios e obstáculos que dificultam a sua implementação eficaz. Foi identificada uma preocupação relevante referente à usabilidade dos portais de OGD, devido às dificuldades que os governos frequentemente enfrentam em tornar os dados acessíveis e compreensíveis para o público, especialmente para aqueles sem experiência em trabalhar com dados. De modo a abordar este problema, esta tese de mestrado adota a abordagem Design Science Research Methodology, com o objetivo de propor uma solução genérica para melhorar o envolvimento dos utilizadores com a exploração de dados usando ferramentas de Business Intelligence (BI). Foi realizada uma demonstração para ilustrar a solução proposta, envolvendo a criação de visualizações em Power BI para três programas administrados pela agência governamental portuguesa ANI – Agência Nacional de Inovação. Posteriormente, uma avaliação dessas visualizações foi realizada através de um processo de entrevistas envolvendo 10 participantes. Os participantes foram divididos em três grupos distintos: aqueles sem experiência em trabalhar com dados, indivíduos com experiência em análise de dados mas com exposição limitada a ferramentas de BI, e especialistas proficientes em Power BI. Os resultados revelaram uma clara divisão na interação do utilizador com base nos níveis de experiência dos participantes. Embora os participantes tenham conseguido extrair as informações necessárias das visualizações, foram observados desafios relacionados à interatividade e intuições, especialmente para aqueles com exposição limitada à análise de dados e ferramentas de BI.

Palavras-Chave: Dados Abertos, Dados Governamentais Abertos, Portal de Dados Abertos, Inovação, Business Intelligence

Table of Contents

Acknowledgments	ii
Abstract	iii
Resumo	iv
List of Figures	vii
List of Tables	viii
Acronyms	ix
1. Introduction	1
1.1. Research Context	3
1.2. Research Methodology	4
1.3. Organization of the Document	4
2. Systematic Literature Review	5
2.1. Research Method	6
2.2. Planning the review	6
2.2.1. Research Questions	6
2.2.2. Search Process	6
2.2.3. Inclusion and Exclusion Criteria	7
2.3. Conducting the review	7
2.4. Reporting the review	9
2.4.1. RQ1 - What is Open Government Data?	9
2.4.2. RQ2 - What are the benefits of implementing OGD?.....	9
2.4.3. RQ3 - What are the challenges governments face in publishing OGD?	11
2.4.4. RQ4 - What are the barriers users face in using OGD portals?	12
2.4.5. RQ5 - What are the requirements for creating an OGD portals?.....	13
3. Research Problem	18
4. Research Proposal	20
4.1. Objectives.....	21
4.2. Potential benefits of BI tools.....	21
4.3. BI tool implementation steps.....	22
5. Demonstration	24
5.1. ANI - Agência Nacional de Inovação	25
5.1.1. SIFIDE	25
5.1.2. PT2020	26
5.1.3. HEurope	26
5.2. Tool – Power BI	26

5.3.	First Demonstration - SIFIDE	27
5.3.1.	Cover Page.....	27
5.3.2.	Summary Page.....	28
5.3.3.	Report Page	30
5.3.4.	Website Link Page	34
5.4.	Second Demonstration – PT2020.....	34
5.4.1.	Cover and Link Pages.....	35
5.4.2.	Summary Page.....	35
5.4.3.	Report Page	35
5.5.	Third Demonstration - HEurope.....	36
5.5.1.	Cover and Link Pages.....	37
5.5.2.	Summary Page.....	37
5.5.3.	Report Page	38
6.	Evaluation	43
6.1.	Description	44
6.2.	Sampling.....	44
6.3.	Data Collection.....	45
6.4.	Results	45
7.	Conclusion.....	49
7.1.	Approach and Results	50
7.2.	Limitations.....	50
7.3.	Future Work	51
	References	53
	Appendix A – Evaluation	56
	A.1 Interview Questions/Tasks.....	56
	A.2 Questionnaire	58

List of Figures

Figure 1 - Screening process	7
Figure 2 - SIFIDE Cover page.....	28
Figure 3 - SIFIDE Summary page	28
Figure 4 – Tooltip for Applications per year	29
Figure 5 – Tooltip for declared and cleared investment per year.....	29
Figure 6 – Tooltip for requested and assigned credit per year	29
Figure 7 - Filter pane of SIFIDE Summary page	30
Figure 8 – “Received Applications” section of SIFIDE Report page.....	30
Figure 9 – Tooltip for received applications per year	31
Figure 10 – Tooltip for declared investment and credit requested per year	31
Figure 11 – Tooltip for received applications per municipality	32
Figure 12 - Tooltip for received applications per district	32
Figure 13 - Tooltip for declared investment and credit requested per municipality	33
Figure 14 - Tooltip for declared investment and credit requested per district	33
Figure 15 - Filter pane of SIFIDE Report page.....	33
Figure 16 – “Approved Applications” section of SIFIDE Report page	34
Figure 17 - SIFIDE Link page	34
Figure 18 - PT2020 Summary page	35
Figure 19 – “Received Applications” section of PT2020 Report page	35
Figure 20 - Bar chart disaggregated into 17 support measures.....	36
Figure 21 - Bar chart aggregated into 3 categories	36
Figure 22 - HEurope Summary page.....	37
Figure 23 - Evolution of approved projects and obtained financing from 2014 until 2023.....	38
Figure 24 – “Analysis per Year” section of HEurope Report page	38
Figure 25 - Tooltip for Success rate of Portuguese proposals per year	39
Figure 26 - Tooltip for obtained financing per year.....	39
Figure 27 - Tooltip for Participations.....	40
Figure 28 - Filter pane of “Analysis per Year” section of HEurope Report page.....	40
Figure 29 – “Analysis per Theme” section of HEurope Report page.....	41
Figure 30 - Tooltip for Success rate of Portuguese proposals per theme	41
Figure 31 - Tooltip for Participations per theme	42
Figure 32 - Filter pane of “Analysis per Theme” section of HEurope Report page	42

List of Tables

Table 1 - Search String 7

Table 2 - Classification definitions 8

Table 3 - Final set of papers 8

Table 4 – Benefits of implementing Open Government Data..... 10

Table 5 - Challenges governments face in publishing Open Government Data 11

Table 6 - Barriers users face in using Open Government Data Portals 12

Table 7 - Five Star Scheme for Linked Open Data..... 15

Table 8 - Requirements for creating an Open Government Data Portal 17

Table A.1. 1 - Tasks related to the SIFIDE program 57

Table A.1. 2 – Tasks realtes to the PT2020 program 57

Table A.1. 3 – Tasks related to the HEurope program..... 58

Acronyms

ANI	Agência Nacional de Inovação
BI	Business Intelligence
DSRM	Design Science Research Methodology
EC	Exclusion Criteria
EDA	Exploratory Data Analysis
ETL	Extract, Transform and Load
FOIA	Freedom of Information Act
IC	Inclusion Criteria
ICTs	Information and Communication Technologies
IRC	Corporate Income Tax
IS	Information Systems
IT	Information Technology
OGD	Open Government Data
RQ	Research Question
SLR	Systematic Literature Review

1. Introduction

Contents

- 1.1. Research Context 3
- 1.2. Research Methodology 4
- 1.3. Organization of the Document 4

In the dynamic digital era of the 21st century, the sharing of information has become a crucial driver for progress, shaping innovation, decision-making, and the evolution of society. Governments have been recognizing the transformative potential of making their data accessible, turning their attention to Open Government Data (OGD) initiatives. These initiatives represent a commitment to transparency, empowering citizens with access to government data and fostering collaboration, innovation, and public accountability.

Publishing data on online open data portals has been the most popular and effective way of sharing OGD enabling easy access to government data for the public (Klievink et al., 2017; Simonofski et al., 2022; Zuiderwijk et al., 2013). However, despite the increase in OGD initiatives, it is still noticed a lack of progress regarding the quality of OGD portals that is preventing their usability, therefore, hindering the achievement of full transparency. In the field of human-computer interaction (HCI), usability relates to the effectiveness and satisfaction derived from the interaction between a user and a computer system through its interface (Chou and Hsiao, 2007). Therefore, OGD portals should be created with a strong emphasis on user-friendliness, intuitiveness and ease of use, to ensure that governments make their data usable for the benefit of all citizens (Wang et al., 2021).

The main issue studied in this thesis is the struggle government agencies have been facing to effectively develop these portals by simply disclosing the data without focusing on its quality, accessibility and understandability. This becomes an obstacle for users, especially for the regular citizen who may lack the technical skills or knowledge to engage with raw data.

As exploring the evolution of data, OGD efforts, and the challenges of digital governance, this thesis serves as a guide, leading towards a future where OGD not only shares data but empowers citizens by providing clear and intelligible data. This thesis promotes a change in how OGD initiatives are implemented, emphasizing that making data accessible and easy to understand should be the top priority.

In the context of websites, usability pertains to the qualitative assessment of user-friendliness and ease of navigation (Lee and Kozar, 2012). Given that OGD portals fall within the category of websites, this study adopts a qualitative approach to assess usability, concentrating on interaction design to ensure a seamless and satisfactory interaction between users and OGD portals.

1.1. Research Context

Data, in its essence, represents information in its rawest form. It includes a wide range of information such as facts, statistics, measurements, and observations collected from various sources, including sensors, devices, human input, and organizational records. What makes data particularly valuable is its potential to be transformed into knowledge and insights, enabling us to better understand the world around us.

The importance of data is multifaceted. In the business world, data is crucial for decision-making processes, helping organizations in identifying trends, optimizing operations, and gaining a competitive advantage. In healthcare, data assists in diagnosing diseases, monitoring public health trends, and improving patient care. In academia, data drives research, enabling scientists to make groundbreaking discoveries and advance human knowledge. Data also plays a crucial role in governance, supporting policy formulation, delivering public services, and ensuring transparency and accountability in government activities (Amalia and Susanto, 2019).

The evolution of data reflects the progress of human society. Over the centuries, data has transitioned from handwritten records and printed documents to digital formats. This shift has been accelerated by the widespread use of technology, resulting in a substantial increase in the amount, speed, and variety of data generated worldwide. We now live in an era of “big data”, where massive datasets are generated and processed on an unprecedented level. Governments around the world collect and maintain large amounts of data, but until recently, data could only be accessed through statistical reports or Freedom of Information Act (FOIA) requests (Noveck, 2016; Schrock, 2016).

In today’s digital age, the rapid growth of the Internet and Information and Communication Technologies (ICTs) has been improving the sharing and delivery of data (Andersen and Henriksen, 2006; Layne and Lee, 2001; Mellouli et al., 2014). Thus, governments have been motivated to make their data publicly available in order to become more transparent about their operations while also enabling citizens to hold their elected officials and public agencies accountable (Linders and Wilson, 2011). Therefore, this growing interest in transparency, openness, and accountability of government agencies has been stimulating an increase in the number of OGD initiatives around the world (Máchová et al., 2018).

The shift of U.S. federal policy towards e-government as “open government”, emphasized in President Barack Obama’s Memorandum on Transparency and Open Government in 2009, played a crucial role in the advancement of OGD. Under Obama’s administration, many OGD initiatives and portals were established in the United States of America, such as the well-known central site, Data.gov. Obama’s commitment to OGD set an example for other governments around the world to follow, advancing the promotion of transparency in government. The three fundamental principles of transparency, participation and collaboration defined in Obama’s memorandum are considered the key foundations of OGD and continue to guide the development of open data initiatives worldwide (Ganapati and Reddick, 2012; Jaeger and Bertot, 2010; McDermott, 2010).

As data has evolved and grown, it has opened up new possibilities for what it is possible to learn and discover from data, but it also has brought new challenges. The amount of data generated requires advanced tools and techniques for storage, management, analysis, and interpretation. Additionally, there is a growing focus on data privacy and security, requiring strong measures to protect sensitive information.

1.2. Research Methodology

This work follows the Design Science Research Methodology (DSRM) due to its adaptability and iterative approach, making it ideal for the dynamic nature of this thesis' research problem. This methodology offers a framework that allows the research to evolve.

Moreover, DSRM aligns well with the research context, with the problem belonging in the field of Information Systems (IS) and Information Technology (IT). DSRM's focus on creating and evaluating IT artifacts makes it a perfect fit for this project (Peppers et al., 2007).

DSRM consists of six distinct phases, each with its role in guiding the research, ensuring that the problem is tackled effectively and systematically (Peppers et al., 2007):

1. **Problem Identification and Motivation:** In this phase, the problem and the value of the proposed solution are defined.
2. **Definition of Objectives for a Solution:** Clear objectives are set for the solution, based on an analysis of related work.
3. **Design and Development:** This phase involves transforming concepts into reality, detailing the functionality of the solution.
4. **Demonstration:** It is shown how the solution proposed can address the identified problem.
5. **Evaluation:** In this phase, the effectiveness of the solution proposed in addressing the identified problem is measured.
6. **Communication:** The final phase involves sharing the research findings, emphasizing the problem's significance, presenting the proposed solution, introducing the artifacts, and highlighting their utility, novelty, and the rigor of our research design.

1.3. Organization of the Document

This Master Thesis is structured as follows: Chapter 2 presents some related work on the subject. Chapter 3 identifies the research problem. In Chapter 4 the research proposal is described and in Chapter 5 it is demonstrated through a real case. In Chapter 6 the work is evaluated through a process of interviews and in Chapter 7, the conclusions are shared, and future work is outlined.

2. Systematic Literature Review

Contents

- 2.1. Research Method..... 6
- 2.2. Planning the review..... 6
- 2.3. Conducting the review..... 7
- 2.4. Reporting the review 9

In this Chapter, a systematic literature review is undertaken to critically examine and synthesize the current state of knowledge within the field of study. The purpose of this review is to identify gaps, patterns, and trends in the existing body of literature. With this review it is possible to understand the main issues faced in the implementation of OGD initiatives and provide a solid foundation for the work that is being studied in this thesis.

This review is structured as follows: Section 2.1 describes the research methodology. Section 2.2 identifies the research questions and describes the search process including the inclusion and exclusion criteria. In Section 2.3 the screening process is described. In Section 2.4 the research questions are answered, and the research results are reported.

2.1. Research Method

According to (Kitchenham, 2004), *“A systematic literature review is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest. Individual studies contributing to a systematic review are called primary studies; a systematic review is a form a secondary study.”* For this study, a systematic literature review (SLR) was conducted following the guidelines suggested by (Kitchenham, 2004). Therefore, the process was divided into three stages: planning the review, conducting the review, and reporting the review.

2.2. Planning the review

In this stage, the search strings, search sources and inclusion/exclusion criteria were established.

2.2.1. Research Questions

The purpose of this study is to investigate the benefits and challenges that arise from sharing government data with the public, as well as to outline the requirements for creating an open data portal that maximizes these benefits. Therefore, this research plans to answer the following research questions:

- RQ1: What is OGD?
- RQ2: What are the benefits of implementing OGD?
- RQ3: What are the challenges governments face in publishing OGD?
- RQ4: What are the barriers users face in using OGD portals?
- RQ5: What are the requirements for creating an OGD portal?

2.2.2. Search Process

The search engine EBSCO was used, and specific search strings were entered to obtain a comprehensive set of publications for this research. The concept of open data was the main topic with a focus on government data, therefore, the strings “open data” and “government” were used. The string “portal” was also added since the evaluation and analysis of OGD portals is essential for this study. The final search expression is presented in Table 1.

Table 1 - Search String

Search String
('open data') AND (government) AND (portal)

2.2.3. Inclusion and Exclusion Criteria

Inclusion criteria (IC) and exclusion criteria (EC) were applied to filter the search in order to obtain the most relevant publications for this research.

- EC1: Articles not written in the English language.
- EC2: Publications not from scientific journals or conferences.
- EC3: Articles lacking full text availability.
- EC4: Articles lacking peer review.
- EC5: Duplicated articles.
- IC1: Articles focused on benefits of OGD, challenges of OGD, best practices, frameworks, OGD Portals case studies and visualization tools.

2.3. Conducting the review

The predefined search string (Table 1) was entered into the search engine EBSCO and the studies were selected according to the exclusion and inclusion criteria mentioned in the previous stage. A total of 250 studies were returned by EBSCO and after automatically detecting and removing duplicates, 129 papers remained. The 129 papers were imported to Zotero and another set of duplicates in addition to articles written in other languages besides English were removed. Then, the screening phase was initiated and the titles and abstracts of the remaining 118 articles were read. From the first selection phase, 37 articles proceeded to the next phase while the other 81 articles were discarded. Then the introductions of the 37 articles were read and 22 papers were selected to be fully read. The 22 papers accepted were analyzed with the aim to answer the research questions previously defined. The process is summarized in Figure 1 **Erro! A origem da referência não foi encontrada.**

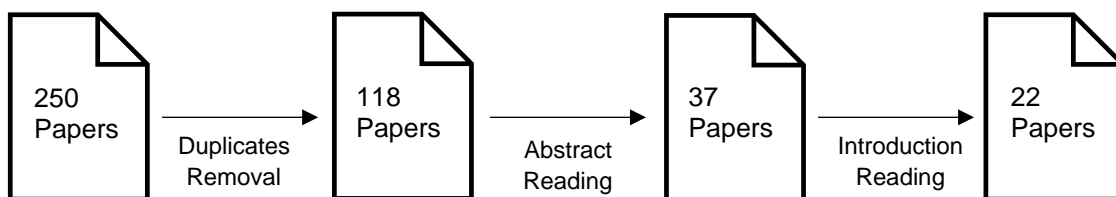


Figure 1 - Screening process

The 22 papers were classified based on their contributions. Table 2 shows the categories created to classify the articles and Table 3 presents the complete list of articles along with their corresponding journals and classifications.

Table 2 - Classification definitions

<i>Type</i>	<i>Description</i>
Research Article	The article presents original research on specific aspects of OGD.
Review Article	The article summarizes and critically analyzes existing literature on OGD.
Case Study	The article focuses on specific OGD initiatives or projects, providing detailed descriptions of the portal's development process.
Comparative Study	The article compares multiple OGD portals or examine the differences in OGD portal implementations.
Usability Study	These articles investigate the usability, accessibility, and user experience of OGD portals.

Table 3 - Final set of papers

<i>Paper</i>	<i>Journal</i>	<i>Type</i>
(Ansari et al., 2022)	Government Information Quarterly	Review Article
(Attard et al., 2015)	Government Information Quarterly	Review Article
(Chokki et al., 2022)	Digital Government Research and Practice	Research Article
(Chua et al., 2020)	ACM International Conference Proceeding Series	Usability Study
(Krismawati and Hidayanto, 2021)	International Conference on Advanced Computer Science and Information Systems	Usability Study
(Lnenicka and Nikiforova, 2021)	Telematics and Informatics	Research Article
(Lněnička et al., 2021)	Online Information Review	Research Article
(Lourenço, 2015)	Government Information Quarterly	Comparative Study
(Máchová et al., 2018)	Aslib Journal of Information Management	Usability Study
(Moradi et al., 2022)	Complexity	Research Article
(Mutambik et al., 2021)	Sustainability (Switzerland)	Comparative Study
(Nikiforova and McBride, 2021)	Telematics and Informatics	Comparative Study
(Paige and Freund, 2019)	JeDEM	Comparative Study
(Park and Gil-Garcia, 2022)	Government Information Quarterly	Case Study
(Sabri et al., 2019)	International Journal of Innovative Technology and Exploring Engineering	Comparative Study
(Safarov et al., 2017)	Information Policy	Review Article
(Saxena and Alexopoulos, 2022)	The Making of Contemporary Maldives: Isolation, Dictatorship and Democracy	Review Article
(Schauppenlehner and Muhar, 2018)	Sustainability (Switzerland)	Case Study
(Simonofski et al., 2022)	International Journal of Information Management	Research Article
(Thorsby et al., 2017)	Government Information Quarterly	Comparative Study
(Wang et al., 2021)	Library Hi Tech	Usability Study
(Zhu and Freeman, 2019)	Journal of the Association for Information Science and Technology	Usability Study

2.4. Reporting the review

The RQs focus on OGD and its benefits, challenges to its implementation, barriers to its usage and the requirements an OGD portal should meet. There was a need to separate barriers from challenges to distinguish between the challenges faced by governments in disclosing the data and the barriers faced by users in accessing and using the data. The following sections report the results for each RQ previously defined.

2.4.1. RQ1 - What is Open Government Data?

The concept of Open Data refers to the practice of sharing data to be freely reused, modified, and distributed by the public. Open Government Data (OGD) is a subset of Open Data referring to the sharing of data made by the government. Governments or government agencies go towards the OGD movement when they make their information freely available to be used, reused, and redistributed by the public without any restrictions or limitations (Krismawati and Hidayanto, 2021).

Open Data Portals are used to make the connection between data publishers and data users. OGD portals are online platforms that serve as central repositories where users can easily find and use a wide range of data that is made available by government agencies and other organizations. Open data portals contain metadata about datasets and enable access to, download, and upload of datasets. They can also provide the tools and resources for analyzing and visualizing data, helping users to better understand and interpret the data (Attard et al., 2015). Additionally, according to (Lnenicka and Nikiforova, 2021), opening and publishing this data in open data portals with open licenses saves costs because it is cheaper than converting them into reports and applications.

2.4.2. RQ2 - What are the benefits of implementing OGD?

There are several benefits that arise from the adoption of OGD. This section summarizes the benefits found in the literature. The list of benefits can be found in Table 4.

Table 4 – Benefits of implementing Open Government Data

<i>Benefits</i>	<i>Papers</i>	<i>#Papers</i>
Transparency	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Chua et al., 2020), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Saxena and Alexopoulos, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Krismawati and Hidayanto, 2021), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	22
Accountability	(Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Saxena and Alexopoulos, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Krismawati and Hidayanto, 2021), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	18
Increased Civic Participation and collaboration	(Moradi et al., 2022), (Attard et al., 2015), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Lněnička et al., 2021), (Chua et al., 2020), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Saxena and Alexopoulos, 2022), (Krismawati and Hidayanto, 2021), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	17
Innovation	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Ansari et al., 2022), (Chua et al., 2020), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Saxena and Alexopoulos, 2022), (Krismawati and Hidayanto, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	15
Economic Development	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Saxena and Alexopoulos, 2022), (Krismawati and Hidayanto, 2021), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Safarov et al., 2017).	14
Better Decision-Making	(Moradi et al., 2022), (Attard et al., 2015), (Chokki et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Krismawati and Hidayanto, 2021), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	12
Public Trust	(Moradi et al., 2022), (Lourenço, 2015), (Ansari et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Krismawati and Hidayanto, 2021), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Safarov et al., 2017), (Attard et al., 2015)	10
Anti-Corruption	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Chokki et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Saxena and Alexopoulos, 2022), (Lnenicka and Nikiforova, 2021), (Safarov et al., 2017).	9
Improved Public Services	(Paige and Freund, 2019), (Attard et al., 2015), (Lourenço, 2015), (Lněnička et al., 2021), (Sabri et al., 2019), (Lnenicka and Nikiforova, 2021), (Safarov et al., 2017)	7
Increased government efficiency	(Paige and Freund, 2019), (Lněnička et al., 2021), (Sabri et al., 2019), (Saxena and Alexopoulos, 2022), (Thorsby et al., 2017), (Safarov et al., 2017).	6

The achievement of transparency is the main outcome from the implementation of OGD initiatives (Lnenicka and Nikiforova, 2021). Increasing transparency in governance by making data openly available enables citizens and other stakeholders to access and analyze information about government activities, policies, and decision-making processes. It gives them a better understanding of how government works and how public resources are being used, leading to greater accountability and trust between citizens and their government (Lněnička et al., 2021).

Achieving governance transparency can also help uncover instances of corruption or other unethical behaviors. According to (Attard et al., 2015), corruption was one of the main reasons for the emergence of OGD initiatives and literature identifies OGD as one of the main forces for anti-corruption movements (Safarov et al., 2017).

OGD can also be used to promote economic growth and innovation. Making data available to entrepreneurs, researchers, businesses and other stakeholders, enables them to utilize and analyze the data creatively, stimulating entrepreneurship and innovation of their products and services, having a positive impact on the economy of a country (Saxena and Alexopoulos, 2022). Citizens can also use the available data to monitor public services and suggest improvements, leading to the identification of areas of need and the development of more effective policies and programs (Safarov et al., 2017).

Another observed benefit is the promotion of citizen participation and collaboration by encouraging citizens to participate in the decision-making processes concerning the development of their society (Chokki et al., 2022). It also creates the opportunity for people with different backgrounds and expertise to share knowledge and work together (Attard et al., 2015).

OGD can also be used by government officials to make more informed decisions and increase efficiency by identifying areas for improvement in government operations and efficiently allocating resources (Lněnička et al., 2021). Providing access to government information also enables citizens and other stakeholders to inform themselves, supporting their own decision-making processes (Schauppenlehner and Muhar, 2018).

2.4.3. RQ3 - What are the challenges governments face in publishing OGD?

The implementation of OGD poses some challenges for governments. The most significant challenges will be addressed in this sub-section and are listed in Table 5.

Table 5 - Challenges governments face in publishing Open Government Data

<i>Challenges</i>	<i>Papers</i>	<i>#Papers</i>
Technical infrastructure and resources	(Attard et al., 2015), (Safarov et al., 2017), (Máchová et al., 2018), (Paige and Freund, 2019), (Lnenicka and Nikiforova, 2021), (Lněnička et al., 2021), (Park and Gil-Garcia, 2022)	7
Legal and regulatory issues	(Attard et al., 2015), (Safarov et al., 2017), (Paige and Freund, 2019), (Lnenicka and Nikiforova, 2021), (Saxena and Alexopoulos, 2022), (Park and Gil-Garcia, 2022)	6
Privacy and security	(Attard et al., 2015), (Safarov et al., 2017), (Paige and Freund, 2019)	3
Data quality	(Attard et al., 2015), (Lnenicka and Nikiforova, 2021)	2

Ensuring the quality of the data published can be challenging for governments. The literature presents many definitions for data quality. Accuracy is the most mentioned aspect regarding quality of data, and it can be divided into categories such as completeness, consistency and timeliness. (Lnenicka and Nikiforova, 2021) suggests that many current OGD portals meet most of the requirements related to data quality but identifies the timeliness of the data and frequency of updates as an unovercome challenge. Ensuring the consistency of the data can also be a challenge for governments due to the absence of standardized data formats, definitions, or schemes that can be inconsistent (Attard et al., 2015).

Another challenge faced by governments when sharing their data is the concern for privacy and security. Government data is sensitive so there is a need to balance transparency with the need to protect it. Appropriate security and privacy protocols are necessary to safeguard against unnecessary access or misuse (Attard et al., 2015).

Additionally, governments need to have the technical infrastructure in place to publish OGD. This includes investing in the necessary software for data analytics and discovery and web-based platforms (Safarov et al., 2017). Investing in technical expertise and other resources is also necessary which can discourage governments from participating in OGD initiatives (Attard et al., 2015).

Another important challenge addressed in the literature is related to legal and regulatory issues. Governments need to comply with laws and regulations related to data privacy, intellectual property, and other areas when publishing OGD. Difficulties in navigating these aspects can discourage governments from implementing OGD (Paige and Freund, 2019).

2.4.4. RQ4 - What are the barriers users face in using OGD portals?

Although governments face some challenges in implementing OGD, the key problem of OGD identified by the literature is not the disclosure of the data itself but its usage, more specifically, the lack of OGD use (Safarov et al., 2017). Despite the increasing number of OGD initiatives and portals, users still face many barriers to using these portals preventing the achievement of the potential benefits of OGD. Table 6 presents the full list of barriers identified.

Table 6 - Barriers users face in using Open Government Data Portals

<i>Barriers</i>	<i>Papers</i>	<i>#Papers</i>
Data quality	(Attard et al., 2015), (Safarov et al., 2017), (Sabri et al., 2019), (Nikiforova and McBride, 2021), (Lnenicka and Nikiforova, 2021), (Krismawati and Hidayanto, 2021), (Lněnička et al., 2021), (Moradi et al., 2022)	8
Technical skills and expertise/ Data understandability	(Safarov et al., 2017), (Paige and Freund, 2019), (Zhu and Freeman, 2019), (Lnenicka and Nikiforova, 2021), (Mutambik et al., 2021), (Simonofski et al., 2022), (Saxena and Alexopoulos, 2022), (Park and Gil-Garcia, 2022)	8
Data discoverability	(Attard et al., 2015), (Lourenço, 2015), (Schauppenlehner and Muhar, 2018), (Máchová et al., 2018), (Lnenicka and Nikiforova, 2021), (Lněnička et al., 2021)	6
Awareness	(Zhu and Freeman, 2019), (Wang et al., 2021), (Saxena and Alexopoulos, 2022)	3

Firstly, a barrier to the usage of OGD is the lack of awareness by citizens. Many potential users may not be aware of the availability or usefulness of OGD and the benefits it can bring (Zhu and Freeman, 2019).

The poor quality or inconsistency is one of the major barriers that impacts the usage of OGD identified by the literature. Data that lacks quality can result in confusion, less transparency and even less trust in government (Lněnička et al., 2021). Governments can help address this barrier by providing clear documentation and metadata about the datasets, as well as working to improve the quality and consistency of the data over time.

Another major barrier identified by the literature is citizens' lack of technical skills and expertise needed to access and analyze OGD. Governments can help address this barrier by providing the necessary tools to help users to analyze and understand the data in a more user-friendly way without the need to go through the raw data (Park and Gil-Garcia, 2022).

Finally, (Máchová et al., 2018) and (Lnenicka and Nikiforova, 2021) mention data discoverability as one of the most significant barriers faced by users of OGD. Even after the data is published in the portal, users often struggle in finding the data which is often associated with the incompleteness and inaccuracy of metadata describing open data.

2.4.5. RQ5 - What are the requirements for creating an OGD portals?

Making the data available is the first step towards implementing OGD. However, simply publishing the data doesn't guarantee the requisite level of transparency and doesn't provide any significant value to the public (Attard et al., 2015). Thus, it is important to provide a user-friendly portal to make the data accessible to a wide range of people while providing them with the necessary features and tools to process, understand and reuse the data.

In 2009, a working group from California wrote the Eight Open Government Data Principles that include the main best practices to share OGD (Attard et al., 2015; Thorsby et al., 2017). These principles include:

1. **Complete:** All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. **Primary:** Data is made available as it is available at the source, not in aggregated or modified forms.
3. **Timely:** Data is made available as quickly as necessary to preserve the value of the data.
4. **Accessible:** Data is available to the widest range of users for the widest range of purposes.
5. **Machine Processable:** Data is reasonably structured to allow automated processing.
6. **Non-Discriminatory:** Data is available to anyone, with no requirement of registration.
7. **Non-Proprietary:** Data is available in a format over which no entity has exclusive control.
8. **License-Free:** Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Nevertheless, although these principles still maintain their relevance, with the growth of the OGD movement, a need for a more detailed set of requirements for the creation of OGD portals has been identified to ensure that all OGD users can take advantage of the data presented in the portals.

(Simonofski et al., 2022) divides OGD portal users into two categories: expert users and lay citizens. Expert users are users that have experience in dealing with raw data while lay citizens are users that have no knowledge in the area. The literature indicates that current OGD portals are not satisfying lay citizens' requirements, for not presenting the information in a way that is easy for them to explore and understand.

Consequently, there has been a surge in studies aimed at identifying the necessary requirements for OGD portals to address these issues. Many articles provided a framework including various criteria and requirements to assess the usability and quality of an OGD portal. (Lnenicka and Nikiforova, 2021) outlined several features, listed in Table 8, that a good open data portal should possess to ensure the maximum use of open data by its users. (Lnenicka and Nikiforova, 2021)'s list of features is presented in this literature review for providing a more complete and detailed set of requirements while incorporating all the requirements identified in other articles.

Data quantity, structure, and general features of the portal

Providing statistical data, such as the total number of datasets and categories can demonstrate that the data is organized and available, as well as to prove that the portal is active. This helps users to get insight into the portal's current state providing a sense of confidence in the data published.

Another good practice in the creation of an OGD portal is to present the content of the portal in different languages, including in English, so that it can be accessed by people from different regions. Using dashboards is also crucial in an OGD portal, as they provide key performance data and allow users to understand the portal's activity level and facilitate interaction. Using standardized vocabulary and definitions is also important to ensure homogeneity and uniformity in the terms used, allowing users to understand the data more clearly and effectively.

Data linkage and data versioning are two other important requirements for OGD portals. By linking data together, users can easily discover and access related data, allowing them to gain a deeper understanding of the information presented, and data versioning ensures that users have access to the latest and most accurate data, while also allowing them to track changes over time.

Data quality

Ensuring the good quality of the data provided is a fundamental requirement. It is crucial to ensure the data is timely and frequently updated, and to perform data quality checks to ensure accuracy, completeness, and consistency. Providing information about each dataset, including a description of the dataset, its attributes, and metadata, is also crucial for users to gain insight into the data and its potential uses.

Additionally, it is also necessary to provide the data in both machine and human readable formats. According to (Simonofski et al., 2022), lay citizens prefer data in a human-readable format, and in some

cases, they don't even expect to download a dataset but simply want to consult it on a web page, while expert users expect a variety of datasets that are raw and exploitable and in a machine-readable format to easily use and combine them without requiring the use of additional software. Indicating conformity with the Open Data Five Star Scheme can help users gain insight into the data format and educate them about the dataset.

The Five Star Scheme for Linked Open Data, suggested by (Berners-Lee, T., 2010), provides a more technical guide towards publishing linked open data (Thorsby et al., 2017):

Table 7 - Five Star Scheme for Linked Open Data

<i>Stars</i>	<i>Interpretation</i>
*	"Available on the web (whatever format) but with an open license, to be Open Data
**	"Available as machine-readable structured data (e.g., excel instead of images can of a table)
***	"As (2) plus non-proprietary format (e.g. CSV instead of excel)
****	"All of the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
*****	"All the above, plus: Link your data to other people's data to provide context"

Data accessibility

Certain requirements must be met to ensure the data is accessible. Firstly, the data must be accessible through open access and a single common license. The data should also be easily downloaded in bulk to facilitate the use of a greater amount of data. Additionally, the portal should provide API and SPARQL endpoints to developers allowing them to have direct access to data catalogues and their content, and to update data using external systems, ensuring that the data can be easily integrated into other applications and tools. Another good practice that helps the accessibility of the data is to offer preview and discoverability options for datasets to allow users gain insight into the content of the datasets without needing to download them. These options could include metadata, data dictionaries, data previews, or data visualization tools. Providing data visualization and analytical tools is also crucial to ensure that data is accessible to all users, regardless of their level of expertise in dealing with raw data.

Data accessibility is closely related to data findability and understandability, as users must be able to locate and understand the data in order to use it effectively. Therefore, to ensure that data can be easily accessed and utilized by users, it is essential that the data is both findable and understandable.

Data findability

Ensuring data findability involves providing a search mechanism that enables users to find relevant data using keywords or by browsing through topics and thematic categories. Additionally, the portal should allow datasets to be filtered and sorted by various criteria such as popularity, newly added, data format, and publisher, to help users limit their search results. Finally, the portal should enable datasets to be categorized and catalogued, making it easier for users to select a subset of datasets based on their interests.

Data understandability

Providing information about each dataset can help users to better understand the data that is being presented. Also, as previously mentioned, the use of visualization and analytical tools makes the data more comprehensible for those who are not used to working with raw data but still want to explore the data provided by governments. Using visuals, such as charts, graphs, or maps to display the data in a more user-friendly way makes the data more comprehensible, accessible, and useful for a wider range of people, enabling lay citizens to get insights about government information of their interest without the need to download datasets and analyze the raw data.

Data usefulness

Creating an open government data portal requires ensuring that the data provided is useful to society. This involves identifying high-value datasets, measuring data interest through views and downloads, and assessing data quality and usefulness through ratings. Additionally, enabling the mapping between use cases and open datasets and allowing users to request new datasets of interest promotes the further reuse of open data.

Public engagement, collaboration, and participation

Promoting public engagement, collaboration and participation is also an important requirement when creating an OGD portal. This can be achieved by allowing users to share use-cases and maps between open datasets, which demonstrates how the data can be reused to produce valuable solutions. Providing comments, forums, and active feedback facilitates interaction between users and the data publisher, which helps improve the quality of the datasets and service. Users can also suggest improvements or report issues through forms and feedback reports. Requesting new datasets and tracking their state helps identify relevant and high-value datasets, while social media can be used to share feedback. Finally, organizing hackathons can increase public involvement and interest in using open data to develop solutions that benefit society.

Service quality

In order to create an effective OGD portal, it is important to ensure high service quality. This includes providing technical guidance and support to users on how to access and use the data, as well as offering helpful resources such as guidelines, tutorials, manuals, and FAQs. It is also important to have responsive personnel who can receive and track user feedback, providing follow-ups and allowing users to monitor the status of their requests or questions.

Platforms

The literature identified two main platforms that are used in the creation of an OGD portal that aim to present OGD in an accessible and user-friendly manner. In the United States, municipal governments work with Socrata, a specialized firm that helps simplify the process of releasing OGD. In other countries, CKAN has become the standard platform for OGD portals, although there are different ways of implementing the front-end of these portals. While CKAN-based portals maintain technical compatibility through standardized metadata and interaction tools on the back-end, variations in the front-end

implementation affect the user-friendliness of these portals. This suggests that the differences in usability of OGD portals mainly come from their front-end design, while the technical and operational aspects on the back-end tend to be similar across portals (Nikiforova and McBride, 2021).

Table 8 - Requirements for creating an Open Government Data Portal

<i>Requirement</i>	<i>Papers</i>	<i>#Papers</i>
Data quality	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Krismawati and Hidayanto, 2021), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	20
Data findability	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Chua et al., 2020), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021).	19
Data accessibility	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021).	18
Data understandability	(Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Chua et al., 2020), (Sabri et al., 2019), (Park and Gil-Garcia, 2022), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021).	18
Data quantity, structure, and general features of the portal	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021), (Safarov et al., 2017).	17
Public engagement, collaboration, and participation	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Zhu and Freeman, 2019), (Chokki et al., 2022), (Ansari et al., 2022), (Lněnička et al., 2021), (Chua et al., 2020), (Sabri et al., 2019), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Schauppenlehner and Muhar, 2018), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021).	17
Data usefulness	(Paige and Freund, 2019), (Moradi et al., 2022), (Attard et al., 2015), (Wang et al., 2021), (Lourenço, 2015), (Zhu and Freeman, 2019), (Ansari et al., 2022), (Lněnička et al., 2021), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Lnenicka and Nikiforova, 2021), (Máchová et al., 2018), (Mutambik et al., 2021).	13
Service quality	(Wang et al., 2021), (Zhu and Freeman, 2019), (Ansari et al., 2022), (Lněnička et al., 2021), (Sabri et al., 2019), (Nikiforova and McBride, 2021), (Simonofski et al., 2022), (Lnenicka and Nikiforova, 2021), (Thorsby et al., 2017), (Máchová et al., 2018), (Mutambik et al., 2021).	11

3. Research Problem

In the perspective of OGD Initiatives, the availability of data represents a significant step towards transparency. However, a critical challenge arises as the mere provision of data does not automatically ensure increased transparency, especially when the design and implementation of open data portals do not align with the diverse expectations of users (Ansari et al., 2022). This discrepancy brings to the forefront a fundamental issue that this research seeks to address: the gap between data availability and effective transparency in OGD.

In the business domain, users make use of information stored in databases to extract knowledge, driving decision-making processes, identifying trends, and gaining a competitive advantage. Translating this dynamic to the public sector, OGD initiatives aim to empower citizens by providing access to government data by making it available in OGD portals that serve as digital gateways to a wealth of information, fostering collaboration, innovation, and public accountability (Janssen et al., 2012). The process of making data available in OGD portals involves extracting information from various sources, such as government databases, organizational records, and public records. This raw data, which can be in formats like Excel, CSV, XML, JSON, or others, is then transformed into a suitable format and loaded into the portal using platforms like CKAN and Socrata (Nikiforova and McBride, 2021). This data is then accessible to users who navigate the portal interface to explore, analyze, and potentially extract valuable insights. However, the effectiveness of these portals is hindered by the absence of robust visualization tools, representing a lack of investment in enhancing the understandability of data which limits user interaction and engagement.

Visualization plays a pivotal role in translating complex datasets into easily understandable insights. There is a diversity of potential users of OGD and while certain user groups are accustomed to working with raw data and may prefer unaltered datasets, the majority of citizens don't have the knowledge necessary to work with raw data and require additional support in making sense of the data presented.

Although platforms such as CKAN and Socrata incorporate some visualization features, their scope is limited, not being able to address the diverse needs of users. These platforms, while serving as the backbone of OGD portals, provide visuals that are static and lack interactivity. This limitation hinders the user experience, discouraging citizens from exploring and interacting with the data.

This research problem involves the intricate relationship between data provision, citizen engagement, and the challenges faced by users in navigating OGD portals. Bridging the gap between data availability and effective transparency requires not only making data openly available but also investing in tools and features that enhance its comprehensibility. Grasping these various challenges is crucial for effectively implementing OGD initiatives and reaching the main objectives of transparency and empowering citizens.

4. Research Proposal

Contents

- 4.1. Objectives..... 21
- 4.2. Potential benefits of BI tools..... 21
- 4.3. BI tool implementation steps 22

In this chapter, the research proposal is defined as the solution for the research problem. In this proposal, a strategy to enhance the usability of OGD portals to make them accessible and comprehensible to a wider audience is outlined. The objectives will be first described.

4.1. Objectives

The main objective of this thesis was to provide a solution to the research problem addressed. More specifically, this thesis proposes the integration of Business Intelligence (BI) tools to create user-friendly visualizations that are accessible and comprehensible to a wider audience enhancing the usability of OGD portals.

4.2. Potential benefits of BI tools

In this section, the potential benefits of integrating BI tools into OGD portals will be defined. This phase explains why this solution is being suggested, clarifying how using BI tools can help solving the usability issues.

Business Intelligence (BI) tools are software applications, such as Power BI, Tableau, and QlikView, designed to transform data into useful insights (Ho and Jakli, 2010). Various sectors have adopted these tools due to their capacity to:

1. **Transform Raw Data:** BI tools excel in extracting, transforming, and loading (ETL) data from diverse sources into a unified, structured format. This transformation is a critical step in making data suitable for analysis and visualization.
2. **Analyze Data:** The analytical capabilities of BI tools empower users to uncover patterns, trends, and correlations within datasets. Advanced algorithms and data modeling facilitate in-depth analysis.
3. **Create visualizations:** Perhaps the most distinctive feature of BI tools is their ability to create visually compelling and interactive data visualizations. These visualizations simplify complex data, making it more accessible and comprehensible to users.
4. **Enable Data Interaction:** BI tools offer interactivity, allowing users to explore data, filter results, and drill down into details, fostering a deeper understanding of the information.
5. **Facilitate Reporting:** The reporting capabilities of BI tools enable the generation of customized reports and dashboards that present data in a user-friendly manner.
6. **User Accessibility:** BI tools are designed with a focus on user accessibility, ensuring that individuals with varying levels of technical expertise can utilize them effectively.

The potential benefits of BI tools integration in OGD portals are multifaceted. By harnessing these tools, government agencies can:

- **Enhance Data Accessibility:** Transform raw government data into user-friendly visualizations, enabling users to access and interpret data with ease.
- **Simplify Data Understanding:** Create visualizations that turn complex datasets into clear, comprehensible representations.
- **Foster User Engagement:** Engage users with interactive dashboards and reports, encouraging exploration and interaction with government data.

- **Support Evidence-Based Decision-Making:** Empower policymakers, researchers, and citizens to make data-driven decisions by providing them with actionable insights.
- **Bridge the Usability Gap:** Address the usability issues that have been a part of OGD portals, making government data more accessible to all.

4.3. BI tool implementation steps

The crux of the research lies in the implementation of BI tools within OGD portals. This phase outlines the three fundamental steps involved in the process:

Step 1: ETL (Extract, Transform, Load)

The ETL phase is the initial step in the implementation of BI tools. It involves a sequence of operations (Ho and Jakli, 2010):

1. **Data Extraction:** In this stage, data is extracted from diverse sources, often in different formats, including databases, spreadsheets, and APIs.
2. **Data Transformation:** The extracted data is then transformed into a uniform format. Data transformation activities encompass data cleaning, structuring, and standardization to ensure consistency and accuracy.
3. **Data Loading:** The transformed data is loaded into the BI tool's data warehouse, creating a centralized repository for analysis and visualization.

The ETL process sets the stage for comprehensive data analysis and visualization, ensuring that the data is in a format that leads to user-friendly presentations.

Step 2: Data Analysis

Following data preparation, the BI tool proceeds to the data analysis phase. BI tools offer analytical techniques and algorithms to uncover valuable insights within the datasets:

1. **Exploratory Data Analysis (EDA):** EDA techniques, including data profiling, summary statistics, and data visualization, are employed to gain a first understanding of the data, identifying trends and patterns.
2. **Statistical Analysis:** Statistical tests and models are applied to investigate relationships, correlations, and dependencies within the data.
3. **Advanced Analytics:** Advanced analytics techniques, such as "What-if scenarios" that allows you to predict what might happen in the future.

The data analysis phase plays a pivotal role in extracting meaningful insights from the transformed data, which will subsequently inform the creation of user-friendly visualizations.

Step 3: Creation of Reports and Visualizations

The culmination of the BI tool implementation process is the creation of user-friendly reports and data visualizations. This phase leverages the capabilities of the BI tool to design compelling and informative presentations of the analyzed data:

1. **Dashboard Design:** Dashboards serve as the user interface for interacting with data. They are designed to present key insights and visuals in a single, customizable view.
2. **Visualization Development:** Various types of visuals including charts, graphs, maps, and infographics, are crafted to convey information effectively. The choice of visual depends on the nature of the data and the insights to be communicated.
3. **Interactivity:** BI tools offer interactive features that allow users to explore data dynamically. Filters, slicers, and drill-down options enhance user engagement.
4. **Accessibility:** Accessibility considerations, such as screen reader support, color contrast, and font size, are integrated to ensure inclusivity.

The creation of user-friendly dashboards and visualizations is the ultimate goal of the BI tool implementation. These visualizations should empower users, including regular citizens, to explore, understand, and derive insights from government data effortlessly.

Recognizing the importance of accessibility and transparency in public data, the main goal of this thesis is to leverage visualization tools to enhance user experience and facilitate data exploration for individuals of varying backgrounds and expertise levels.

5. Demonstration

Contents

- 5.1. ANI - Agência Nacional de Inovação 25
- 5.2. Tool – Power BI..... 26
- 5.3. First Demonstration - SIFIDE 27
- 5.4. Second Demonstration – PT2020..... 34
- 5.5. Third Demonstration - HEurope 36

In this chapter, the demonstration step of the Design Science Research Methodology (DSRM) will be addressed, presenting the practical application of BI tools in improving data accessibility in the context of OGD Initiatives. This chapter provides a detailed illustration of the BI-driven visualizations created during the research, focusing on three key programs of a Portuguese governmental agency - *Agência Nacional de Inovação* (ANI): SIFIDE, PT2020 and HEurope.

The first section gives a brief description of the agency (ANI) and the three programs to which the visualizations will be made. Then, the tool used to create the visualizations is identified and a brief explanation on why that tool was chosen is given. The next three sections explain the visualizations created for each program, providing insights into the design and development processes. This phase involved collaborative efforts between the authors and ANI to ensure alignment with the needs of stakeholders.

5.1. ANI - Agência Nacional de Inovação

ANI - *Agência Nacional da Inovação* is a Portuguese governmental agency that aims to support scientific and technological innovation in Portugal. ANI occupies a central position in the relationship between science and economy, promoting the sharing of knowledge through cooperation and collaboration between institutions and companies.

ANI's primary data repository resides in OutSystems serving as the initial source of information. However, due to certain constraints, the direct uploading of data from OutSystems to Power BI proved unattainable. As a result, a transitional step involves extracting the data from OutSystems and storing it in Excel sheets that can be uploaded in Power BI.

While the ideal scenario was to perform the ETL operation directly within Power BI, technical limitations prevented this seamless integration. Consequently, the ETL stage continues to be executed during the extraction process to Excel sheets.

5.1.1. SIFIDE

SIFIDE – *Sistema de Incentivos Fiscais à I&D Empresarial*, is an initiative to provide tax incentives for research and development (R&D) in the business sector. It was created in 1997 with the purpose to stimulate innovation and competitiveness of companies in Portugal, encouraging investment in R&D activities. Over time, the incentive system has undergone several revisions to improve its characteristics. In 2011 SIFIDE II was implemented introducing significant changes to the legislation, making it even more attractive for companies and facilitating the active participation of companies in R&D activities.

SIFIDE offers tax benefits to companies that spend on research and technological development projects, through the deduction of R&D expenses when collecting Corporate Income Tax (*IRC*). SIFIDE also encourages adequate documentation and proof of R&D activities, ensuring transparency and compliance. The program is essential to promote innovation in companies and improve Portugal's competitiveness on the international stage.

5.1.2. PT2020

The PT2020 program refers to the Partnership Agreement between Portugal and the European Commission. This agreement encompasses the implementation of the five European Structural and Investment Funds during the period from 2014 to 2020. The five included are the European Regional Development Fund, the Cohesion Fund, the European Social Fund, the European Agricultural Fund for Rural Development, and the European and Fisheries Fund.

The Partnership Agreement establishes programming principles that guide Portugal's economic, social, and territorial development policy during the specified period. These principles are aligned with Smart, Sustainable and Inclusive Growth which is part of the Europe 2020 strategy.

5.1.3. HEurope

The HEurope program is dedicated to fostering the internationalization of higher education institutions and research centers in Portugal. The primary focus is on enhancing international mobility for students, teachers, and researchers, thereby increasing collaboration between Portuguese institutions and their foreign counterparts. The program strives to achieve these objectives through various mobility projects and by encouraging participation in international academic and research networks.

One of HEurope's key goals is to contribute to the internationalization of the Portuguese higher education system by promoting academic enrichment, knowledge dissemination, and the exchange of experiences on a global scale. The program places a strong emphasis on training qualified human resources, actively encouraging the mobility of students, researchers, and teachers.

Horizon Europe, the European Union's current Framework Program for Research and Innovation, is in operation from January 1, 2021, to December 31, 2027. With a substantial budget of 95.5 billion euros, the program is designed to support research and innovation activities throughout the European Union. The HEurope program came to replace the H2020 program that was active from 2014 until 2020.

In Portugal, ANI plays a pivotal role in coordinating the participation of the scientific, technological, and innovation community in Horizon Europe, specifically within Pillars II (Global Challenges and European Industrial Competitiveness) and III (Europe Innovative). To streamline and facilitate these activities, ANI has appointed National Contact Points (NCPs) responsible for promoting and disseminating actions related to their specific areas of intervention within the program. NCPs also provide guidance and assistance to interested parties during the application process.

ANI's NCPs have an additional responsibility in monitoring European Partnerships, particularly industrial ones, and Missions established under Pillar II of Horizon Europe. They serve as National Delegates in the Horizon Europe Pillars II and III Program Committees, actively participating in discussions and contributing to the definition of research and innovation priorities. This involvement spans across various biennial Work Programs launched from 2021 to 2027.

5.2. Tool – Power BI

Microsoft Power BI was chosen as the preferred tool for the creation of the visualizations due to its user-friendly interface and extensive list of features. Power BI stands out as an excellent for both visualization

creators and users due to its ease of use and comprehensive features. They can effortlessly drag-and-drop elements without the need to verify data completeness or correctness, streamlining the focus on data and dedicate more time to analyzing it. Its user-friendly interface empowers visualization creators to efficiently design and present data, while its simplicity and interactivity enhances the experience for end-users accessing and interpreting the data presented on portals. This dual benefit makes Power BI a valuable asset in fostering effective communication and understanding between data providers and the public.

Power BI supports data import from various sources, including databases, data warehouses, Microsoft Excel, or manual entry. The data loaded to Power BI Desktop can then be transformed in Power Query. The scripting language M facilitates simple and complex operations, handling large datasets, creating calculated columns, and addressing data issues. In Power Query, there is the flexibility to perform operations through a user-friendly interface, allowing the achievement of tasks with simple clicks. This visual approach is beneficial for users who may not be proficient in coding. However, if a more customized approach is preferred, Power Query also provides the option to write code using the M language. This gives greater control and ability to tailor operations according to specific requirements and accommodating to users with varying levels of technical expertise.

For the creation of reports, Power BI provides a diverse range of visualization types, including charts, tables, maps, and more. These visualizations can be customized to suit the specific needs of the audience, enhancing the overall presentation of data and making it more accessible and understandable. The reports created can be published to the web in Power BI Service to be shared and can be integrated into dashboards, consolidating essential information and key performance indicators (KPIs).

ANI had previously established Power BI visualizations for these programs, streamlining the ETL phase. The pre-existing visualizations served as the starting point, with the foundation ETL already completed. The main focus was to update the data and optimize visual representations for a more engaging interface. More interactive features were incorporated in order to improve the user experience, leveraging Power BI's flexibility.

5.3. First Demonstration - SIFIDE

This chapter provides an in-depth examination of the visualization process implemented for the SIFIDE program during the internship at ANI. The objective was to effectively communicate key project characteristics, including the number of applications, types of companies involved, investment details, and other pertinent metrics. The visualization structure for the SIFIDE program comprises four main pages: Cover, Summary, Report, and Website Link.

5.3.1. Cover Page

The Cover Page serves as an introduction, offering a succinct overview of ANI and introducing the SIFIDE program. This initial page establishes the context for stakeholders to comprehend the purpose and significance of the following pages. The Cover Page for the program SIFIDE is represented in Figure 2.



Figure 2 - SIFIDE Cover page

5.3.2. Summary Page

The Summary Page, represented in Figure 3, serves as a centralized hub for key indicators and KPIs related to the SIFIDE program, designed for quick reference. "Card" visuals highlight essential metrics, including received applications, approved applications, application approval rate, declared investment, cleared investment, investment approval rate, credit requested, credit assigned, tax credit approval rate, candidate companies, companies supported, and support fee. Additional visual elements, such as column charts and area charts, offer a graphical representation of data trends.

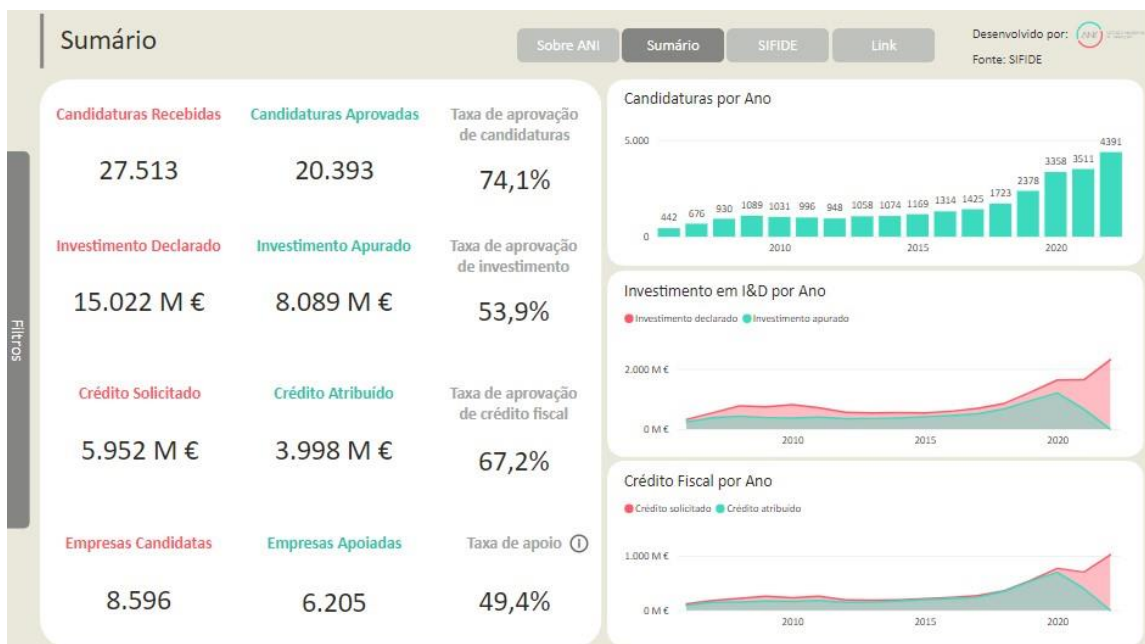


Figure 3 - SIFIDE Summary page

Interactive Features

- Column Chart:** Illustrates the number of applications received per year. As shown in Figure 4, hovering over a column triggers a tooltip displaying detailed information, including the number of applications received, approved, not approved, still being evaluated, and the application approval rate for each year.



Figure 4 – Tooltip for Applications per year

- Area Charts:** Illustrate trends in declared and cleared investment, as well as credit requested and assigned over the years. Figure 5 and Figure 6 show the tooltips displayed when hovering each chart.

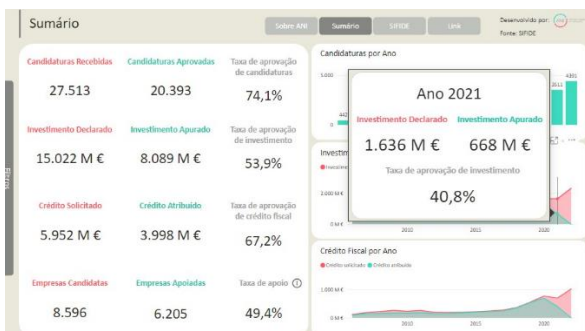


Figure 5 – Tooltip for declared and cleared investment per year



Figure 6 – Tooltip for requested and assigned credit per year

- Filter Pane Button:** Allows users to filter data on the page by year, sector, company dimension, company age, and district, as shown in Figure 7.



Figure 7 - Filter pane of SIFIDE Summary page

5.3.3. Report Page

The Report Page offers a more detailed exploration, providing users the option to focus on either received applications or approved applications. This section encompasses an array of visuals and interactive elements. Figure 8 shows the “Received Applications” section of the Report page.

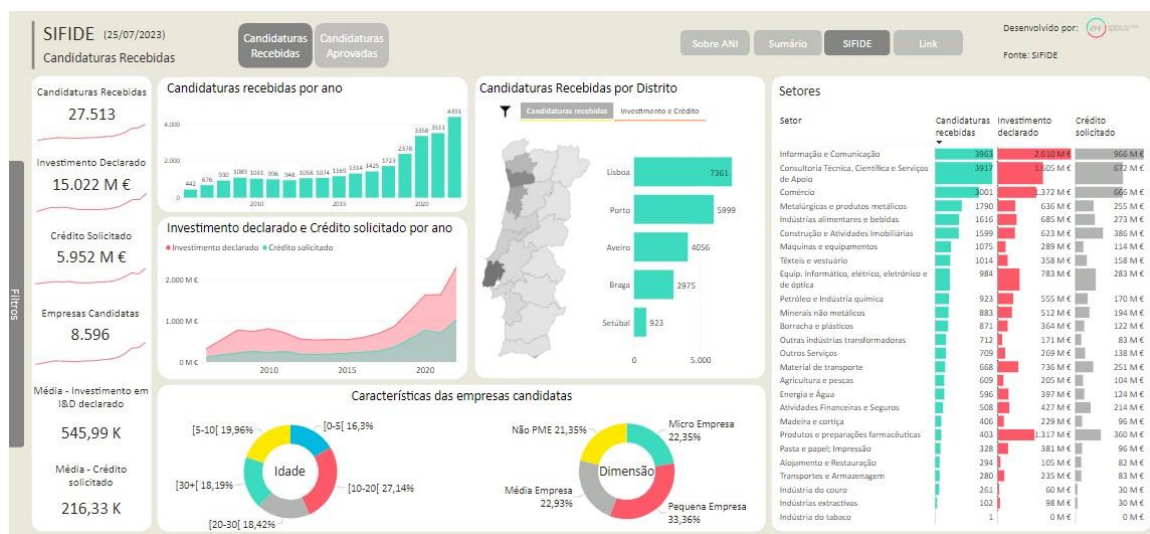


Figure 8 – “Received Applications” section of SIFIDE Report page

“Received Applications” Section

- **Key Indicators:** Highlights metrics such as the number of applications received, declared investment, credit requested, number of candidate companies, average declared investment, and average credit requested.
- **Column Chart:** Displays the number of received applications per year. As shown in Figure 9, hovering over a column triggers a tooltip displaying additional information, including the top 5

sectors with the most received applications and the number of candidate companies by company dimension and age.



Figure 9 – Tooltip for received applications per year

- **Area Chart:** Illustrates trends in declared investment and credit requested per year. As shown in Figure 10, hovering over a year triggers a tooltip with information on the top 5 sectors with the most declared investment and credit requested, as well as breakdowns by company dimension and age.

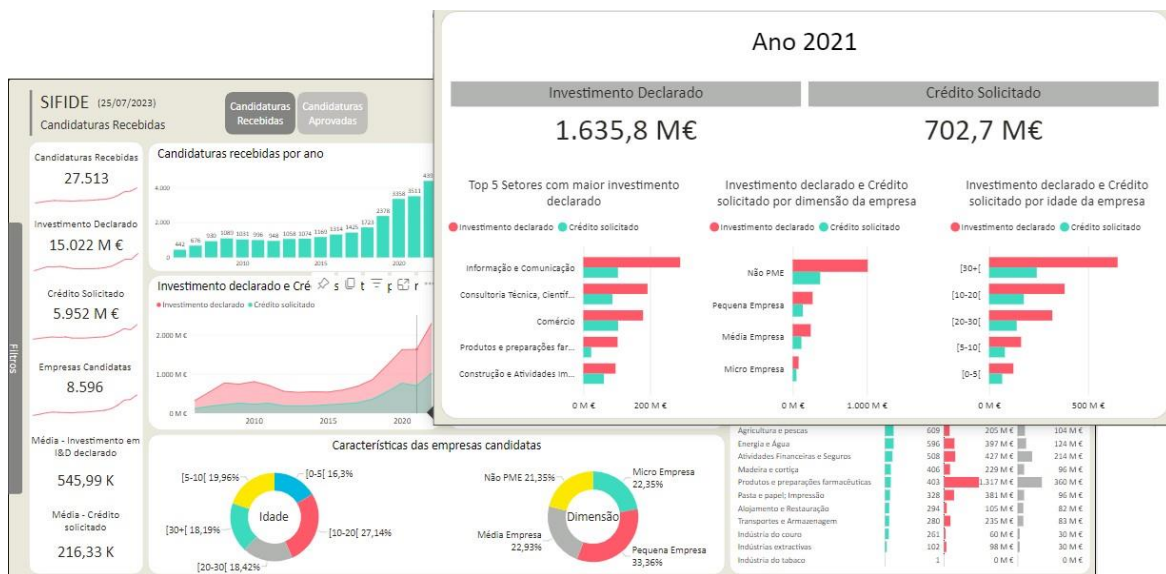


Figure 10 – Tooltip for declared investment and credit requested per year

- **Pie Charts:** Represent the distribution of candidate companies by age and dimension.
- **District-wise Application Map:** The page has a section where information is presented by district. It can be divided in two sections by selecting one of the options presented, “Received applications” or “Investment and Credit”. If the “Received applications” option is selected, this section shows the data for applications by district. It utilizes a map visual where darker districts indicate higher received application volumes. As shown in Figure 11, hovering over a district

triggers a tooltip showing the number of received application and candidate companies and a bar chart displaying the number of received applications by municipality. Next to the map, another bar chart depicts the top 5 districts with the most received applications. As shown in Figure 12, hovering a bar triggers a tooltip with information on the top 5 sectors with the most received applications in the district in question, as well as breakdowns by company dimension and age. If the “Investments and Credit” option is selected, it shows the same visuals but related to Investment declared and Credit requested, as shown in Figure 13 and Figure 14.

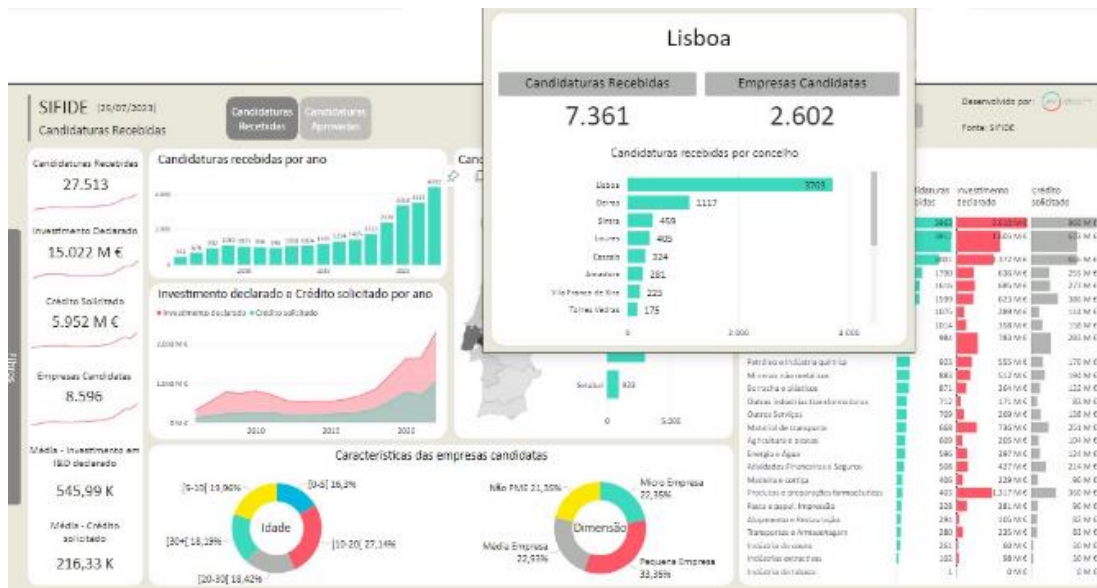


Figure 11 – Tooltip for received applications per municipality

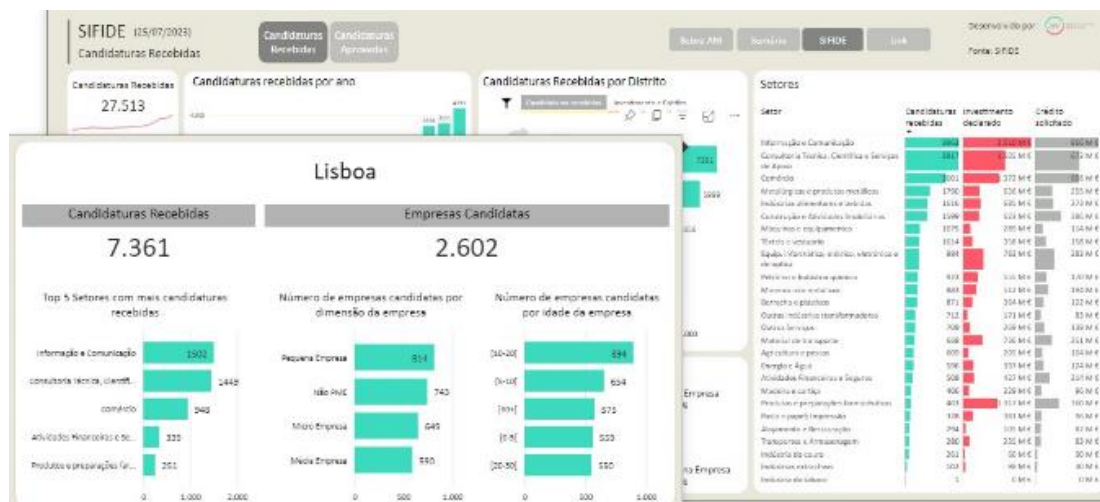


Figure 12 - Tooltip for received applications per district

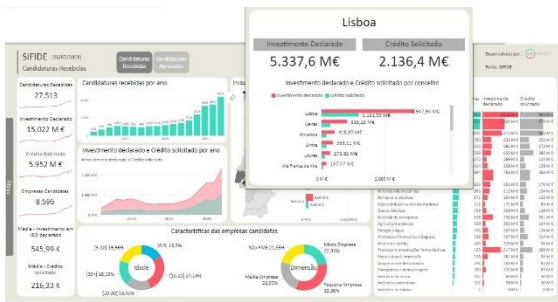


Figure 13 - Tooltip for declared investment and credit requested per municipality

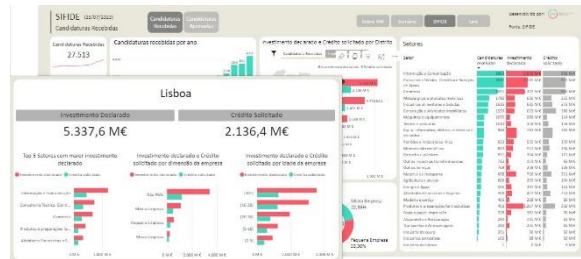


Figure 14 - Tooltip for declared investment and credit requested per district

- **Table:** A table visual is embedded in the Report Page, presenting detailed information on all sectors. This includes the number of received applications, declared investment, and requested credit for each sector. The table provides stakeholders with a granular view of sector-specific data, enhancing the comprehensiveness of the information presented.
- **Filter Pane Button:** Allows users to filter data on the page by year, sector, company dimension, company age, and district, as shown in Figure 15.

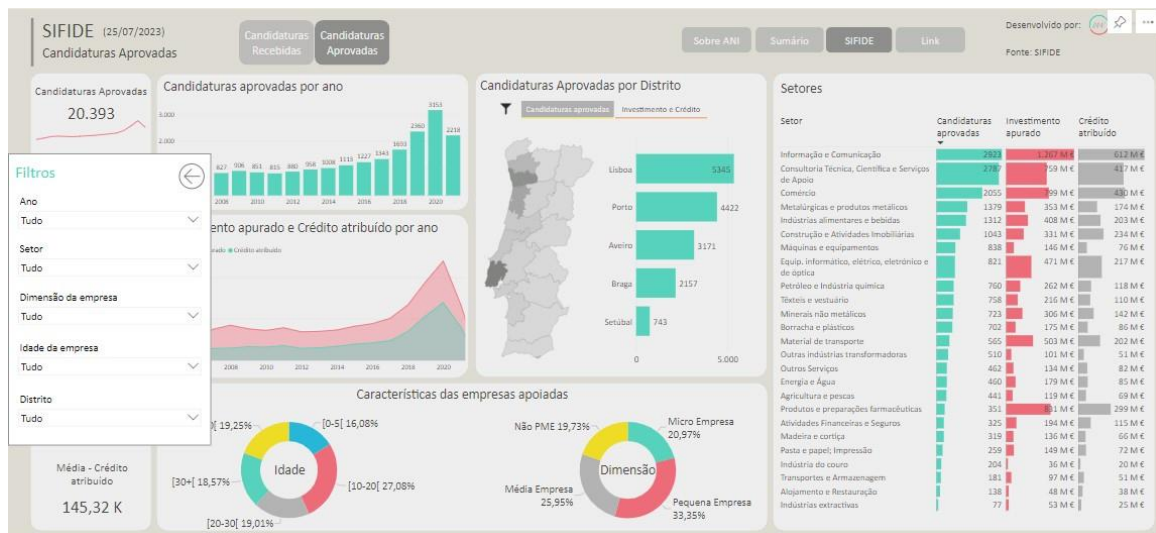


Figure 15 - Filter pane of SIFIDE Report page

“Approved Applications” Section

All the visuals mentioned for the “Received Applications” section are replicated for the “Approved Applications” section, providing detailed insights into approved applications, cleared investment, credit assigned, and companies supported. Figure 16 shows the “Approved Applications” section of the Report page.

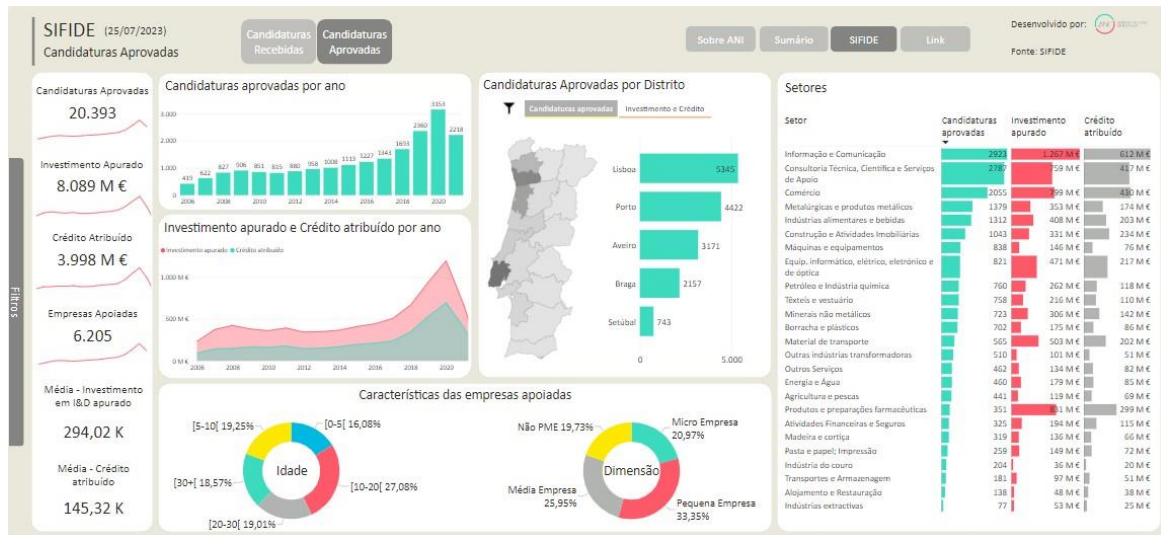


Figure 16 – “Approved Applications” section of SIFIDE Report page

5.3.4. Website Link Page

The final page, represented in Figure 17, exclusively contains a link to the official SIFIDE program website, offering users direct access to additional information and resources.



Figure 17 - SIFIDE Link page

The design of each page prioritizes seamless navigation, with buttons facilitating movement between the Cover, Summary, Report, and Website Link Pages. The interactive features included in the visuals enhance user engagement and data exploration, ensuring a user-friendly experience.

5.4. Second Demonstration – PT2020

The second demonstration refers to the PT2020 program. The structure of the visualization for this program is very similar to the one for the SIFIDE program.

5.4.1. Cover and Link Pages

The structure of the Cover and Link pages for PT2020 closely mirror that of SIFIDE, providing a consistent framework for program introductions and navigation.

5.4.2. Summary Page

The Summary page, represented in Figure 18, provides key indicators and a column chart for users to quickly grasp some of the most important information.



Figure 18 - PT2020 Summary page

5.4.3. Report Page

Similar to the SIFIDE program, this page also gives users the option to analyze the data for received applications or approved projects. The structural layout and incorporated features closely resemble those utilized in the SIFIDE program. The “Received Applications” section of the report page is shown in Figure 19. All the visuals included in the “Received Applications” section are replicated for the “Approved Projects” section.



Figure 19 – “Received Applications” section of PT2020 Report page

Nevertheless, a distinctive feature was added in the PT2020 program – the drill-down functionality. This feature allows users to either aggregate or disaggregate data, introducing a dynamic element to the analysis process.

Figure 20 and Figure 21 are both representations of the bar chart that displays the number of received applications and the number of promoters per support measure. There are 17 different support measures that can be aggregated in 3 categories, as shown in Figure 21. The drill-down functionality empowers users to adjust the level of detail based on their analytical preferences, enabling them to explore and understand the data at varying levels of granularity. Figure 20 represents the bar chart after the drill-down functionality is accessed, disaggregating the three categories into the 17 support measures.

To access this functionality, the user has to click on the arrow icons presented in the action bar that is displayed when hovering the chart. An information icon is positioned next to the chart to give the users an explanation on how to access the functionality.

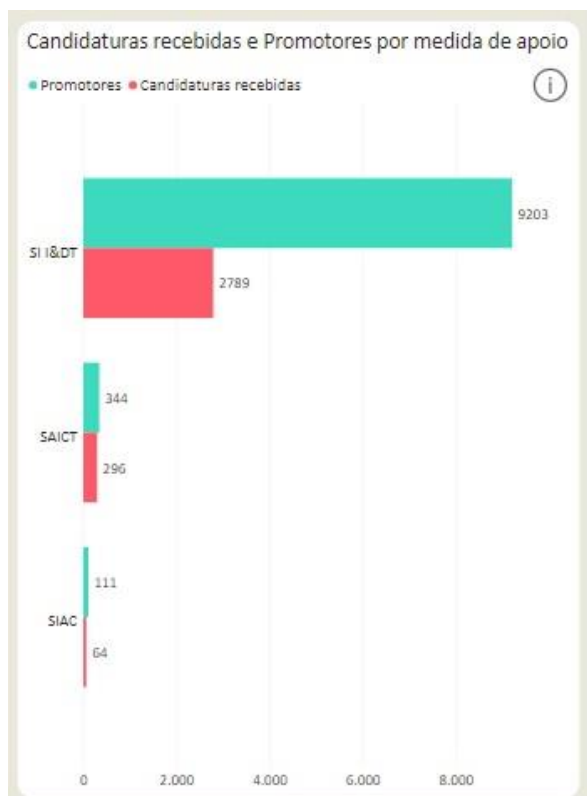


Figure 21 - Bar chart aggregated into 3 categories



Figure 20 - Bar chart disaggregated into 17 support measures

5.5. Third Demonstration - HEurope

This chapter delineates the distinct visualization approach employed for the HEurope program, highlighting the unique aspects that differentiate it from the previous programs. For HEurope, the data extraction from OutSystems to Excel differs from the other programs. Instead of extracting the original data, only the results were captured. This limitation reduced the granularity of information available for visualization, restricting the level of interactivity that could be achieved in Power BI.

5.5.1. Cover and Link Pages

The structure of the Cover, Summary and Link pages for HEurope closely mirror that of PT2020 and SIFIDE, providing a consistent framework for program introductions and navigation.

5.5.2. Summary Page

The Summary Page for HEurope, represented in Figure 22. provides key indicators for both H2020 and HEurope programs, to compare the H2020 program that had ended in 2020 with the HEurope program that started upon the end of H2020, in 2021. This page provides the users with an option to access a chart displaying the evolution of approved projects and financing obtained since 2014 until 2023. This chart, represented in Figure 23, is accessed by clicking on the arrow at the top right corner of the Summary Page.

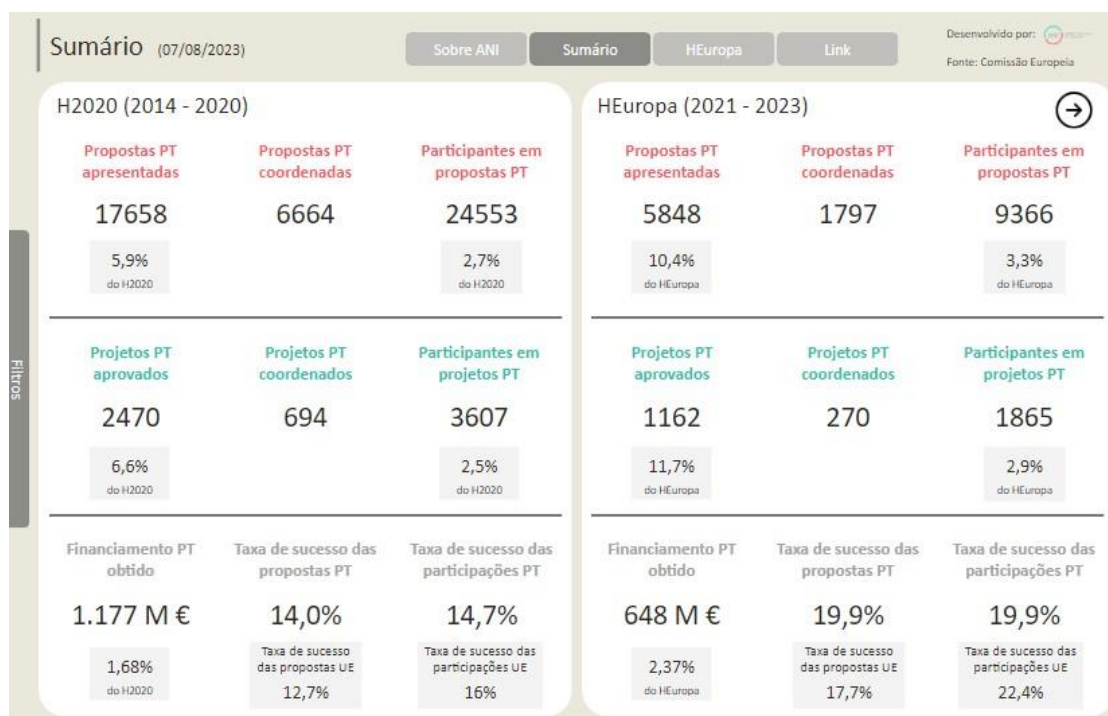


Figure 22 - HEurope Summary page

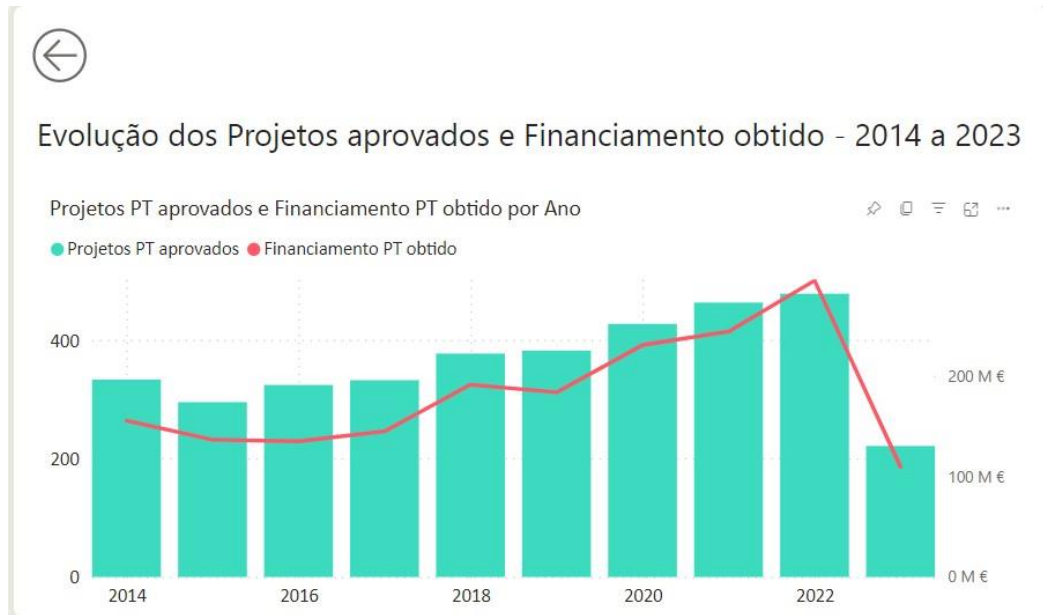


Figure 23 - Evolution of approved projects and obtained financing from 2014 until 2023

5.5.3. Report Page

The Report Page for HEurope necessitated a distinctive approach due to the limitations imposed by the extracted data. Two main sections were developed – one providing an analysis per year and the other per theme. Both sections show information on received applications, approved applications, obtained financing and participations, but the “Analysis Per Year” section show this information per year and the “Analysis Per Theme” show this same information per theme.

Analysis Per Year

Figure 24 shows the “Analysis Per Year” section of the Report page.



Figure 24 – “Analysis per Year” section of HEurope Report page

- Column Charts:** Figure 25 represents the column chart displaying the number of presented proposals and approved projects per year and the tooltip triggered when hovering over a column, showing a card visual with the success rate of Portuguese proposals. Figure 26 represents the column chart displaying the obtained financing per year and the tooltip triggered when hovering a column, revealing a pie chart depicting obtained financing per entity type. Another column chart is presented comparing the number of presented proposals with the number of coordinated proposals. The user is given the option to also compare the number of approved projects with the number of coordinated projects.



Figure 25 - Tooltip for Success rate of Portuguese proposals per year



Figure 26 - Tooltip for obtained financing per year

- Pie Chart:** Represents participations in proposals or projects per entity type. As shown in Figure 27, hovering over a slice reveals a tooltip with a table visual displaying participant numbers and success rates per year.



Figure 27 - Tooltip for Participations

- Filter Pane Button:** Allows users to filter data on the page by year and entity type, as shown in Figure 28.

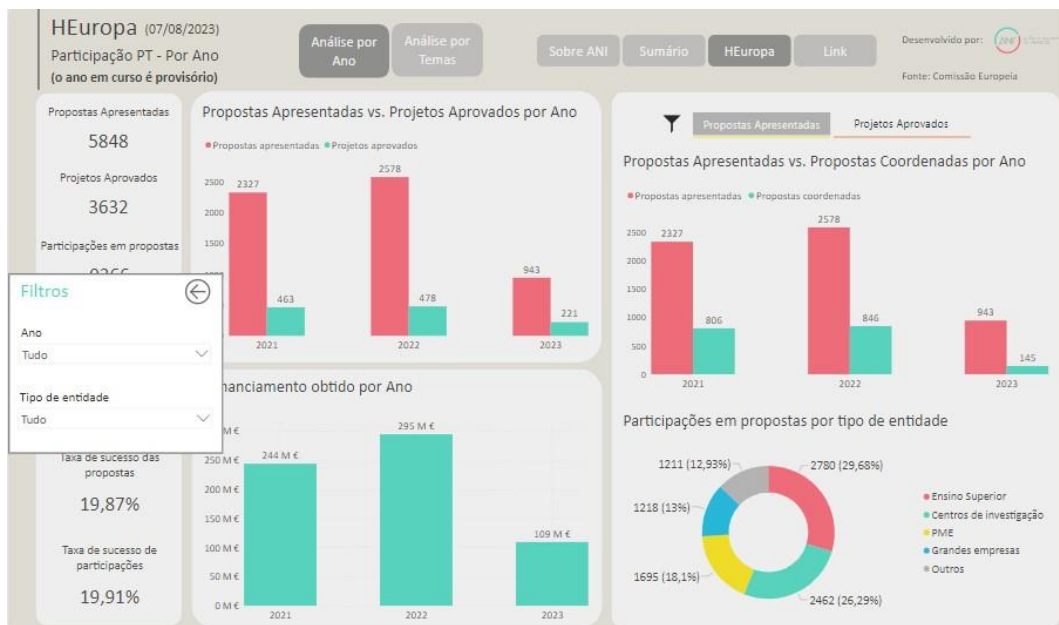


Figure 28 - Filter pane of "Analysis per Year" section of HEurope Report page

Analysis Per Theme

Figure 29 represents the “Analysis Per Theme” section of the Report page.



Figure 29 – “Analysis per Theme” section of HEuropa Report page

- Column Chart:** Displays the number of presented proposals and approved projects per theme. As shown in Figure 30, hovering over a column triggers a tooltip showing a card visual with the success rate of Portuguese proposals.

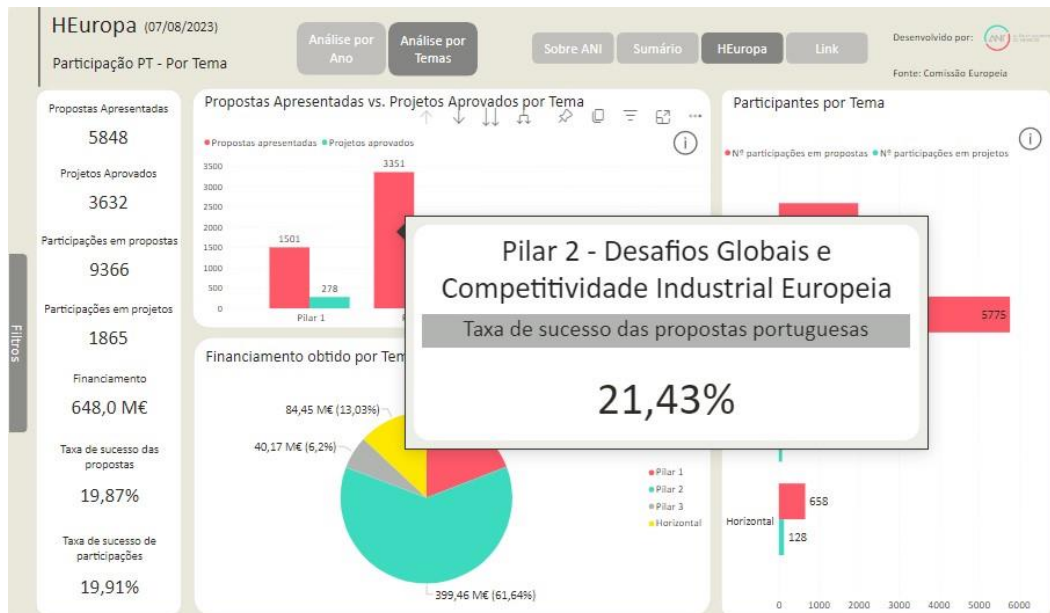


Figure 30 - Tooltip for Success rate of Portuguese proposals per theme

- Pie Chart:** Illustrates obtained financing per theme.

- Bar Chart:** Shows the number of participants per theme. As shown in Figure 31, hovering over a bar reveals a tooltip with a table visual displaying participant numbers and success rates per entity type.



Figure 31 - Tooltip for Participations per theme

- Filter Pane Button:** Allows users to filter data on the page by theme and entity type, as shown in Figure 32.

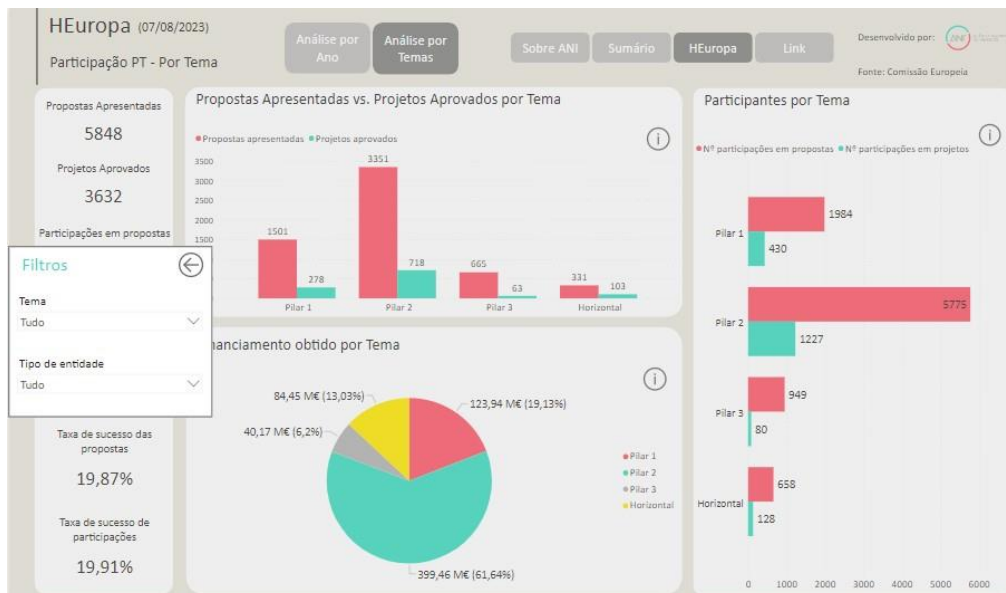


Figure 32 - Filter pane of “Analysis per Theme” section of HEurope Report page

All visualizations on this sector offer drill-down functionality. Each theme is categorized, allowing users to view information per category or per theme. Icons with explanatory tooltips facilitate user understanding and navigation through the drill-down and drill-up features.

6. Evaluation

Contents

- 6.1. Description 44
- 6.2. Sampling 44
- 6.3. Data Collection 45
- 6.4. Results 45

In this chapter, the methodology used to evaluate the efficiency of the visualizations developed for the three programs within ANI is outlined. (Venable et al., 2016) discussed in an article two paradigms for evaluation studies: artificial evaluation and naturalistic evaluation.

Artificial evaluation focuses on controlled experiments, simulations, and theoretical arguments to test design hypotheses. It tends to be positivist and reductionist, aiming to prove or disprove the utility of design artifacts (Venable et al., 2016).

In contrast, naturalistic evaluation explores the performance of a solution technology in its real environment, embracing the complexities of human practice within organizations. It tends toward interpretivism and includes methods such as case studies, field experiments, surveys, and ethnography (Venable et al., 2016).

The type of approach in this work to evaluating the efficiency of the visualizations in Power BI aligns with the naturalistic evaluation paradigm. By conducting interviews with a diverse group of participants, including those unfamiliar with data analysis tools and those experienced with Power BI, it is sought to understand how the visualizations perform in real-world scenarios. This approach acknowledges the complexities of human interaction with the visualizations and allows for a more holistic assessment of their effectiveness. By collecting qualitative data on users' navigation patterns, preferences, and challenges, this work aimed to provide insights into the real-world usability and utility of the visualizations.

6.1. Description

In the interview process, a specific set of questions and tasks, that can be consulted on section A.1 (Appendix A), were tailored for each program's visualization. These questions were designed to request the participants to find specific information within the visualizations, allowing the interviewer to observe their navigation process, time taken, and overall interaction with the visualizations. This approach provided valuable insights into how users engage with the visualizations and allowed for the identification of potential usability issues or areas for improvement.

6.2. Sampling

Unlike statistical surveys, which prioritize statistical reliability through probability sampling, qualitative studies aim to explore the diversity of phenomena within the target population. Qualitative research emphasizes the need to capture the diversity of the phenomenon under study within the target population. Rather than aiming for a large random sample, qualitative sampling prioritizes purposive selection to ensure coverage of all relevant varieties of the phenomenon, a concept known as saturation (Jansen, 2010). In qualitative surveys with open-ended questions, each response is unique, being necessitating a focus on diversity in the sample to ensure comprehensive coverage of the population under study. By prioritizing diversity over quantity in sampling for qualitative research, researchers can achieve a comprehensive understanding of the phenomenon under study. This approach allows for the exploration of rich, nuanced insights and ensures that the research findings are grounded in the diverse perspectives and experiences of the target population (Jansen, 2010).

Therefore, the selection for this study aimed to ensure a diverse representation of individuals with varying levels of familiarity with data analysis tools, educational backgrounds, professions, ages, and genders. A total of 10 interviewees participated in the evaluation process, providing diverse perspectives on the effectiveness of the visualizations across the three programs.

Participants with limited or no experience with working with data were included to assess the accessibility of the visualizations to a wider audience, particularly focusing on the common citizen who may encounter difficulties in understanding data presented on portals.

On the other hand, participants familiar with data analysis were also included. Those who have knowledge of working with data but are inexperienced with Power BI or similar business intelligence tools were included to evaluate the intuitiveness of the tool and its effectiveness in improving data navigation. Experts in both data analysis and Power BI usage were included to gather constructive feedback and suggestions for improvement from experienced users.

Furthermore, participants from various educational backgrounds and professions were included to capture a broad range of perspectives and insights. This is also related to the varying levels of knowledge in working with data due to the specific skills, training, and experiences acquired within each field or discipline.

Participants with different age groups were also included to explore potential age-related differences in interaction with the visualizations and to assess the impact of age on usability and effectiveness. Lastly, both male and female participants were included to ensure gender diversity within the sample.

6.3. Data Collection

Data was collected during the interviews. The interviews were conducted in person, allowing for direct observation of participants' reactions and behaviors. Observations were made regarding participants' navigation patterns, their ability to locate desired information within the visualizations, and the time taken to complete assigned tasks. Throughout the interviews, participants were encouraged to verbalize their thoughts and experiences, articulating any challenges or difficulties encountered while navigating through the visualizations. This verbal feedback offered additional context and depth to the observations made during the interview sessions.

At the end of the interview, participants were asked to complete a questionnaire aimed at evaluating the user-friendliness and intuitiveness of the visualizations, as well as their experience with Power BI as a tool. It was also asked feedback on aspects such as preferred features, challenges encountered, and suggestions for improvement. The questionnaire can be consulted on section A.2 (Appendix A).

Experts in working with Power BI were not asked to find information, but rather just explore the visualizations and provide constructive feedback while exploring.

6.4. Results

In this section, the results obtained from the interviews are presented. Gender was found not to impact interactivity with the visualizations. However, the analysis revealed a distinct difference between the

elderly participants and those in other age groups. This was expected, as working with technology is more challenging for the elderly group.

Furthermore, participants' level and type of education, as well as their profession, were found to have an impact. Nevertheless, these factors are correlated with the level of expertise in data analysis.

So, the evaluation of results places particular emphasis on discerning differences among participants with varying levels of experience in working with data.

Completion of tasks

All participants successfully found the required information within the visualizations. However, participants with limited experience in working with data required additional guidance from the interviewer to access certain types of information. A common complaint among all participants was the drill-down and drill-up feature. Participants with no experience in working with data required guidance to locate and activate this feature from the action bar. Participants with some experience in data analysis could activate the feature with assistance, although taking longer than it should. Despite the presence of an information icon intended to provide guidance on how to use this feature, most of participants did not recognize its purpose clearly, indicating its lack of explicitness. This hindered participants' ability to utilize the feature effectively, highlighting the need for improved clarity in user interface design. Experts in Power BI acknowledged the feature's lack of intuitiveness. One participant who regularly works with Power BI admitted not using the feature frequently, and he wasn't totally familiar on how to use it. It was then concluded that this feature may not be suitable for visualizations intended to be used by individuals who have no experience with Power BI.

Exploration and Time Efficiency

Participants with no experience in working with data spent less time exploring the visualizations but required more time to locate the required information. Initially, there was some hesitation among these participants in exploring the visualizations deeply, possibly due to feeling overwhelmed by the amount of information provided and a lack of familiarity with using such platforms. However, once the questions were posed and they started to navigate to find the information, they became more comfortable with the software. In contrast, experienced participants didn't find the amount of information overwhelming and navigated the visualizations more swiftly, leveraging their familiarity with data analysis tools to expedite information retrieval. Some information was found in tooltips when hovering over charts. Participants who didn't explore the visualizations deeply took longer to understand that they could access additional information by hovering over charts, while those who explored the visualizations understood this feature. A suggestion made by participants with experience in data analysis was to provide information on the additional details available within each chart.

Additionally, participants with no experience working with Power BI or similar tools, regardless of their experience in working with data, did not realize that they could filter the entire page by clicking on charts. They would always use the filter pane to filter the page. In contrast, participants with experience in Power BI were faster to access information because they would filter the page by clicking on charts instead of opening the filter pane.

Design

Participants provided predominantly positive feedback regarding the overall organization of the visualizations, expressing appreciation for their user-friendly and intuitive nature. They particularly enjoyed the clarity of the charts and commended the effectiveness of the titles. However, they also offered constructive suggestions for improvement.

Some participants identified design issues within the HEurope visualization, particularly on the Summary page where H2020 and HEurope programs are compared. One participant commented, *“Do not present data slightly unrelated to the core dashboard's purpose with the same 'visual strength' as other more relevant data.”* However, experts in Power BI did not find this issue troubling, attributing its clarity to the titles in each section. Nevertheless, they suggested providing a brief description of the H2020 program on the cover page, clarifying its transition to HEurope from 2020 onwards. Additionally, in the report page dedicated to the HEurope program, they suggested including dates to avoid confusion wondering if the page included the H2020 program as well.

Additionally, some participants with experience in data analysis recommended enhancing the visibility and prominence of the buttons labeled "Candidaturas Recebidas" (Received Applications) and "Candidaturas Aprovadas" (Approved Applications), treating them more like prominent titles rather than mere buttons. This suggestion implies that participants felt these elements were crucial focal points within the visualization interface and should be visually emphasized to draw users' attention effectively. However, experts in Power BI did not encounter issues with these buttons, quickly grasping their purpose and function within the visualization interface.

It was also suggested among the participants the inclusion of a "clear all filters" button allowing users to reset filters effortlessly, bigger font size across the interface to improve readability and ensure a more user-friendly experience, and the inclusion of a comprehensive quick guide outlining all functionalities offered by Power BI, aiming to familiarize users with the tool's capabilities and enhance their overall navigation experience.

Among the experts in Power BI, the feedback provided included:

- **Static Header:** Maintaining a static header across all pages improves user experience with visualizations. However, a trade-off may be necessary between a static header and optimizing page layout. For instance, on the "Summary" page a smaller frame was used compared to the report page, that had to accommodate more detailed information, so optimization of the pages was prioritized.
- **Graphs for Evolution of Success/Approval Rates:** Experts found charts showing the evolution of success/approval rates per year valuable for providing insight, noting their effectiveness in conveying relevant information. For instance, they suggested adding a button next to the "Received applications per year" column chart, allowing users to click on it to view the evolution of applications' approval rates per year.
- **Consistency in Colors and Charts:** Experts emphasized the importance of consistency in the use of colors and charts. For example, while hovering over the "Received applications per year"

column chart, bar charts displaying the number of candidate companies by dimension and age are shown for each year. However, below the “Received applications per year” chart, total candidate companies by dimension and age are represented in pie charts. It was recommended to maintain consistency by using bar charts throughout and ensuring the same order of categories.

The overall feedback from the experts was positive across the visualizations. They expressed satisfaction with the simplicity and clarity of summary pages, noting their effectiveness in conveying information concisely. Additionally, the report pages received praise for their structural layout and intuitiveness, facilitating smooth navigation and comprehension. They appreciated the transparent presentation of data, leaving no ambiguity on what information users are viewing. Suggestions for improvement mainly focused on enhancing the user experience, reflecting a constructive approach to refinement.

7. Conclusion

Contents

- 7.1. Approach and Results 50
- 7.2. Limitations 50
- 7.3. Future Work 51

In the final chapter of this thesis, the key findings of this work are summarized, the limitations encountered are discussed and the potential directions for future research are outlined. The first section of this chapter reviews the methodology employed and the outcomes achieved. The second section identifies the constraints and challenges faced during the study. Finally, the third section proposes areas for further investigation and potential improvements.

7.1. Approach and Results

Following the DSRM, this study undertook a demonstration phase involving the creation of visualizations for three key programs of ANI (SIFIDE, PT2020, HEurope) using Power BI. These visualizations were created with a focus on user-friendliness, interactivity, visual appeal, and ease of use.

Upon completion of the visualizations, interviews were conducted with participants to evaluate their effectiveness in enhancing data accessibility and user experience. The findings from these interviews provided valuable insights into the strengths and weaknesses of the visualizations created.

The results highlighted the significant role that visualization tools can play in making data more accessible and user-friendly. However, several areas for improvement were identified. A concern mentioned was the overwhelming amount of information presented on a single page, which could be intimidating for users with limited data analysis experience. Additionally, participants noted the need for instructions regarding the basic functionalities of the tool, such as filtering the page by clicking on charts and accessing additional information through tooltips when hovering over charts. Another feature that created confusion among most participants, was the drill-down/drill-up functionality. While these features were available within the visualizations, participants indicated a lack of awareness or understanding of how to utilize them effectively. This highlights the importance of providing clear and concise instructions within the visualization interface to guide users in maximizing the utility of the tool.

Some concerns raised by participants with low expertise did not appear to be problematic for experts, who found the interface intuitive. Nevertheless, the primary objective of this work was to enhance user interaction for those with limited expertise. Therefore, even if experts did not perceive certain issues as problematic, the feedback from less experienced users was given significant consideration to address their needs effectively.

Overall, participants expressed positive feedback regarding the accessibility of the visualizations. They appreciated the clarity of the charts, explicit titles, and intuitive navigation features, which facilitated their ability to explore and understand the data effectively. Additionally, the interactivity of the visualizations was highlighted as a key factor in enhancing the user experience, allowing participants to interact with the data dynamically and gain deeper insights.

7.2. Limitations

A limitation of this work is the limited number of iterations conducted to fully optimize the efficacy of the visualizations. More iterations would have been crucial to refine the visualizations developed while addressing the identified improvements from the interviews. Inadequate iteration may result in some shortcomings or missed opportunities for improvement to be adequately addressed.

However, within the constraints of an academic project, it was developed a proposal that demonstrates the potential for improving data accessibility and usability. It is believed that this work offers valuable insights into the challenges and opportunities in this domain. It serves as a foundational step towards future research and development endeavors aimed at improving the accessibility and usability of data visualization tools.

Another limitation relates to the sample used for the evaluation of the visualizations. The primary users of ANI's visualizations are individuals who actively seek specific information and are already familiar with the terminology and language used within ANI's programs (SIFIDE, PT2020, HEurope). Contrarily, the participants involved in this study were not familiar with ANI's programs or the specialized terminology associated with them.

This discrepancy between ANI's main users and the participants poses a challenge in assessing the effectiveness of the visualizations, as the participants lacked the contextual understanding typically possessed by regular users of ANI's data portals. Despite this limitation, the study aimed to evaluate the intuitiveness of the visualizations and their ability to facilitate information retrieval, even among individuals unfamiliar with the specific terminology used in ANI's programs. While this limitation may have influenced the participants' interaction with the visualizations, the study was able to evaluate whether the visualizations could overcome this barrier and remain intuitive for users with limited familiarity with ANI's programs.

Another notable limitation is related to data extraction processes. The inability to directly connect ANI's database in OutSystems to Power BI hindered the examination of Power BI's efficiency in performing ETL processes. Power BI offers significant advantages not only to end-users but also to developers during the development process. Power BI offers direct connectivity to a wide range of data sources, including databases, cloud services, and online platforms. This feature enables real-time data analysis and automatic updates, allowing users to access the most up-to-date information effortlessly. Additionally, Power BI provides advanced functionalities for data modeling, cleaning and transformation, streamlining the ETL process and enhancing data quality. Unfortunately, due to the limitation in data extraction methods, the study could not fully explore in depth these capabilities of Power BI. However, it is essential to acknowledge the platform's potential in simplifying data integration and analysis tasks, which could significantly benefit both end-users and developers.

It's crucial to underscore the significance of presenting high-quality, accessible data for governmental agencies, as it fosters transparency and trust among citizens. However, achieving this objective requires extensive research and effort to develop portals that effectively cater to diverse user needs and seamlessly integrate with data analysis tools. Addressing these limitations and refining methodologies in future studies will be essential in advancing the accessibility and usability of governmental data portals.

7.3. Future Work

Following this work's limitations, there is still future work that could be made. Future research could delve deeper into the development of a comprehensive open data portal. While this study primarily

focused on enhancing data visualization using Power BI, a holistic approach to open data portals involves considering more aspects beyond visualization. As mentioned in the literature review, creating an effective open data portal entails fulfilling various requirements.

Furthermore, to advance the field, future research endeavors could explore strategies to integrate disperse data sources into a unified platform. This involves addressing technical challenges such as data interoperability, ensuring compatibility across different data formats and systems. Integrating and connecting visualization tools such as Power BI with the underlying infrastructure of data portals is also a relevant step, ensuring smooth integration and interoperability across the entire data ecosystem. By creating a cohesive and interconnected environment, organizations can enhance data accessibility, streamline analysis processes, and empower users to derive actionable insights more effectively.

Collaborative efforts between researchers, policymakers, and technologists are essential to driving innovation and progress in the realm of open data portals. By fostering interdisciplinary collaborations and leveraging emerging technologies, such as machine learning and natural language processing, future endeavors can unlock new possibilities in data exploration and analysis, empowering users to derive actionable insights from vast datasets.

In conclusion, while this thesis represents a foundational step towards enhancing data accessibility and visualization within governmental open data portals, there remains a wealth of opportunities for further research and innovation. Embracing a multidimensional approach and leveraging the collective expertise of diverse stakeholders enables the ongoing advancement of open data, unlocking its transformative potential to promote transparency, accountability, and informed decision-making within the public sector.

References

- Amalia, D.N., Susanto, T.D., 2019. Analysis of Motivation and Perceived Risk Factors in Open Data Measurement: A Conceptual Model. *Procedia Comput. Sci.* 161, 308–315. <https://doi.org/10.1016/j.procs.2019.11.128>
- Andersen, K.V., Henriksen, H.Z., 2006. E-government maturity models: Extension of the Layne and Lee model. *Gov. Inf. Q.* 23, 236–248. <https://doi.org/10.1016/j.giq.2005.11.008>
- 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. *Gov. Inf. Q.* 39. <https://doi.org/10.1016/j.giq.2021.101657>
- Attard, J., Orlandi, F., Scerri, S., Auer, S., 2015. A systematic review of open government data initiatives. *Gov. Inf. Q.* 32, 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Chokki, A.P., Simonofski, A., Frénay, B., Vanderose, B., 2022. Engaging Citizens with Open Government Data: The Value of Dashboards Compared to Individual Visualizations. *Digit. Gov. Res. Pract.* 3. <https://doi.org/10.1145/3558099>
- Chou, J.-R., Hsiao, S.-W., 2007. A usability study on human–computer interface for middle-aged learners. *Comput. Hum. Behav.* 23, 2040–2063. <https://doi.org/10.1016/j.chb.2006.02.011>
- Chua, U.C., Santiago, K.L., Ona, I.B.M., Peña, R.M.N., Marasigan, G.Z.S., Reyes, P.G.A.D., Samson, B.P.V., 2020. From Access to Effective Use: Open Data Portals for Everyday Citizens, in: *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 61–64. <https://doi.org/10.1145/3391203.3391219>
- Ganapati, S., Reddick, C.G., 2012. Open e-government in U.S. state governments: Survey evidence from Chief Information Officers. *Gov. Inf. Q.* 29, 115–122. <https://doi.org/10.1016/j.giq.2011.09.006>
- Ho, B., Jakli, J., 2010. ASSESSING BENEFITS OF BUSINESS INTELLIGENCE SYSTEMS – A CASE STUDY 15.
- Jaeger, P.T., Bertot, J.C., 2010. Transparency and technological change: Ensuring equal and sustained public access to government information. *Gov. Inf. Q.* 27, 371–376. <https://doi.org/10.1016/j.giq.2010.05.003>
- Jansen, H., n.d. *The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods*.
- Janssen, M., Charalabidis, Y., Zuiderwijk, A., 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Inf. Syst. Manag.* 29, 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Kitchenham, B., 2004. *Procedures for Performing Systematic Reviews*.
- Klievink, B., Neuroni, A., Fraefel, M., Zuiderwijk, A., 2017. Digital Strategies in Action: a Comparative Analysis of National Data Infrastructure Development, in: *Proceedings of the 18th Annual International Conference on Digital Government Research*. Presented at the dg.o '17: 18th Annual International Conference on Digital Government Research, ACM, Staten Island NY USA, pp. 129–138. <https://doi.org/10.1145/3085228.3085270>

- Krismawati, D., Hidayanto, A.N., 2021. The User Engagement of Open Data Portal, in: 2021 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2021. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICACSIS53237.2021.9631357>
- Layne, K., Lee, J., 2001. Developing fully functional E-government: A four stage model. *Gov. Inf. Q.* 18, 122–136. [https://doi.org/10.1016/S0740-624X\(01\)00066-1](https://doi.org/10.1016/S0740-624X(01)00066-1)
- Lee, Y., Kozar, K.A., 2012. Understanding of website usability: Specifying and measuring constructs and their relationships. *Decis. Support Syst.* 52, 450–463. <https://doi.org/10.1016/j.dss.2011.10.004>
- Linders, D., Wilson, S.C., 2011. What is open government?: one year after the directive, in: Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times. Presented at the dg.o '11: The 12th Annual International Conference of Digital Government Research, ACM, College Park Maryland USA, pp. 262–271. <https://doi.org/10.1145/2037556.2037599>
- Linked Data - Design Issues [WWW Document], n.d. URL <https://www.w3.org/DesignIssues/LinkedData.html> (accessed 6.26.23).
- Lněnička, M., Machova, R., Volejníková, J., Linhartová, V., Knezackova, R., Hub, M., 2021. Enhancing transparency through open government data: the case of data portals and their features and capabilities. *Online Inf. Rev.* 45, 1021–1038. <https://doi.org/10.1108/OIR-05-2020-0204>
- Lnenicka, M., Nikiforova, A., 2021. Transparency-by-design: What is the role of open data portals? *Telemat. Inform.* 61, 101605. <https://doi.org/10.1016/j.tele.2021.101605>
- Lourenço, R.P., 2015. An analysis of open government portals: A perspective of transparency for accountability. *Gov. Inf. Q.* 32, 323–332. <https://doi.org/10.1016/j.giq.2015.05.006>
- Máchová, R., Hub, M., Lnenicka, M., 2018. Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. *Aslib J. Inf. Manag.* 70, 252–268. <https://doi.org/10.1108/AJIM-02-2018-0026>
- McDermott, P., 2010. Building open government. *Gov. Inf. Q.* 27, 401–413. <https://doi.org/10.1016/j.giq.2010.07.002>
- Mellouli, S., Luna-Reyes, L.F., Zhang, J., 2014. Smart government, citizen participation and open data. *Inf. Polity* 19, 1–4. <https://doi.org/10.3233/IP-140334>
- Moradi, M., Mazoochi, M., Ahmadi, M., 2022. A Comprehensive Method for Improving the Quality of Open Government Data and Increasing Citizens' Willingness to Use Data by Analyzing the Complex System of Citizens and Organizations. *Complexity* 2022. <https://doi.org/10.1155/2022/5876035>
- Mutambik, I., Almuqrin, A., Lee, J., Zhang, J.Z., Alomran, A., Omar, T., Floos, A., Homadi, A., 2021. Usability of the G7 open government data portals and lessons learned. *Sustain. Switz.* 13. <https://doi.org/10.3390/su132413740>
- Nikiforova, A., McBride, K., 2021. Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telemat. Inform.* 58. <https://doi.org/10.1016/j.tele.2020.101539>

- Noveck B.S., 2016. Is Open Data the Death of FOIA. *Yale L.J.F.* 126, 273. - Beth Simone Noveck, Is Open Data the Death of FOIA?, 126 *Yale L.J. F.* 273 (2016). <http://www.yalelawjournal.org/forum/is-open-data-the-death-of-foia>
- Paige, B., Freund, L., 2019. A Case Study of New England Open Data Portals. *JeDEM* 11, 2019.
- Park, S., Gil-Garcia, J.R., 2022. Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap. *Gov. Inf. Q.* 39. <https://doi.org/10.1016/j.giq.2020.101456>
- Peppers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S., 2007. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* 24, 45–77.
- Sabri, N.A.M., Emran, N.A., Harum, N., 2019. Government open data portals: A measurement of data veracity coverage. *Int. J. Innov. Technol. Explor. Eng.* 8, 1975–1983. <https://doi.org/10.35940/ijitee.L2908.1081219>
- Safarov, I., Meijer, A., Grimmelikhuijsen, S., 2017. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Inf. Polity* 22, 1–24. <https://doi.org/10.3233/IP-160012>
- Saxena, S., Alexopoulos, C., 2022. Open Government Data (OGD) Research: Retrospection and Forethoughts. *Mak. Contemp. Maldives Isol. Dictatorship Democr.* 3, 1–19. <https://doi.org/10.52823/idws4806>
- Schauppenlehner, T., Muhar, A., 2018. Theoretical availability versus practical accessibility: The critical role of metadata management in open data portals. *Sustain. Switz.* 10. <https://doi.org/10.3390/su10020545>
- Schrock, A.R., 2016. Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media Soc.* 18, 581–599. <https://doi.org/10.1177/1461444816629469>
- Simonofski, A., Zuiderwijk, A., Clarinval, A., Hammedi, W., 2022. Tailoring open government data portals for lay citizens: A gamification theory approach. *Int. J. Inf. Manag.* 65. <https://doi.org/10.1016/j.ijinfomgt.2022.102511>
- Thorsby, J., Stowers, G.N.L., Wolslegel, K., Tumbuan, E., 2017. Understanding the content and features of open data portals in American cities. *Gov. Inf. Q.* 34, 53–61. <https://doi.org/10.1016/j.giq.2016.07.001>
- Venable, J., Pries-Heje, J., Baskerville, R., 2016. FEDS: a Framework for Evaluation in Design Science Research. *Eur. J. Inf. Syst.* 25, 77–89. <https://doi.org/10.1057/ejis.2014.36>
- Wang, D., Richards, D., Bilgin, A.A., Chen, C., 2021. Advancing open government data portals: a comparative usability evaluation study. *Libr. Hi Tech.* <https://doi.org/10.1108/LHT-10-2020-0250>
- Zhu, X., Freeman, M.A., 2019. An evaluation of U.S. municipal open data portals: A user interaction framework. *J. Assoc. Inf. Sci. Technol.* 70, 27–37. <https://doi.org/10.1002/asi.24081>
- Zuiderwijk, A., Janssen, M., Parnia, A., n.d. The complementarity of open data infrastructures: an analysis of functionalities.

Appendix A – Evaluation

A.1 Interview Questions/Tasks

SIFIDE

Table A.1. 1 - Tasks related to the SIFIDE program

<i>Question/Task</i>	<i>Answer/Action</i>
Click on the button to see the general and summarized data of the program.	Click on button “Sumário” (“Summary”).
In 2020, what was the credit approval rate?	90,7%
See all the data provided in this page only for the year 2021 and the Lisbon district.	Go to the Filter pane, select Year “2021” and District “Lisboa” (“Lisbon”).
Go to the next page.	Click on button “SIFIDE”.
In 2021, what was the sector that obtained a higher declared investment?	Sector “Informação e comunicação” (“Information and communication”)
In 2015, what was the dimension of the majority of candidate companies?	Dimension “Médias Empresas” (“Medium Companies”).
See data for Approved applications.	Click on button “Candidaturas Aprovadas” (“Approved Applications”).
What is the district with more approved applications?	District “Lisboa” (“Lisbon”).
Go to the next page.	Click on button “Link”.

PT2020

Table A.1. 2 - Tasks related to the PT2020 program

<i>Question/Task</i>	<i>Answer/Action</i>
Click on the button to see the general and summarized data of the program.	Click on button “Sumário” (“Summary”).
How many cancelled applications were there in 2020?	41.
Go to the next page.	Click on button “PT2020”.
What is the percentage of universities that participated in received applications?	21,8%.
What is the district that obtained more approved investment? How much?	District “Porto”. 525,6 M€.
In Porto, what is the sector that obtained more approved investment?	Sector “TIC”.

HEurope

Table A.1. 3 - Tasks related to the HEurope program

<i>Question/Task</i>	<i>Answer/Action</i>
Click on the button to see the general and summarized data of the program.	Click on button “Sumário” (“Summary”).
In the HEurope program, what is the success rate of Portuguese proposals?	19,9%.
See the evolution of the number of approved projects from 2014 to 2023.	Click on arrow icon. This will take the user to a graph displaying the evolution of approved projects from 2014 to 2023.
Go the next page.	Click on button “HEuropa”.
In 2022, how much financing was obtained for research centers?	107,96 M€.
Analyze the program per theme.	Click on button “Análise por Tema” (“ <i>Analysis per Theme</i> ”)
At the moment, the data is presented by pilar, which is a group of themes. See the graph “Propostas vs. Projetos” (“ <i>Proposals vs. Projects</i> ”) by theme.	Go to the graph’s Action bar, click on the icon with the 2 arrows facing down to go to the next level of hierarchy.
See only the themes of Pilar 2.	In the Action bar, click on the 1 arrow facing up to aggregate themes by pilar, then click on the 1 arrow facing down to activate drill-down feature, and then click on column Pilar 2.
How many big companies participated in approved projects with the theme security?	2.
Aggregate themes by pilars.	In the Action bar, click on the 1 arrow facing up to aggregate.

A.2 Questionnaire

Power BI User Interface

Age *

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 or over

Gender *

- Male
- Female

Experience with Power BI *

- Beginner
- Intermediate
- Advanced
- Never used it

How would you rate the overall intuitiveness of the Power BI interface? *

	1	2	3	4	5	
Not intuitive at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very intuitive

How easy was it to navigate through the visualizations in Power BI? *

	1	2	3	4	5	
Very difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very easy

Did you find any specific features confusing or difficult to understand? *

Yes

No

If Yes, specify

A sua resposta _____

Rate the user-friendliness of the Power BI visualizations. *

Not user-friendly at all 1 2 3 4 5 Very user-friendly

Did you feel that all the features of the visualizations were well explained? *

Yes

No

Partially

What were the main difficulties you encountered while interacting with the Power BI visualizations? *

A sua resposta _____

What did you like most about the visualizations? *

A sua resposta _____

If you could suggest one improvement to the Power BI visualizations, what would it be?

A sua resposta

What are your overall thoughts on Power BI as a tool for presenting data, based on your experience? *

	1	2	3	4	5	
Very negative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very positive

Are there any additional comments or feedback you would like to provide regarding your experience?

A sua resposta

Thank you for participating!