THESIS / THÈSE

DOCTOR OF SCIENCES

Open Government Data

Empowering Reuse through Collaboration, Data Quality, and Data Storytelling

Chokki, Abiola Paterne

Award date:
2023

Awarding institution:
University of Namur

Open Government Data: Empowering Reuse through Collaboration, Data Quality, and Data Storytelling

Abiola Paterne Chokki

Jury

Prof. Wim Vanhoof
University of Namur, Belgium

Prof. Adegboyega Ojo
Carleton University, Canada

Prof. Anastasija Nikiforova
University of Tartu, Estonia

Prof. Anthony Simonofski
University of Namur, Belgium

Prof. Xavier Devroey
University of Namur, Belgium

Prof. Bruno Dumas
University of Namur, Belgium

Prof. Benoît Frénay
University of Namur, Belgium

Prof. Benoît Vanderose
University of Namur, Belgium

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the subject of Computer Science

Supervised by Prof. Benoît Vanderose and Prof. Benoît Frénay

University of Namur

PReCISE Research Center

UNIVERSITÉ DE NAMUR
FACULTÉ D'INFORMATIQUE

Cover design: ©Presses universitaires de Namur

Rue Grandgagnage 19
B - 5000 Namur (Belgium)
pun@unamur.be - www.pun.be

Registration of copyright: D/2023/1881/20
ISBN : 978-2-39029-184-8
Printed in Belgium.

ABSTRACT

Open Government Data (OGD), spanning various government activities, has been made publicly accessible through online portals in the aim of increasing transparency and accountability. Although OGD offers myriad benefits, the full potential is only realized when the data is effectively reused, a facet hindered by multiple barriers. This thesis focuses on these challenges, targeting three specific barriers to develop artifacts and practical solutions that aid OGD stakeholders, primarily citizens, infomediaries, and OGD publishers.

The first barrier, concerning the lack of communication and collaboration among OGD stakeholders, was addressed by exploring communication methods and socio-collaborative needs. Following a literature review and an online survey, we devised recommendations for effective communication channels. We also identified the requirements for an application to raise user awareness about OGD, and then developed the application integrating those requirements. Through testing, we confirmed that the application was successful in enhancing user awareness about OGD. Further, to bridge the gap between stakeholders, we identified requirements for an application that could address the socio-collaborative impediments, and upon its development, we confirmed its effectiveness in ameliorating the technical issues surrounding collaboration.

The second barrier tackled was the lack of (meta)data quality on OGD portals, a problem that can severely affect the reliability of services developed based on these data. We initiated our approach by identifying quality dimensions to assess data and metadata quality. Subsequently, we developed a conversational agent integrating these dimensions and user preferences, which was evaluated and confirmed to be effective in providing a comprehensive assessment of (meta)data quality.

The third barrier confronted was the difficulty for users to turn data into stories, a process crucial for making informed decisions. The approach to this challenge involved three interactive methods. An analysis of the usage statistics of the Namur OGD portal and an online survey informed us of user preferences in data types. Interviews gave us insights into the most suitable data visualization methods. Following this, a literature review enabled us to identify key dashboard design principles and the requirements for data storytelling tools. Eventually, we developed a prototype integrating these requirements, which was evaluated to ascertain its usefulness in aiding users to create engaging data narratives regardless of their technical proficiency.

In summary, this thesis primarily focuses on tackling the challenges related to the reuse of OGD by outlining recommendations and practical prerequisites for feasible solutions. It further proposes effective and user-friendly solutions to these challenges through the development of tools designed to facilitate communication and collaboration, improve data quality, and promote data storytelling. These contributions not only provide a robust theoretical foundation for future research but also practical solutions that can be leveraged by OGD stakeholders to optimize data reuse.

Keywords: open government data, reuse, collaboration, data quality, data story-

telling

RÉSUMÉ

Les données gouvernementales ouvertes (DGO), couvrant diverses activités gouvernementales, ont été rendues accessibles au public par le biais de portails en ligne dans le but d'augmenter la transparence et la responsabilité gouvernementale. Bien que les DGO offrent une myriade d'avantages, le plein potentiel n'est réalisé que lorsque les données sont efficacement réutilisées, une facette entravée par de multiples obstacles. Cette thèse se concentre sur ces défis, en ciblant trois obstacles spécifiques pour développer des artefacts et des solutions pratiques qui aident les parties prenantes des DGO, principalement les citoyens, les infomédiaires et les éditeurs de DGO.

Le premier obstacle, concernant le manque de communication et de collaboration entre les parties prenantes des DGO, a été abordé en explorant les méthodes de communication et les besoins socio-collaboratifs. Suite à une revue de la littérature et à une enquête en ligne, nous avons élaboré des recommandations pour des canaux de communication efficaces. Nous avons également identifié les exigences pour une application visant à sensibiliser les utilisateurs aux DGO, puis nous avons développé l'application en intégrant ces exigences. Par le biais de tests, nous avons confirmé que l'application réussissait à améliorer la sensibilisation des utilisateurs aux DGO. De plus, pour combler le fossé entre les parties prenantes, nous avons identifié les exigences pour une application qui pourrait aborder les entraves socio-collaboratives, et après son développement, nous avons confirmé son efficacité à améliorer les problèmes techniques entourant la collaboration.

Le deuxième obstacle abordé était le manque de qualité des données sur les portails DGO, un problème qui peut affecter sévèrement la fiabilité des services développés sur la base de ces données. Nous avons initié notre approche en identifiant les dimensions de qualité pour évaluer la qualité des données et des métadonnées. Par la suite, nous avons développé un agent conversationnel intégrant ces dimensions et les préférences des utilisateurs, qui a été évalué et confirmé comme étant efficace pour fournir une évaluation complète de la qualité des (méta)données.

Le troisième obstacle affronté était la difficulté pour les utilisateurs de transformer les données en tableaux de bord facilement compréhensibles, un processus crucial pour prendre des décisions éclairées. L'approche de ce défi a impliqué trois méthodes interactives. Une analyse des statistiques d'utilisation du portail DGO de Namur et une enquête en ligne nous ont informés sur les préférences des utilisateurs en matière de types de données. Les entretiens nous ont donné des informations sur les méthodes de visualisation de données les plus appropriées. Suite à cela, une revue de la littérature nous a permis d'identifier les principes clés de la conception des tableaux de bord et les exigences pour les outils de narration de données. Finalement, nous avons développé un prototype intégrant ces exigences, qui a été évalué pour vérifier son utilité à aider les utilisateurs à créer des récits de données engageants, quelle que soit leur compétence technique.

En résumé, cette thèse se concentre principalement sur la résolution des défis liés à la réutilisation des DGO en définissant des recommandations et des prérequis pratiques pour des solutions réalisables. Elle propose également des solutions efficaces et conviviales à ces défis grâce au développement d'outils conçus pour faciliter la communication et la collaboration, améliorer la qualité des données et promouvoir la narration de données. Ces contributions fournissent non seulement une base théorique solide pour les futures recherches, mais aussi des solutions pratiques qui peuvent être utilisées par les parties prenantes des DGO pour optimiser la réutilisation des données.

Mots clés : données gouvernementales ouvertes, réutilisation, collaboration, qualité des données, narration de données

ACKNOWLEDGEMENTS

The path to completing this PhD thesis was lengthy and filled with many twists and turns. Thanks to the assistance, encouragement, and advice from numerous individuals, I managed to reach the end of this journey:

First and foremost, I express my deepest appreciation to my supervisors, Benoît F. and Benoît V. for selecting me as their PhD student under the CERUNA scholarship. Your constant encouragement, and thorough review of my work have made this academic journey fulfilling and enriching. I'm particularly thankful for the freedom you gave me to explore my own thoughts. This trust helped me to develop both as a person and in my studies, encouraging me to take charge of my own research.

Anthony and Antoine, your multifaceted roles in my thesis journey have been invaluable, and I cannot overstate how grateful I am for your consistent and unstinting support. From the very onset, you've been my guiding lights in this research endeavor, always ready to brainstorm, to offer insightful feedback, and to engage in enlightening discussions. In every way imaginable, you've been instrumental to my thesis journey. Your unwavering commitment, insightful guidance, and genuine friendship have made a profound impact on me, both academically and personally.

I extend my heartfelt thanks to the esteemed members of my supervisory committee - Wim, Anthony, Xavier, and Bruno - and external examiners, Adegboyega, and Anastasija. I'm grateful that you accepted the responsibility of being part of my PhD committee. Your critiques and input on this thesis have been of enormous value and have contributed significantly to its refinement.

I am deeply indebted to my co-authors and collaborators - Anthony S., Antoine, Rabeb, Anthony C., Stuti, and Charalampos. Your support, expressed in myriad ways, including ideation, manuscript composition, and meticulous proofreading, enriched my doctoral research experience.

I am especially appreciative of Ahmed, Antoine, Edilton, Geraldin, Jerome, Nicolas, Pamphile, Rabeb, Sacha, and Valentin. Your insightful contributions to user testing sessions, your thoughtful responses to surveys, and your invigorating discussions played a crucial role in enhancing this research. Your enthusiasm and dedication in participating in these activities have been instrumental in shaping and honing the ideas and concepts embodied in this thesis.

My sincerest gratitude goes to the University of Namur for providing an inviting environment conducive to learning for international students, and for the financial support that underpinned my research.

A special word of thanks to my family - my father Grégoire, my late mother Irene, my sisters Eunice, Gwladys, Lisa and my brothers Ulrich, Patrick. Your love and unwavering support provided a strong foundation that held me steady throughout this journey.

Above all, my deepest gratitude is reserved for my wife, Noutsara, and our precious daughter, Emma. Noutsara, your decision to join me in Belgium, stands as one of my greatest blessings. Your constant love, support, and assistance were the pillars that upheld me throughout this journey. To my precious daughter, you have been my guiding light throughout this journey. Even on days when I couldn't spend

as much time with you as I wanted, your beautiful smile and endless love were my biggest inspiration and strength.

PUBLICATIONS

The contributions outlined in this thesis have been published in peer-reviewed journals and conferences. This section provides a comprehensive list of these publications and elucidates their specific contributions to the overall objectives of this thesis.

Improving Communication and Collaboration among OGD Stakeholders

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Increasing Awareness and Usefulness of Open Government Data: An Empirical Analysis of Communication Methods. In **16th International Conference on Research Challenges in Information Science**, 2022f

- This paper examines the preferences of users regarding the channels utilized for promoting OGD. Through a combination of a literature review and an online survey, effective communication methods were identified and the results were aggregated to offer recommendations to governments.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open government data awareness: eliciting citizens' requirements for application design. **Transforming Government: People, Process and Policy**, 2022e. doi: https://doi.org/10.1108/TG-04-2022-0057

- Building upon the previous paper (Chokki et al., 2022f), this study expands its scope by identifying and validating the 11 requirements for infomediaries to develop a functional tool aimed at increasing user awareness of OGD. This is achieved through a comprehensive approach that includes a literature review, interviews, and user testing.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, Benoît Frénay, and Benoît Vanderose. Fostering Interaction between Open Government Data Stakeholders: An Exchange Platform for Citizens, Developers and Publishers. In **Electronic Government**, 2022c

- In this paper, a set of 16 requirements is identified for the development of a tool that enhances communication and collaboration among stakeholders involved in OGD. Additionally, practical contributions are made by presenting a use case diagram that outlines the necessary features to fulfill these requirements, along with the implementation of a functional tool that incorporates them. These outcomes are the result of a thorough investigation that encompasses a literature review, interviews, and user testing.

Assessing (Meta)data Quality of OGD

Abiola Paterne Chokki, Charalampos Alexopoulos, Stuti Saxena, Benoît Frénay, Benoît Vanderose, and Mohsan Ali. Metadata quality matters in Open Government Data (OGD) evaluation! An empirical investigation of OGD portals of the GCC constituents. **Transforming Government: People, Process and Policy**, pages 1–21, 2022a. doi: 10.1108/TG-09-2022-0118

- This paper undertakes an evaluation of the metadata quality within the GCC countries, aiming to identify the best-performing metric for each country while also investigating potential disparities in metadata quality across the region. To accomplish this, we adopt a quantitative research approach comprising a series of steps: defining the metrics for assessing metadata quality, gathering datasets from the portals of GCC countries, calculating the metadata quality of these portals, and conducting a comparative analysis of metadata quality among the GCC countries.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, and Benoît Vanderose. Evaluating a Conversational Agent for Open Government Data Quality Assessment. In **29th Americas Conference on Information Systems**, pages 1–10, 2023

- This paper presents a comprehensive compilation of dimensions and metrics (sub-dimensions) that are essential for evaluating the quality of metadata and data content independently for each open government dataset. Furthermore, it offers a collection of features that are necessary for the development of a quality assessment tool for (meta)data. These findings are the outcome of an extensive investigation that encompasses a literature review and user testing.

Facilitating Data Storytelling with OGD

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open Government Data for Non-expert Citizens: Understanding Content and Visualizations' Expectations. **Research Challenges in Information Science**, 415 LNBIP: 602–608, 2021. ISSN 18651356. doi: 10.1007/978-3-030-75018-3_42

- This paper presents the datasets that are of interest to users and the visualizations that they prefer on OGD portals. These insights are valuable for infomediaries who are searching for suitable datasets for their applications, as well as for OGD publishers who need to determine which datasets to make available on their portals. Furthermore, these findings offer guidance on the types of visualizations that enhance user comprehension. These findings were obtained through a multi-method approach, which included an analysis of the usage statistics of the OGD portal of Namur, a complementary online survey to understand the dataset needs of end-users, and interviews to gather their opinions on correctly designed and well-presented visualizations of datasets.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Engaging Citizens with Open Government Data: The Value of Dashboards compared to Individual Visualizations. **Digital Government: Research and Practice**, 2022d. doi: https://doi.org/10.1145/3558099

- This paper presents a compilation of design principles for dashboards in the OGD context, aiming to enhance their usability and user understanding. These principles are identified through a systematic literature review. Furthermore, the paper validates the hypothesis that the use of well-designed dashboards can promote citizen engagement with OGD. This validation is achieved through the implementation of these design principles in a usable tool and the assessment of their impact on citizen engagement using the tool as a proxy.

Abiola Paterne Chokki, Benoit Frenay, and Benoît Vanderose. Open Data Explorer: An End-to-end Tool for Data Storytelling using Open Data. In **Americas Conference**

on Information Systems 2022 Proceedings. 6, 2022b

- This paper introduces a collection of essential features required for designing a generic and end-to-end data storytelling tool that enables users with varying levels of data manipulation skills to effortlessly transform OGD into engaging narratives. The identification of these features is based on a comprehensive literature review and interviews conducted with potential users. To validate these features, a functional tool named ODE (Open Data Explorer) is utilized as a proxy, incorporating the identified features.

Abiola Paterne Chokki and Benoît Vanderose. From Conventional Open Government Data Portals to Storytelling Portals : The StoryOGD Prototype. In **24th Annual International Conference on Digital Government Research**, pages 1–2, 2023

- This paper introduces StoryOGD, an enhanced iteration of the ODE tool, which is explicitly tailored to aid publishers in effectively presenting their portals in a user-friendly fashion. Additionally, it outlines the requirements necessary for implementing such a tool, which were identified through an exploratory study of websites that aggregated open datasets and presented them in a user-friendly manner.

CONTENTS

LIST OF FIGURES

List of Tables

CHAPTER 1

INTRODUCTION

Aiming to increase transparency and accountability, many governments at different levels, such as national, regional, and local, are providing public access to their data through online portals. These data are referred to as Open Government Data (OGD), covering a broad spectrum of government activities, including political boundaries, education, health, finance, and mobility (Graves and Hendler, 2013). Major cities like London in 2010, Paris and New York in 2011, and Namur (Belgium) in 2018 have established online portals for sharing OGD. The benefits of OGD include increased transparency and accountability of government operations, enhanced citizen engagement and participation, improved decision-making, and economic growth through innovation and entrepreneurship (Bertot et al., 2010; Davies, 2010; Johnson and Robinson, 2014; Janssen et al., 2017; Purwanto et al., 2020).

One significant prerequisite to unlock the aforementioned benefits of OGD is its reuse (Attard et al., 2015). By making government data accessible and usable, OGD enables citizens, researchers, businesses, and organizations to leverage the data for a wide range of purposes (Gonzalez-Zapata and Heeks, 2015; Safarov et al., 2017; Graves and Hendler, 2014). For example, OGD can be used to develop new services and applications that improve the efficiency of public services, address societal challenges, and create new economic opportunities. Reusing OGD can also enhance accountability and transparency by enabling citizens to monitor government actions and hold officials accountable for their decisions. Moreover, the reuse of OGD can facilitate collaboration among stakeholders, including the public sector, private sector, and civil society organizations, to address common challenges and achieve shared objectives. By sharing data, stakeholders can build on each other's expertise and knowledge to create innovative solutions and generate new insights. Therefore, the potential for reuse of OGD is a critical advantage that contributes to the broader goal of promoting transparency, accountability, and citizen participation in government operations.

Despite the availability of OGD and the potential benefits of their reuse, only few people are using them (Toots et al., 2017; Abdelrahman, 2021). This can be explained by the many barriers impeding the OGD reuse process. These barriers can be classified into four categories: technical, social, legal, and economic (Janssen et al., 2012; Crusoe and Melin, 2018; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013). Technical barriers include difficulties in accessing and processing the data, while social barriers are related to a lack of awareness or interest in using the data. Legal barriers refer to restrictions on the reuse of OGD, such as copyright or privacy concerns. Economic barriers arise from the cost of accessing or processing the data. While all these barriers are important, this thesis focuses on technical and social barriers, and more specifically on three main barriers. Therefore, this thesis aims to develop artifacts and tools that can assist OGD stakeholders in overcoming these barriers. The selection of three specific barriers was based on a literature review and their relevance to the context of Wallonia (Belgium).

The first barrier is the **lack of communication and collaboration among stakeholders involved in OGD reuse**. This thesis focus on three pivotal stakeholders involved in the OGD reuse process: citizens, infomediaries, and OGD publishers. Infomediaries, also known as OGD intermediaries, are a distinct category of OGD users who manipulate data, extracting, aggregating, and transforming it into a format that is readily usable, beneficial, and of value to the general public (Shaharudin et al., 2023; Johnson and Greene, 2017). OGD publishers are vital in this process as they are responsible for releasing data to enable its reuse. Each of these stakeholder groups plays a crucial role in maximizing the potential benefits of OGD. While citizens can benefit from OGD, infomediaries leverage OGD to create innovative services that deliver value to citizens, and OGD publishers release data to enable infomediaries to create these services. The term “users” refers to both citizens and infomediaries, as both may use OGD. However, these stakeholders often operate in silos and lack effective communication and collaboration, leading to a lack of awareness and understanding about the potential benefits of OGD (Toots et al., 2017; Abdelrahman, 2021). For instance, citizens may not be aware of the availability of OGD, while infomediaries may face challenges in accessing the necessary data or processing it. Similarly, OGD publishers may not have a clear understanding of the needs of users and may not provide data in useful formats, further hindering OGD reuse.

To tackle the first barrier, the thesis explores two areas: communication methods needs and socio-collaborative needs. Regarding communication methods needs, the goal is to understand users’ preferences in terms of channels used to promote OGD. A literature review and online survey were conducted to identify effective communication methods, and the results were used to provide recommendations to governments. The results indicated that users preferred practical applications, leading the researchers to identify the requirements for a mobile application to improve user awareness of OGD. The application was developed and evaluated through user testing, which confirmed its usefulness in enhancing user awareness of OGD. As for the socio-collaborative needs, the objective is to identify impediments associated with the lack of collaboration between OGD stakeholders and suggest technical requirements and solution to mitigate these challenges. Initially, we identified the requirements for an application that would address the technical aspects of the lack of collaboration between stakeholders through literature review and stakeholder interviews. After developing the application, we used it as a proxy to confirm the usefulness of the proposed requirements in addressing the technical challenges of the lack of collaboration between OGD stakeholders.

The second barrier is the **insufficient (meta)data quality on OGD portals**. Data quality on OGD portals refers to the accuracy, completeness, consistency, timeliness, and relevance of the data made available to the public (Reiche et al., 2014; Wang and Strong, 1996). Poor data quality can have severe consequences for both OGD users and publishers (Kubler et al., 2018). For OGD users, poor data quality can lead to incorrect or incomplete analysis, decision-making, and research. For OGD publishers, poor data quality can result in reputation damage, decreased public trust, and decreased demand for data. Poor data quality can also lead to an increased cost of data management, as organizations may need to spend more resources to correct and improve data quality. Moreover, poor data quality can also reduce the potential benefits of open data, as poor data quality can limit the ability of data users to derive insights, innovate, and create value from the data (Crusoe et al., 2019; Kubler et al., 2018). Therefore, ensuring data quality is essential for achieving the potential benefits of open data.

As the focus of this thesis is on reuse, we emphasize the importance of assessing the data quality of individual datasets before using it in service development. To achieve this, we first identified quality dimensions for assessing both data and metadata quality through a literature review. We then developed a conversational agent that integrated these dimensions and allowed for user preferences to be included in the quality assessment. A usability evaluation with users demonstrated the effectiveness of this approach in obtaining overall and detailed (meta)data quality, as well as its ease of use. By doing so, users can make informed decisions about the quality of the data they intend to use.

The third barrier is the **difficulty for users to turn data into stories**. Once the users are aware of the existence and the usefulness of OGD and informed about the quality of these datasets, there remains the question of how they can turn these datasets into stories. The ability to transform data into meaningful stories is crucial for making informed decisions. This process is called data storytelling and consists of using data visualizations, charts, and graphs, combined with text and images, to communicate a story to an audience (Brolcháin et al., 2017). Data storytelling can help users understand complex data sets by making them more accessible and engaging. Moreover, data storytelling can help to promote transparency and accountability by making it easier for users to understand and interpret government data. When government data is presented in a clear and compelling way, it can help to increase public trust in government and encourage citizen engagement. In this way, data storytelling can help to support open government initiatives and promote a more participatory democracy and yet it can be a difficult task for many users, especially those who do not possess specialized technical skills, as it requires not only an understanding of the data but also the ability to analyze it, identify patterns, create visual representations, and weave a cohesive and compelling narrative (Graves and Bustos-Jiménez, 2015; Aanderud et al., 2020; Brolcháin et al., 2017). The process requires a combination of skills that span both the technical (e.g., data analysis, statistics) and the creative (e.g., narrative construction, visual design), which may not be within the skill set of all users. Consequently, many users might find the process daunting and complex, limiting their ability to fully harness the potential of the available data.

To address this third barrier, we have employed three interactive approaches. Firstly, we conducted an analysis of the usage statistics of the Namur OGD portal, coupled with an online survey to identify the specific requirements of users regarding the data they would like to see published on portals. This enabled us to gather valuable insights and generate recommendations for OGD publishers and infomediaries regarding the types of data that should be made available in their services. Additionally, we conducted interviews to determine the most suitable visualization methods for different categories of data, taking into account user ratings. This allowed us to gain a deeper understanding of user preferences and needs in terms of data presentation and visualization on portals. Secondly, we conducted a thorough literature review to determine the key principles that should be incorporated into dashboards to enhance their comprehensibility and usability. By implementing these principles, we aimed to create user-friendly dashboards that facilitate data understanding and exploration. Once we obtained insights from these initial two steps, we proceeded to identify the requirements for data storytelling tools. These requirements were gathered through a comprehensive literature review, focusing on the features and functionalities that would enable both technical and non-technical users to effortlessly create engaging data stories. Subsequently, we developed a prototype that integrated these requirements and conducted user testing to evaluate its ease of use and usefulness in helping users, regardless of their technical skills, transform data into compelling narratives. By following this approach, we aimed to bridge the gap between raw data and meaningful stories, providing users with the tools and capabilities to communicate their insights effectively. Through user-centered design and iterative refinement, we strived to develop a data storytelling solution that empowered users to harness the potential of OGD and share their findings with others. The goal was to enable both technical experts and non-technical individuals to leverage data to tell impactful stories that could drive informed decision-making, foster transparency, and promote engagement with government initiatives. By facilitating the creation and dissemination of data-driven narratives, we sought to enhance the accessibility and understanding of OGD, ultimately empowering users to extract valuable insights.

1.1 Outline

The remainder of this thesis is divided into four parts to organize the content effectively.

Part I focuses on the relevance of the research. Chapter 2 discusses the relevance of the research to the field by providing an overview of the Open Government Data concept, with a particular focus on its deployment in Wallonia, Belgium. Additionally, it enumerates the different OGD stakeholders under consideration in this thesis and elucidates how the three barriers were identified through a literature review.

Part II presents the research design used to address the barriers. In Chapter 3, the identified barriers and research gaps from Part I are mapped to the research questions. Chapter 4 provides a detailed explanation of the overall research methodology as well as its underlying methodologies.

Part III consists of dedicated chapters that present the contributions made for each research question. Chapter 5 addresses the first barrier by investigating users' preferred communication methods when accessing and utilizing OGD, leading to recommendations for governments on effective OGD communication strategies. Additionally, it identifies users' requirements for application design to enhance OGD awareness. Furthermore, Chapter 5 identifies and validates technical requirements through user testing and interviews, aiming to mitigate challenges related to collaboration among OGD stakeholders. In Chapter 6, the focus shifts to the second barrier, where a literature review is conducted to identify dimensions and metrics that infomediaries can employ to assess (meta)data quality on OGD portals at dataset level. The chapter also outlines the requirements for tools that integrate these dimensions and allow for user preferences to be considered in the quality assessment. Chapter 7 addresses the third barrier. Firstly, it explores users' preferences for data contents on portals and data visualizations when engaging with OGD, providing recommendations to infomediaries regarding data needs and effective visualization techniques for user understanding. Then, it discusses the design principles that infomediaries should consider when developing data storytelling dashboards with OGD. Lastly, Chapter 7 analyzes and validates the requirements of data storytelling tools through user testing and interviews, ensuring the creation of engaging narratives using OGD for both technical and non-technical users. Each chapter includes specific methodological details that supplement Chapter 4.

Part IV reflects on the results presented in the previous part. Chapter 8 serves as a comprehensive summary, discussing the contributions made, their implications for research, and how they align with the research questions. The chapter further examines the practical impact of these contributions and acknowledges the general limitations of the thesis. Moreover, it identifies future research directions based on the identified limitations and presents additional avenues for future work, which were not derived solely from the limitations. Chapter 9 closes this thesis with brief concluding remarks.

Part I

Background and Problem Statement

CHAPTER 2

RESEARCH BACKGROUND

In this chapter, the origin and definition of open government data are initially explored, along with its maturity in the Wallonia, Belgium context. The chapter then delves into the core vision of this thesis, which is the reuse of OGD, particularly examining its benefits, and the stakeholders involved. Furthermore, three barriers that hinder OGD reuse are extensively discussed. These barriers include the absence of effective communication and collaboration among stakeholders engaged in OGD reuse, the insufficient (meta)data quality on OGD portals, and the challenges faced by users in transforming data into meaningful narratives. These three barriers serve as the foundation for the problem statement addressed in this thesis.

2.1 Open Government Data (OGD)

Open data is a core principle that centers on the idea of providing data to the public in a machine-readable format, free from constraints on usage or redistribution (Attard et al., 2015). The goal of open data initiatives is to unleash the full potential of data by eliminating barriers such as restrictive licenses or limited accessibility. This facilitates its broad utilization for societal, economic, and research objectives.

The concept of open data government emerged as an expansion of the broader open data movement, aiming to proactively disclose government-held data and make it accessible to the public. These datasets encompass a wide range of government activities, spanning areas such as political boundaries, education, health, environment, demographics, finance, transportation, and more (Graves and Hendler, 2013). The origins of open data government initiatives can be traced back to the early 2000s when visionary pioneers recognized the value of opening up government data for public scrutiny and active engagement. A significant catalyst for this movement was the introduction of the Open Government Directive by President Barack Obama in 2009, followed by the memorandum on Open Data Policy in 2013. These directives placed a strong emphasis on the foundational principles of transparency, participation, and collaboration in the context of an open government. They laid the groundwork for governments worldwide to proactively release and share their data with the public, fostering increased accountability and citizen involvement. As a result, open data government initiatives have flourished globally, with numerous countries and regions acknowledging the value and potential benefits of open data, leading to the establishment of online portals for sharing open government

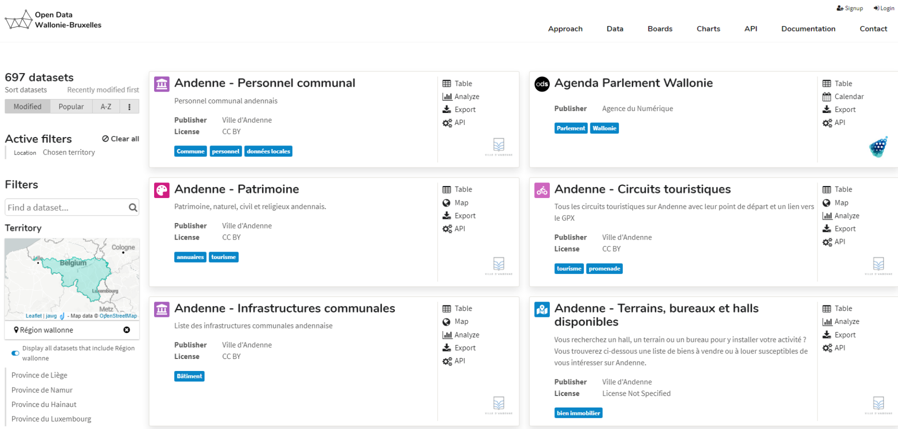


Figure 2.1: Screenshot of the Open Data Wallonia Portal

data. Major cities like London in 2010, Paris and New York in 2011, and Namur (Belgium) in 2018 have exemplified this trend by implementing dedicated platforms for the publication of open government data. Furthermore, out of the 27 European countries, only two countries lack local portals, as indicated by (Carsaniga et al., 2022).

Open government data reuse refers to the practice of utilizing and repurposing data made available by governments for various purposes. As part of the broader open government movement, the reuse of government data aims to maximize the value and impact of public information by allowing individuals, organizations, and communities to access, analyze, and leverage this data in innovative ways. The benefits of open government data reuse are manifold. First and foremost, it promotes transparency and accountability by providing citizens with access to information about government activities, policies, and outcomes. This enables citizens to hold governments accountable for their actions and decisions. Additionally, open government data reuse fosters innovation and economic growth by creating opportunities for entrepreneurs, startups, and businesses to develop new products, services, and solutions. It also facilitates evidence-based decision-making, as policymakers and researchers can draw insights from comprehensive and up-to-date data to inform policy development, planning, and evaluation.

In the specific context of Wallonia, Belgium, the maturity of open data government has been steadily progressing, although advancements in certain areas remain modest, particularly among smaller municipalities (FuturoCité, 2022). This can be attributed to a lack of political awareness and limited allocation of human and financial resources to the field, as highlighted by Futurocite's 2022 open data barometer. The barometer reveals that only 15% of the 96 surveyed municipalities actually adhere to open data principles when publishing their data, indicating a slight increase compared to previous years. Furthermore, there has been a positive development in terms of data format, with a majority of the published open data now being available in non-proprietary structured formats, signifying progress compared to previous assessments. However, it is crucial to acknowledge that although local authorities possess a wealth of valuable data for their operations and communications, this data remains largely underutilized. Surprisingly, over 50% of the surveyed municipalities are not effectively harnessing their data, with only 10% implementing visualization solutions and a mere 2% leveraging analytics and artificial intelligence to derive insights from their data, as reported by Futurocite in 2022.

In order to enhance the publication and reuse of open data in the Wallonia region, the government of Wallonia has implemented significant measures. These include the establishment of standardized data formats and the creation of dedicated open data portals, such as the Wallonia Open Data portal. This portal offers a diverse range of datasets spanning various domains, including transportation, environment, demographics, and more (refer to Figure 2.1). Additionally, the region actively engages with the open data community through the organization of events, hackathons, and collaborative initiatives, fostering a culture of innovation and encouraging the reuse of data. The government has also introduced specific Calls for Projects that facilitate the initiatives utilizing administrative data to enhance public policies. Notable examples include the Intelligent Territory Call for Projects[1], Start IA[2], and Tremplin IA[3]. Furthermore, the recent adoption of the Decree on the dissemination and re-use of public sector information (PSI/OD Directive) by the Wallon region in November 2022 demonstrates a firm commitment to open data and its reusability. This decree aligns with Directive (EU) 2019/1024[4] of the European Parliament and of the Council, reinforcing the region's dedication to advancing the maturity of open data government in Wallonia and harnessing the potential of open data to drive positive societal and economic outcomes.

2.2 OGD Stakeholders

Previous research (Gonzalez-Zapata and Heeks, 2015; Graves and Hendler, 2013; Safarov et al., 2017; Lassinantti et al., 2019) has identified a wide array of user types, including developers, open data researchers, large corporations, students, city managers, NGOs, civic tech communities, journalists, startups, entrepreneurs, media industry professionals, civil society organizations, citizens, and researchers, as key stakeholders in the use and propagation of open data. Drawing upon user category terminologies already defined in the literature (Graves and Hendler, 2013; Shaharudin et al., 2023; Johnson and Greene, 2017) and interviews with some stakeholders involved in the OGD reuse during our research study (Chokki et al., 2022c), we consolidated the aforementioned user types into three categories based on their roles in the context of OGD reuse: citizens (Graves and Hendler, 2013), infomediaries (Shaharudin et al., 2023; Johnson and Greene, 2017), and publishers (Graves and Hendler, 2013). Figure 2.2 summarizes the retained stakeholders and the interaction between them.

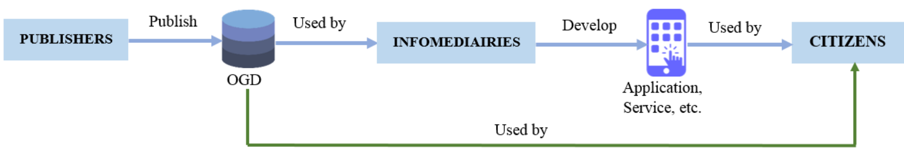


Figure 2.2: OGD Stakeholders. The green link illustrates that citizens can occasionally use open data directly without requiring the involvement of infomediaries.

Citizens: Citizens are key stakeholders in the OGD ecosystem and describe a group of people that are not directly interested in OGD, although they may consume it via applications, visualizations or dashboards (Graves and Hendler, 2013). They are the end-users and beneficiaries of open data, utilizing it to make informed decisions, engage with government activities, and participate in public affairs. Nevertheless, it's worth pointing out that a subset of these citizens, endowed with technical expertise,

[1] https://www.digitalwallonia.be/fr/programmes/smart-region/
[2] https://www.digitalwallonia.be/fr/publications/start-ia/
[3] https://www.digitalwallonia.be/fr/publications/tremplin-ia/
[4] https://eur-lex.europa.eu/eli/dir/2019/1024/

are keenly interested in OGD. These people transcend the typical citizen role and are referred to as infomediaries in this thesis.

Infomediaries: Infomediaries, also referred to as OGD intermediaries, serve as a bridge connecting the government and its citizens (Shaharudin et al., 2023; Johnson and Greene, 2017). They are specific categories of open data users who extract, aggregate, and transform data (i.e., unstructured or structured data), altering it into a format that is seen as valuable, beneficial, and, most importantly, usable to the general public (i.e., data in user-friendly format for example excel format, visualization, dashboard, service, application) (Shaharudin et al., 2023; Johnson and Greene, 2017). They play a crucial role in bridging the gap between the supply and demand of open data. These stakeholders include developers, startups, researchers, students, civil society organizations, media industry professionals, non-governmental organizations, large corporations, civic tech communities, journalists, and data intermediaries. Infomediaries analyze, interpret, and visualize open data to make it more accessible, understandable, and relevant to citizens. They play an essential role in providing feedback on data quality, identifying data gaps, and suggesting areas for improvement. They work to promote data literacy, engage citizens with open data, and advocate for the use of open data for social impact.

Publishers: Publishers are primarily government entities responsible for collecting, managing, and releasing open data (Graves and Hendler, 2013). They play a central role in the OGD ecosystem as data custodians and providers. Government agencies, city managers and departments at various levels generate a wide range of datasets that are made available to the public. Publishers should ensure open data is accessible, usable, and of high quality. Additionally, they should set policies, standards, and guidelines for data publication while putting in place measures for data privacy and security.

These three stakeholder categories work in synergy to foster the growth and impact of OGD. Citizens benefit from the availability of open data, infomediaries help bridge the gap between data and citizens' understanding, and publishers facilitate the release and management of OGD. It's important to recognize that citizens might sometimes source data directly from the publishers without involving infomediaries, especially when the data is neatly organized and accompanied by clear metadata (as indicated by the green link in figure 2.2). Collaboration and engagement among these stakeholders are essential for promoting transparency, accountability, and innovation within the OGD ecosystem.

2.3 Barriers to OGD Reuse

The reuse of open government data is not without its challenges, and several barriers hinder its effective utilization. These barriers can be classified into four categories: technical, social, legal, and economic (Janssen et al., 2012; Crusoe and Melin, 2018; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013). Technical barriers include difficulties in accessing and processing the data, while social barriers are related to a lack of awareness or interest in using the data. Legal barriers refer to restrictions on the reuse of OGD, such as copyright or privacy concerns. Economic barriers arise from the cost of accessing or processing the data. While all these barriers are important, this thesis focuses on technical and social barriers, and more specifically on three main barriers. Three specific barriers were identified based on semi-structured discussions with at least three individuals from each OGD stakeholder group in Wallonia (Belgium) during our research study (Chokki et al., 2022c). The following sections will present more about each barrier and their related research gaps.

2.3.1 Lack of communication and collaboration between OGD stakeholders

In order for open government data to be effectively reused, it is crucial that users are aware of its existence and understand its value. However, despite the availability of a vast amount of data, only a small number of users are aware of OGD and its potential benefits (Toots et al., 2017; Abdelrahman, 2021). This highlights the need to examine appropriate communication methods that can raise citizen awareness about the existence and usefulness of OGD. While several studies have explored specific methods such as social media applications (Gunawong, 2015), training programs (Gascó-Hernández et al., 2018), OGD portals, hackathons, and newspapers (Berends et al., 2020; Simperl and Walker, 2020; European Environment Agency, 2019; Michael et al., 2014), none of them have comprehensively evaluated multiple methods with users to recommend suitable approaches for increasing awareness and usefulness of OGD to citizens. Therefore, Research Gap 1a (RG1a) can be identified as follows.

RG1a: Limited understanding of effective communication methods with users to determine the most suitable approaches for OGD awareness.

Based on these previous studies (Michael et al., 2014; European Environment Agency, 2019; OECD, 2019; Berends et al., 2020; Chokki et al., 2022f), it appears that the use of applications could be an appropriate method to aware citizens of OGD. However, existing applications aimed at raising citizen awareness about OGD have certain limitations.For example, many OGD portals, which are primarily used to raise citizen awareness, focus more on publishing data than presenting the potential reuses of such data, which are more relevant to citizens, rather than the raw data. Moreover, when the OGD are used in some reuses (e.g., applications or services), there is no highlighting of the use of OGD in the reuse or any information about the OGD used. In addition, to the best of our knowledge, there are no studies addressing the requirements and impact of an application on citizen awareness of OGD. Hence, Research Gap 1b (RG1b) emerges as follows.

RG1b: Limited understanding of the necessary requirements to design a tool that effectively raises user awareness of OGD.

While raising awareness is an essential step, it does not guarantee extensive usage of OGD. The limited utilization of OGD can be partially attributed to the current publication process, which lacks effective collaboration and communication among stakeholders (Beno et al., 2017). Publishers control the release of data without adequately considering user needs, while infomediaries develop applications without aligning them with citizen requirements. Various methods and platforms have been proposed in the literature, such as hackathons, to promote communication and collaboration, but they often exclude citizens from the process (Gebka et al., 2019). Existing OGD portals also do not provide a shared space for stakeholders to collaborate, report issues, and co-create new services (Janssen et al., 2012; Zuiderwijk et al., 2012; Beno et al., 2017). Despite these proposed methods, none of them fully satisfies the requirements for effective communication and collaboration among OGD stakeholders. Research Gap 1c (RG1c) is formulated as follows.

RG1c: Limited understanding of the essential requirements for designing a tool that facilitates communication and collaboration between OGD stakeholders.

2.3.2 Insufficient (meta)data quality on OGD portals

Once users are aware of open government data, they can utilize it for various purposes. However, it is crucial to assess the quality of the data before using it, as inadequate metadata or data quality can result in inaccurate or incomplete analyses, decision-making, and research (Crusoe et al., 2019; Kubler et al., 2018). Previous studies have proposed methods to evaluate the quality of metadata, data, or both. However, these methods have several limitations. Firstly, most of the methods (Bhandari et al., 2021; Chokki et al., 2022a; Neumaier et al., 2016) focus on evaluating the overall quality of portals rather than individual datasets, making it challenging to apply them to assess a single dataset. Secondly, some methods (Li et al., 2018; Nogueras-Iso et al., 2021; Vetrò et al., 2016; Wenige et al., 2021) rely on manual assessment, which is time-consuming and costly to perform, to evaluate quality dimensions. Thirdly, certain methods (Raca et al., 2022; Reiche et al., 2014; Wenige et al., 2021) can only evaluate either metadata or data, consider a limited number of dimensions, or combine metadata and data dimensions, making it difficult for users to identify specific data quality issues (Raca et al., 2022; Reiche et al., 2014; Wenige et al., 2021). Lastly, most of the methods (Bhandari et al., 2021; Chokki et al., 2022a; Chu and Tseng, 2016; Raca et al., 2022) do not take into account user needs in the assessment, despite the fact that the acceptance of data can vary among users based on the context of data reuse (Even and Shankaranarayanan, 2007). Furthermore, to the best of our knowledge, no study has explicitly presented a comprehensive list of dimensions that are applicable separately to OGD and its metadata. Therefore, Research Gap 2 (RG2) can be identified as:

RG2: Limited understanding of comprehensive dimensions tailored individually to OGD and its metadata, and the need to incorporate user needs.

2.3.3 Difficulty to turn data into stories

Once users have become aware of the existence and usefulness of OGD and have gained knowledge about the quality of these datasets, the challenge remains in effectively transforming this data into compelling narratives.

Before delving into the task of creating data-driven stories, it is crucial to identify the data that will be of interest to a wide range of users. Previous studies show that user engagement with OGD depends on how relevant the content of OGD portals is to them (Wirtz et al., 2018; Jurisch et al., 2015). However, despite the numerous datasets published on OGD portals, only a few of them are utilized. A reason for the lack of reuse by users is that many datasets available on portals do not meet the information expectations of end-users (Crusoe et al., 2020). Prior research, such as those by (EUROPEAN COMMISSION, 2020; Nikiforova, 2022b), has enumerated high-value datasets deemed as core or specific to certain regions. However, this thesis narrows its focus to the Wallonia context to deeply comprehend the requirements of Wallonia's users. This focus aims to curate use cases for the user testing sessions of our developed applications. Moreover, as (Nikiforova et al., 2023) highlighted, a region or country's distinctiveness—be it geographical, socio-economic, cultural, or market-driven—can dictate which datasets are deemed of high value. In addition, to our knowledge, there is currently no study that has specifically identified datasets of interest to users using a methodology that actively involves them and incorporates portal usage statistics in the Wallonia context. As a result, Research Gap 3a (RG3a) can be identified as follows:

> **RG3a**: Insufficient insights into datasets that captivate users' interest, especially when emphasizing methodologies that engage them and lean on portal usage statistics, within the Wallonia context.

Another potential reason for the lack of data reuse is the inadequacy of visualizations used to represent datasets and the manner in which they are presented. Existing studies related to OGD have explored only a limited set of visualization types (Ornig et al., 2017) or proposed general interactivity methods (Khan and Shah Khan, 2011) without focusing on users' expectations with different types of datasets. Hence, Research Gap 3b (RG3b) can be formulated as:

> **RG3b**: Limited understanding of users' visualization expectations for diverse dataset types within the OGD context.

Acknowledging that individual visualizations may not offer a comprehensive representation of dataset content, we propose the adoption of dashboards as a means of aggregating and presenting multiple indicators through a unified graphical interface (Kitchin et al., 2015). Dashboards have been extensively used in the private sector, facilitated by various tools such as business intelligence (BI) platforms (e.g., Google Data Studio, Tableau, Power BI). Recently, they have found applications in the public sector through city dashboards, such as those employed in Dublin[5], London[6], and New York[7]. These city dashboards offer an innovative way to present several indicators using multiple city open and private (i.e., collected from private organizations and whose reuse requires a license) data to users. However, despite their potential benefits, dashboard systems in the public sector have faced criticism, as they often fail to meet the essential requirements for improving citizen engagement due to various challenges (Kitchin and Mcardle, 2016; Few, 2006a). Notable issues identified in dashboards, such as London City Dashboard and Boston's City Score, include difficulties in understanding and utilizing the provided information due to the lack of comprehensive details on the tracked tasks, insufficient interpretation of visualizations with confusing terminologies and abbreviations, inadequate information about data quality and veracity, and the absence of mechanisms for collecting citizen feedback. To address these challenges, the deployment of dashboards should adhere to best practices, as highlighted by (Matheus et al., 2020). Properly designed dashboards have the potential to foster transparency, accountability, and enhanced citizen engagement. Nevertheless, there is currently no study that presents dashboard principles tailored to ensure user understanding in the context of OGD and evaluates the impact of applying these design principles on citizen engagement with OGD. Therefore, Research Gap 3c (RG3c) can be identified as follows:

> **RG3c**: Insufficient insights about essential dashboard principles in the OGD context that promote user understanding and their influence on citizen engagement with OGD.

After addressing the previously mentioned research gaps, our subsequent objective is to delineate the prerequisites for data storytelling tools that would facilitate user-driven narrative creation. One example of such tools lies in business intelligence software (e.g., Tableau, Power BI) primarily tailored for businesses to monitor their performance, yet they can be adapted for open data utilization as well. Nonetheless, these tools necessitate users to download the data prior to use, and their operation entails a steep learning curve (Graves and Hendler, 2013). On the

[5] https://dublindashboard.ie/
[6] https://citydashboard.org/london/
[7] https://datausa.io/profile/geo/new-york-ny

other hand, there exist web-based alternatives (e.g., Datawrapper, Google Studio Data, The Gamma (Petricek, 2017)), which offer a more user-friendly experience. Nevertheless, similar to their predecessors, these tools do not permit direct connections with open data portals or facilitate the collection of feedback on the produced narratives. Additionally, there are generic tools intended for open data utilization, such as OpenDataVis (Graves and Hendler, 2013), SPOD (Cordasco et al., 2017), ChartViz (Pirozzi and Scarano, 2016), and YDS (Brolcháin et al., 2017), allowing integration with open data. However, they possess limited functionalities (e.g., SPOD lacks data processing capabilities, such as obtaining quick overviews of columns or data quality, or combining datasets), compelling users to rely on multiple tools to complete their tasks. In essence, although numerous data storytelling tools have been proposed, none have fully encompassed the essential features required to handle open data throughout the entire storytelling process (i.e., from data collection to deployment and feedback stages). Furthermore, to the best of our knowledge, no studies have been conducted to propose a comprehensive list of features for the design of a generic and end-to-end data storytelling tool tailored to open data. Consequently, Research Gap 3d (RG3d) can be identified as follows:

RG3d: Limited understanding of the essential features needed for the design of a generic and end-to-end data storytelling tool tailored for OGD.

Part II

Research Design

CHAPTER 3

RESEARCH QUESTIONS

The primary aim of this thesis is to explore strategies to support stakeholders in the reuse of open government data. Due to the broad scope of this objective, we narrowed our research focus to three specific barriers, which were identified through an extensive analysis of existing literature (as discussed in Section 2.3). For each of these selected barriers, we have associated corresponding research gaps and formulated research questions to address them.

Barrier 1 pertains to the "lack of communication and collaboration among stakeholders involved in the reuse of open government data." This barrier is connected to Research Gaps 1a, 1b, and 1c. The corresponding research question that addresses Barrier 1 is formulated as follows:

- **RQ1**: What facilitates effective communication and collaboration among OGD stakeholders in data reuse?

This research question serves as an overarching inquiry that will be further elaborated into three sub-research questions, each addressing Research Gaps 1a, 1b, and 1c respectively. The first sub-question directly focuses on RG1a and aims to identify appropriate communication methods to enhance user awareness of the existence and usefulness of OGD. The second sub-question addresses RG1b and seeks to identify the requirements for designing a tool that raises user awareness of OGD. The third sub-question tackles RG1c and aims to identify the requirements for designing a tool that promotes communication and collaboration among OGD stakeholders. Therefore, RQ1 is refined into the following sub-research questions:

- **RQ1a**: What communication methods effectively raise user awareness about the existence and utility of OGD?
- **RQ1b**: What are the key design considerations for a tool that promotes OGD awareness among users?
- **RQ1c**: What elements are crucial for a tool that enhances communication and collaboration among OGD stakeholders?

Once users are aware of OGD, they can then utilize it. However, a significant number of datasets on OGD portals lack good quality, making it crucial to provide users with an overview of the data and metadata quality before utilizing them. Thus, Barrier 2 centers on assessing the quality of individual open government datasets and is associated with Research Gap 2. The research question addressing Barrier 2 is as follows:

- **RQ2**: What supports automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user needs?

Subsequently, after users have acquired awareness of OGD and are familiar with their quality, they require suitable applications to transform their data into engaging narratives. Thus, Barrier 3 centers on data storytelling tools and is associated with Research Gaps 3a, 3b, 3c, and 3d. The research question that addresses Barrier 3 is as follows:

- **RQ3**: What facilitates the transformation of OGD into stories?

This research question constitutes an umbrella question that will be refined into 5 sub-research questions regarding their implication on addressing the Research Gaps 3a, 3b, 3c and 3d. The first sub-question directly addresses RG3a and aims to identify the data that users are interested in. The second sub-question tackles RG3b and focuses on identifying the appropriate visualization types for different data categories. The third sub-question addresses part of RG3c by identifying the design principles for effective dashboards in the context of OGD. The fourth sub-question explores the impact of applying the identified dashboard design principles on citizen engagement with OGD. The fifth and final sub-question addresses RG3d and seeks to identify the essential features for designing a comprehensive end-to-end data storytelling tool in the context of OGD. Therefore, RQ3 is refined into the following research sub-questions:

- **RQ3a**: What datasets are of primary interest to users in the Wallonia context?
- **RQ3b**: What visualization optimization approaches are most effective for specific type of datasets?
- **RQ3c**: What are the design principles for effective dashboards in the context of OGD?
- **RQ3d**: In what ways do well-structured dashboards, incorporating best dashboard design principles, enhance citizen engagement with OGD compared to individual visualizations?
- **RQ3e**: What essential features should be incorporated in a comprehensive, end-to-end data storytelling tool tailored for open data?

Table 3.1 provides a summary of the mapping between the formulated research questions, the research gaps and the barriers.

Table 3.1: Mapping between Research Questions, Research Gaps and Barriers.

RQ	**Research Gaps**								**Barriers**		
	RG1a	**RG1b**	**RG1c**	**RG2**	**RG3a**	**RG3b**	**RG3c**	**RG3d**	**B1**	**B2**	**B3**
RQ1	X	X	X						X		
RQ1a	X								X		
RQ1b		X							X		
RQ1c			X						X		
RQ2				X						X	
RQ3					X	X	X	X			X
RQ3a					X						X
RQ3b						X					X
RQ3c							X				X
RQ3d							X				X
RQ3e								X			X

CHAPTER 4

METHODOLOGY

This chapter outlines the research methodology utilized to investigate the research questions introduced in Chapter 3. In Section 4.1, a concise overview of the comprehensive methodology applied consistently across all three research questions is presented. Further details about the specific methodologies employed for each research question can be found in their respective sections within the Part III. Additionally, Section 4.2 provides a description of the underlying research methodologies that support this overall approach.

4.1 Overall Methodology

Given the objective of this thesis, which involves the development of tools to address various research questions, we have chosen to employ the Design Science Research (DSR) approach (Hevner et al., 2004). This well-established methodology allows for the creation of relevant tools (the "Design Cycle") that align with the knowledge base (the "Rigor Cycle") and the environment (the "Relevance Cycle"). Figure 4.1 illustrates the visual representation of the DSR methodology. For each research question, we predominantly applied the DSR approach, which consists of three main stages: information or requirements collection (Rigor Cycle), prototyping (Design Cycle), and evaluation (Relevance Cycle).

In the information or requirements collection stage (**Inf/Req Collection**), we gathered the necessary specifications for the tool or information corresponding to address the research question. This involved conducting a literature review (**LR**), exploratory study (**ES**), administering questionnaires (**Q**), conducting interviews (**I**) and analyzing the consultation statistics of Namur portal (**CS**). Subsequently, in the prototyping stage (**Proto**), we implemented the identified requirements into a functional and user-friendly tool. Finally, in the evaluation stage (**Eval**), we utilized the implemented prototype as a proxy to validate the requirements. This validation process involved user testing (**UT**) along with questionnaires to gather valuable feedback and ensure the tool's effectiveness and usability.

Table 4.1 offers a concise overview of how the formulated research questions were mapped to the research methods employed at each stage of the Design Science Research (DSR) approach. It is essential to mention that for research questions RQ1a, RQ3a, and RQ3b, a full DSR was not applied, as these questions required specific information collection on certain aspects of the study.

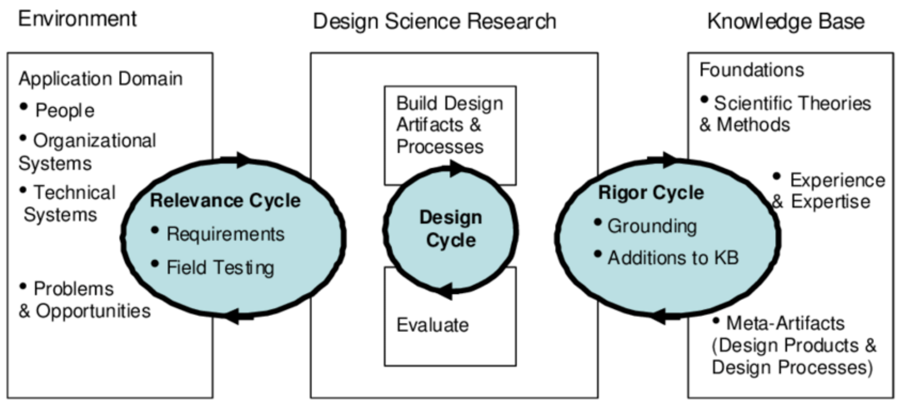


Figure 4.1: Design Science Research Methodology (Hevner et al., 2004).

Table 4.1: Mapping between Research Questions and Research Methods.

	Inf/Req Collection (Rigor)					Proto	Eval (Relevance)	
RQ	**LR**	**Q**	**I**	**ES**	**CS**	**(Design)**	**UT**	**Q**
RQ1a	X	X						
RQ1b			X			X	X	X
RQ1c	X		X			X	X	X
RQ2	X					X	X	X
RQ3a		X			X			
RQ3b			X					
RQ3c	X							
RQ3d						X	X	X
RQ3e	X			X		X	X	X

4.2 Underlying Research Methods

With the primary objective and methodology at the forefront, our intention was to maintain a harmonious and cohesive approach throughout the research journey. To ensure this coherence, we will utilize the research "onion" framework, as presented by Saunders et al., 2009. This framework, depicted in Figure 4.2, emphasizes the importance of making informed decisions across various layers of the research process before proceeding with data collection. By employing this framework, we strive to establish a consistent and well-grounded research methodology that aligns with our overall goals. The subsequent sections provide a detailed overview of each layer of the research onion.

4.2.1 Philosophy

The research philosophy constitutes the outer layer of the research onion, shaping the researcher's perspective and beliefs concerning the nature of knowledge and reality (Saunders et al., 2009). Considering our primary aim of offering practical solutions for OGD reuse, adopting the pragmatic approach appears fitting. This choice is also supported by our utilization of a blend of quantitative and qualitative methods, aligning with the pragmatic approach's typical research methods.

4.2.2 Approaches

This layer of the research onion pertains to the overarching strategy or methodology employed in conducting the research (Saunders et al., 2009). It encompasses two primary research approaches: the deductive approach and the inductive approach. However, for this thesis, we have chosen the abductive approach, which blends elements from both deductive and inductive approaches. In adopting the abductive approach, we utilize existing theories from the literature review and observations from user testing and questionnaires to generate plausible explanations for the observed phenomena. This approach allows us to iteratively reason and develop new insights, combining existing knowledge with empirical data to address our research questions effectively.

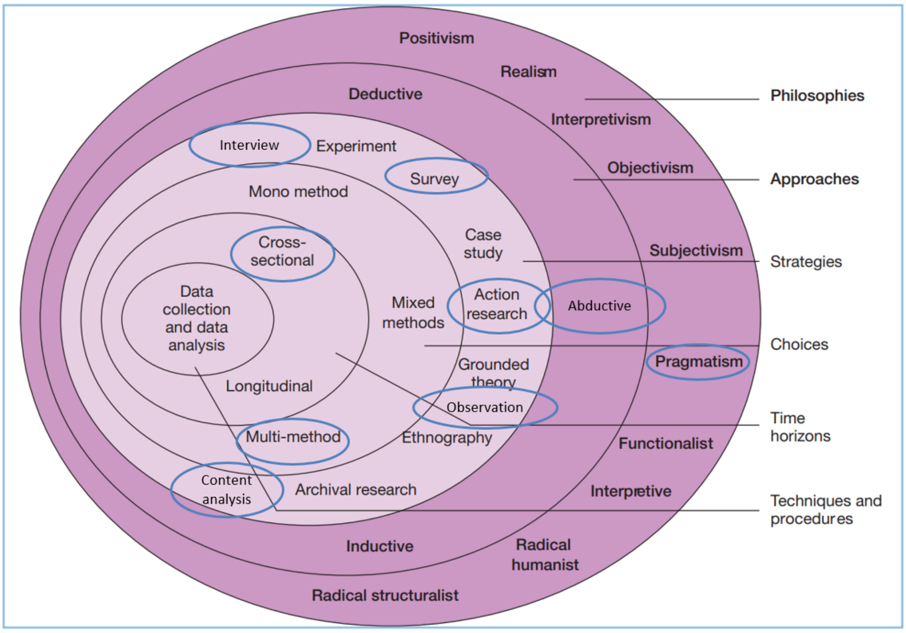


Figure 4.2: Research Onion adapted from (Saunders et al., 2009). The blue ellipses denote our selections for each layer within the research onion.

4.2.3 Strategies

The research strategies focus on how the research objectives will be achieved (Saunders et al., 2009). Therefore, in this thesis, five research strategies are used and each of them is used based on the research questions and the type of data to be collected.

Surveys: This research strategy involves the collection of quantitative data through structured questionnaires or surveys from a large number of participants. Surveys are utilized primarily during the evaluation of the developed tools, but they also serve to gather users' preferences, as demonstrated in Table 4.1. For instance, in addressing RQ1a, which centers on identifying appropriate communication channels for OGD awareness, surveys are employed to collect users' preferred communication channels.

Interviews: This research strategy entails conducting one-on-one or group discussions with participants to obtain in-depth qualitative data (Saunders et al., 2009). In this thesis, interviews are primarily utilized to gather requirements for the tools and to collect users' opinions on specific topics. For instance, when addressing

RQ3b, which focuses on identifying appropriate visualization types for specific data categories, interviews are employed as a means of data collection.
Observation: This research strategy involves observing and recording behaviors or events in their natural settings to gather qualitative data (Saunders et al., 2009). In this thesis, observation is primarily utilized during the evaluation of the tools, particularly during user testing, to assess the ease of use of the implemented tools.
Content Analysis: Content analysis is a research strategy that involves analyzing written or visual materials, such as documents, articles, or media, to derive insights (Saunders et al., 2009). In this thesis, content analysis is primarily used to extract requirements needed for the development of the tools from the existing literature.
Action Research: Action research is a collaborative approach where researchers work closely with practitioners to address real-world problems and implement solutions (Saunders et al., 2009). In line with this approach, we have actively engaged and collaborated with various OGD stakeholders throughout the research process. For example, we have involved stakeholders in the development of a tool that facilitates collaboration among them, ensuring that the solutions are tailored to their needs and effectively address the challenges they face.

4.2.4 Choices

According to Saunders et al., 2009, there are two main research choices: the mono-method and the multi-method. The mono-method involves using either quantitative or qualitative data collection and analysis. Quantitative methods generate or utilize numerical data, while qualitative methods use non-numerical data, collected through techniques like interviews or analyzed through categorization. On the other hand, the multi-method approach involves using more than one data collection technique and analysis procedure. It allows researchers to gather a broader range of insights and perspectives.

In this thesis, we adopted the multi-method approach. For example, in evaluating the tools, we used surveys that included both qualitative and quantitative questions. The qualitative questions aimed to gather users' overall opinions on the tools, while the quantitative questions were based on established questionnaires like System Usability Scale (SUS) questionnaire (Brooke, 1986) and Technology Acceptance Model (TAM) questionnaire (Davis, 1989; Moreno Cegarra et al., 2014), providing a more comprehensive assessment.

4.2.5 Time horizons

Time horizons in research refer to the timeframe within which a study is conducted. According to (Saunders et al., 2009), there are two classifications: cross-sectional and longitudinal. In a cross-sectional study, the investigation focuses on a specific time period, aiming to provide a snapshot of a particular situation. On the other hand, longitudinal research extends over an extended period, observing and examining changes and developments in a given topic.

For this thesis, the chosen time horizon was cross-sectional. Both qualitative and quantitative data collection took place at specific points in time, as the investigation needed to be completed within a limited period of four years. Therefore, the research aimed to gain insights within a particular timeframe rather than tracking changes over an extended duration.

4.2.6 Techniques and Procedures

Table 4.1 provides an overview of the primary data collection techniques utilized in this thesis, including interviews, surveys, literature reviews, and focus groups.

Specific information regarding the data collection and analysis for each research contribution can be found in their respective sections in the Part III. In relation to the focus groups, as outlined in Section 2.2, our focus was on engaging OGD stakeholders, particularly citizens, infomediairies, and publishers.

Part III

Contributions

CHAPTER 5

Tackling Barrier 1: Improving Communication and Collaboration among OGD Stakeholders

5.1 General Introduction

This chapter presents the contributions developed to address the first barrier, that is, the lack of communication and collaboration among stakeholders involved in the reuse of open government data. In order for OGD to be effectively reused, users need to be aware of its existence (awareness) and know their added value (usefulness) (Zuiderwijk et al., 2015). Yet, even though a lot of data are published, only a few users are aware of its existence and usefulness (Abdelrahman, 2021; Toots et al., 2017), which hinders achieving the goal of OGD initiatives. To tackle this issue, the research question **(RQ1a)** is formulated: **"What communication methods effectively raise user awareness about the existence and utility of OGD?"** A literature review and online survey were conducted to identify effective communication methods, and recommendations were provided to OGD publishers based on the findings. The results revealed that users preferred practical applications, leading to the formulation of another research question **(RQ1b)**: **"What are the key design considerations for a tool that promotes OGD awareness among users?"**. This question was addressed through a literature review to gather requirements for an awareness-raising tool, which were subsequently validated through user testing and interviews.

In addition, it is essential to recognize that awareness alone does not ensure the utilization of open government data. Although raising awareness is a critical first step in the data reuse process, it does not automatically result in extensive usage. This is evident from the limited actual utilization of OGD, even when users are aware of its existence and benefits. One possible explanation for this is the current publication process, which is controlled by publishers who have the authority to determine which data is made available. In this thesis, the approach is reversed, suggesting that data should be published in response to user requests, with infomediaries developing applications and services based on citizen needs. To facilitate this interaction between stakeholders, the research question **(RQ1c)** is posed: **"What elements are crucial for a tool that enhances communication and collaboration among OGD stakeholders?"** This question was addressed through a literature review to gather requirements for a tool that promotes communication and collaboration among OGD stakeholders, which were validated through user testing and interviews.

5.1.1 Publications

The content of this chapter is based on the following peer-reviewed scientific publications:
Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Increasing Awareness and Usefulness of Open Government Data: An Empirical Analysis of Communication Methods. In **16th International Conference on Research Challenges in Information Science**, 2022f

- This paper examines the preferences of users regarding the channels utilized for promoting OGD. Through a combination of a literature review and an online survey, effective communication methods were identified and the results were aggregated to offer recommendations to governments.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open government data awareness: eliciting citizens' requirements for application design. **Transforming Government: People, Process and Policy**, 2022e. doi: https://doi.org/10.1108/TG-04-2022-0057

- Building upon the previous paper (Chokki et al., 2022f), this study expands its scope by identifying and validating the 11 requirements for infomediaries to develop a functional tool aimed at increasing citizen awareness of open government data. This is achieved through a comprehensive approach that includes a literature review, interviews, and user testing.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, Benoît Frénay, and Benoît Vanderose. Fostering Interaction between Open Government Data Stakeholders: An Exchange Platform for Citizens, Developers and Publishers. In **Electronic Government**, 2022c

- In this paper, a set of 16 requirements is identified for the development of a tool that enhances communication and collaboration among stakeholders involved in OGD. Additionally, practical contributions are made by presenting a use case diagram that outlines the necessary features to fulfill these requirements, along with the implementation of a functional tool that incorporates them. These outcomes are the result of a thorough investigation that encompasses a literature review, interviews, and user testing.

5.1.2 Outline

This chapter is organized as follows: Section 5.2 presents the communication methods to enhance OGD awareness. Section 5.3 discusses the requirements for designing a tool that raises user awareness of OGD. Section 5.4 outlines the requirements for a tool that facilitates communication and collaboration among OGD stakeholders. Finally, Section 5.5 concludes the chapter by summarizing its contributions.

5.2 Users' Preferences for Communication Methods about OGD Awareness

Previous studies (Gunawong, 2015; Gascó-Hernández et al., 2018; Simperl and Walker, 2020) have made efforts to investigate the suitability of specific methods or provide recommendations to increase awareness and usefulness of open government data to users. However, these studies have not thoroughly evaluated multiple methods with users to effectively recommend appropriate approaches for governments to enhance awareness and usefulness of OGD. This section aims to bridge this gap by identifying suitable methods to raise awareness and usefulness of OGD among users and providing recommendations to governments regarding the most appropriate methods to promote OGD to users.

The remainder of this section is divided into four subsections. In Section 5.2.1, we outline the research methodology employed in this study. Section 5.2.2 examines existing methods for raising user awareness of the existence and usefulness of OGD and provides an overview of the survey results. Finally, Section 5.2.3 summarizes the contributions made in this section, highlights the limitations of the study, and suggests potential directions for future research.

5.2.1 Research Methodology

To address the research question of this study **(RQ1a) "What communication methods effectively raise user awareness about the existence and utility of OGD?"**, we combined two methods: literature review and questionnaires.

First, from September to October 2021, we conducted a literature review, on methods used to raise user awareness of the existence and usefulness of OGD. The literature review was conducted using the databases “Scopus” and “Science Direct” with the keywords (“open government data” OR “open data”) AND (“citizen” OR “user”) AND (“promote” OR “raise”) AND (“awareness” OR “usefulness” OR “existence” OR “utility”). In our research, we expanded our scope to include grey literature and policy reports, recognizing that scientific publications have addressed the topic only minimally. Interestingly, grey literature and policy reports provide a broader perspective, discussing multiple communication channels. This contrasts with the narrower focus often found in scientific publications, which tend to concentrate on a singular communication channel. By tapping into these alternative resources, we hoped to gain a more comprehensive understanding of the various communication channels being employed. Most of the publications found dated from 2011-2021. From these publications, an additional selection was made based on their relevance to our research, leaving altogether 15 academic articles, web pages and policy reports which were looked at more thoroughly. The retained articles were then used to collect appropriate methods for raising user awareness of the existence and usefulness of OGD.

Second, we created an online survey to collect users’ preferred methods. The survey was pretested with two users and later shared via the following communication channels: UNAMUR mailbox, Facebook and Twitter to recruit users. Appendix A presents the questions contained in the survey. After collecting users’ ratings, the median was used to identify the most appropriate methods. In total, 30 participants completed the survey. The literature review along with the user feedback were used to improve the current knowledge base and provide recommendations to governments.

5.2.2 Results

In this section, we first describe previous work on methods used to raise user awareness of the existence and usefulness of OGD. Then, we present the survey results.

5.2.2.1 Communication Methods Identified

By performing a literature review described above, we were able to identify a set of eight methods that could be used to raise user awareness of the existence and usefulness of OGD. These methods were drawn from studies such as Berends et al., 2020; Simperl and Walker, 2020; Michael et al., 2014; European Environment Agency, 2019. The following paragraphs explain each of these methods in more detail.

Social Media. According to (Mergel, 2011), social media applications are new types of information and sharing tools, used in digital environments. They have been adopted by a few governments with different purposes: sharing information, interacting with citizens, promoting citizen participation in public issues or improving transparency (Gunawong, 2015; Mergel, 2011). The most commonly used social media in the public sector are: blogs, collaborative projects (e.g., wikis, online forums), social networking sites (e.g., Facebook, Twitter) and content communities (e.g., YouTube) (Gunawong, 2015; Mergel, 2011). Although social media applications offer many benefits, they can only reach specific citizens. For example, in the case of social networking sites, only the citizens who have an account and who fall within the criteria used for campaigns can be reached.

Public outreach campaigns. Apart from social media, a few governments have used methods such as radio, television, newspapers, newsletters and poster campaigns to inform citizens, especially of some applications built on the basis of OGD (European Environment Agency, 2019; Michael et al., 2014). The problem with these methods is that the content of the advertisement focuses on the implemented application without telling citizens that the application was implemented using open data. Therefore, citizens may use the service without knowing that it was built using open data.

Workshops and Conferences. These types of events aim to bring together various open data stakeholders to discuss the adoption and use of open data (Berends et al., 2020; European Environment Agency, 2019; Ojo and Janssen, 2014). Two well-known examples of these types of events are the Open Government Data Camp and Open Data Day. The main advantage of these types of communication is that they help governments to have a direct discussion with citizens and also gather their feedback (e.g., needs, barriers) for improvement (Cook and Jurkat, 2011). However, there are limitations, such as the potential restriction in participant numbers due to these communication methods. Some participants might be keen to attend the conference or workshop but might be unavailable at the designated event time.

Hackathons. Like the previous method, this method is an event that allows developers to design, implement and present services for a specific issue (Johnson and Robinson, 2014; Berends et al., 2020; European Environment Agency, 2019; Simonofski et al., 2020). This method allows for the promotion OGD to participants and the development of some services that can be published later to help a wider range of citizens. However, this method faces the following problems. It primarily target developers, though recently, citizens with limited technical skills have begun participating in these events (Purwanto et al., 2019; Nikiforova, 2022a). However, many outcomes from hackathons are not implemented or made available online post-event, influencing awareness mainly among developers rather than the broader citizenry (Simperl and Walker, 2020; Gebka et al., 2019).

Training and Education. This method consists of bringing together different stakeholders to inform or instruct them on a certain task with the aim of improving their performance or knowledge (Berends et al., 2020). In (Berends et al., 2020; European Environment Agency, 2019), they suggested enabling the creation of a "culture for Open Data" to students by integrating the use of open data (e.g. building apps) into academic programs. This method has been experimented in (Gascó-Hernández et al., 2018) and by Thessaloniki's Digital Strategy (Berends et al., 2020) but this method only attracts a specific and limited part of digitally literate users.

Public Displays. These are mainly outdoor displays, as well as indoor displays in public spaces, which offer different benefits to users (refer to "passersby"): collecting feedback such as voting system, displaying information or accessing services (Clarinval et al., 2021). The main advantage of public displays is that they help citizens to interact directly with services in real life and can be easily accessible (visible) to "passersby". However, public displays face the following problems: difficulty of interaction by a certain range of users and limited access (only available to a specific location) (Coenen et al., 2019).

Applications. These include platforms developed to help users to easily access government data and also tangible examples of what can be done with published data (Berends et al., 2020; European Environment Agency, 2019). For example, some OGD portals have offered visualizations, dashboards and success stories built from the data in addition to raw data, raising awareness of the benefits of Open Data and showing what can be done with particular datasets (Berends et al., 2020). Apart from these features, some OGD portals have also proposed News and Events sections in their portals, which helps to increase traffic to the portal (Berends et al., 2020). Another way to raise awareness among citizens is the development of practical applications and services accessible mainly on the web or mobile that use the data provided by governments and facilitate the daily life of citizens (e.g., the mobility application which helps Namur citizens to see the location of available parking in a specific area[1]). These practical applications are mainly developed either for a specific purpose or as federative applications to promote existing applications developed from open data (e.g., Datafruit[2] which summarizes in a mobile application the reuses of open datasets on the French portal). The problem with this method is that without awareness campaigns like the ones presented above, these applications cannot be acknowledged by citizens. Another problem is that each government or developer promotes their applications separately, which increases the funds used to promote the different applications and the need for citizens to go through (or install) different applications in order to use them.

Word of mouth. This method involves citizens talking to their friends, family and other people with whom they have close relationships about a topic of open data (HAYES, 2021; Chen and Yuan, 2020). This method was less discussed in the literature of open data awareness but was proven to be one of the most powerful forms of awareness in general (e.g., e-commerce) as 92% of citizens trust their friends over traditional media (HAYES, 2021).

Figure 5.1 presents an overview of the methods that can be used to raise user awareness of OGD.

5.2.2.2 Survey Results

Through questionnaires that participants completed, we were able to gather their opinions related to the research question (RQ1a) "What communication methods effectively raise user awareness about the existence and utility of OGD?" In total, 30 participants (22 are aware of OGD and 8 are not) responded to the survey. All participants are between the ages of 18 and 50 and have at least a high school degree.

After collecting the participants' responses on the channels through which they have been informed or wish to be informed of the existence and usefulness of OGD, we associated each participant response with one of the methods presented in Section 3.1, where possible. Some participants' responses were ambiguous (e.g., Google search or from municipality) and therefore were not considered. Figure 5.2 summarizes the percentage of participants' responses on the channels through which they have been informed (A1) or wish to be informed of the existence and usefulness of OGD (A2).

[1] https://sti.namur.be/

[2] http://opendatatales.com/%f0%9f%a5%a5-datafruit-un-argumentaire-de-poche-de-lopen-data/

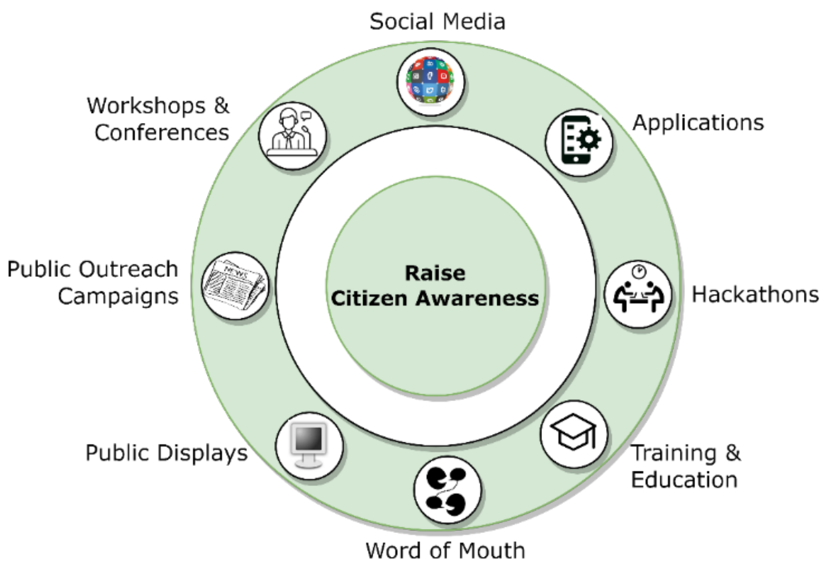


Figure 5.1: Communication methods to raise awareness of OGD.

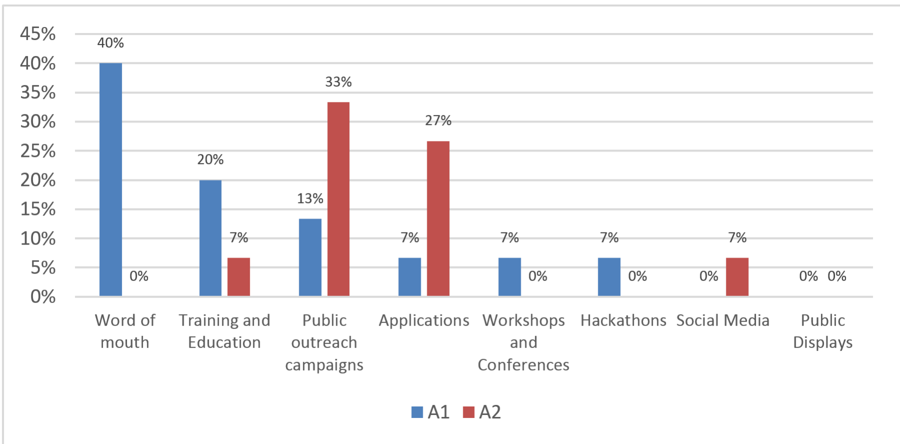


Figure 5.2: Percent of respondents for each method of awareness and usefulness of OGD applied in the practice (A1) and preference by citizens (A2).

Regarding the methods used in practice to raise awareness and usefulness of OGD, Figure 5.2 (ref. A1) shows that the “word of mouth” channel (mainly through friends or colleagues), is the channel through which most citizens have been informed about OGD. This channel is followed by “training and education”, especially “education”, as many users indicate that they have heard about OGD in their classes. After this method comes public outreach campaigns, and after applications, workshops and conferences and hackathons. The methods “social media” and “public displays” were not mentioned by users. Regarding users' preferred methods, Figure 5.2 (ref. A2) shows that users' preferred channel is "public outreach campaigns", followed by applications (suggested by citizens: OGD portals and OGD-based applications), training and education, and social media.

5.2.3 Conclusion

The purpose of this study was to identify the appropriate methods for raising awareness and usefulness of OGD to users. To achieve this objective, we first conducted a literature review to identify methods used to raise citizen awareness of OGD. Then, we used an online survey completed by 30 users to compare the results of the literature review with the citizens' perception.

This study contributes to the knowledge base in the following aspects. First, this study provides an inventory of communication methods that have been used in the literature to promote OGD to users. Second, unlike these previous studies that examine whether a specific method was appropriate (Gunawong, 2015) or suggest methods to increase awareness and usefulness of OGD to users (Berends et al., 2020; Simperl and Walker, 2020; European Environment Agency, 2019; Michael et al., 2014) without providing an evaluation, this study presents the benefits and challenges of each of the methods used in the literature and evaluates each through an online survey completed by users. Third, based on the survey results, this study highlights the discrepancies between the channels used in practice and those preferred by users to raise awareness of OGD (See Figure 5.2). Fourth, based on Figure 5.2, we recommend for governments to use public outreach campaigns and applications to inform users about the existence and the usefulness of OGD. The "word of mouth" method appears to be the most effective method used in practice for spreading awareness of OGD. Governments should therefore use the methods suggested above and encourage users to disseminate them to those around them.

However, a key limitation of this study pertains to the representativeness of the participants in the questionnaires. To address this limitation and enhance representativeness, we propose utilizing alternative communication channels or collecting data in-person at universities or public spaces. Unfortunately, in this study, such approaches were not feasible due to the constraints imposed by the COVID-19 pandemic. Another limitation is the lack of consideration for user profiles in this study. Future work will focus on assessing user awareness among specific user profiles, including citizens and infomediaries, and exploring whether there are notable disparities between the communication channels commonly used in practice and the channels preferred by these user profiles. A potential avenue for future research involves delving deeper into the content of the identified communication methods. While the current study focused primarily on identifying these communication channels, a comprehensive exploration of their content would be valuable. This deeper analysis would enable the creation of a detailed procedure for governments, offering guidance on how to optimally implement these communication methods to achieve the best outcomes. By understanding the content nuances, governments can refine their approach to effectively raise awareness and drive engagement. Given that applications have emerged as a preferred method for promoting OGD, and considering the limited number of studies investigating the requirements for a usable tool to enhance OGD promotion, our forthcoming research will aim to fill this gap by developing and evaluating a usable tool and assessing its impact on citizen awareness.

5.3 OGD Awareness: Eliciting Users' Requirements for Application Design

Numerous communications methods (e.g., social media, public outreach campaigns, workshops and conferences, hackathons, training and education, public displays, applications, and word of mouth) have been proposed in previous studies to make users aware of OGD (Michael et al., 2014; Amugongo et al., 2015; Gunawong, 2015;

European Environment Agency, 2019; Clarinval et al., 2021; Chokki et al., 2022f). In this study, we will primarily emphasize the applicative aspect, driven by the following motivations. First, based on the results on previous section 5.2, the use of practical applications is one of the best way to aware users of OGD. Second, existing applications to make users aware of OGD have limited features. For example, many OGD portals, which are primarily used to raise user awareness, focus more on publishing data than presenting the potential reuses of such data, which are more relevant to users, rather than the raw data. Furthermore, when the OGD are used in some reuses (e.g., applications or services), there is no highlighting of the use of OGD in the reuse or any information about the OGD used. Third, to the best of our knowledge, there are no studies in the literature addressing the requirements and the impact of an application on the awareness of OGD to users. This study aims to address these gaps by identifying the list of requirements needed in the design of an application to raise user awareness of OGD, implementing these requirements into a usable tool called CitizenApps and used it as proxy to evaluate the identified requirements.

The remainder of this section is divided into five main subsections. Section 5.3.1 explains the methodology used to address the research question. Section 5.3.2 presents the results of the tasks performed during research method. In Section 5.3.3, we present the contributions of this study and limitations of this study, and then propose some avenues for future work.

5.3.1 Research Methodology

To address the research question of this study **(RQ1b)**: **"What are the key design considerations for a tool that promotes OGD awareness among users?"**, we followed the Design Research Science (DSR) methodology (Hevner et al., 2004; Peffers et al., 2007; Vaishnavi and Kuechler, 2007; Baskerville, 2008; Dresch et al., 2015; Hivon and Titah, 2017). This research paradigm aims to develop solutions (artefacts) that meet defined objectives, contribute to the scientific knowledge base (rigor) and provide utility in the environment (relevance). The following paragraphs will explain the different tasks performed for each cycle of the DSR: rigor, design and relevance cycles.

In the **design cycle**, we recruited 10 students and researchers of different departments at the University of Namur in order to collect the users' requirements for the application. An incremental approach based on the agile methodology (Fowler and Highsmith, 2001) was later used to validate the requirements and progressively collect features, which is described as follows. First, we collected basic features from two users and then implemented them in the application. The developed application was then presented to another user for feedback and additional features. This process was repeated several times, with different users each time, until no additional features were suggested by the users (Guest et al., 2006; Lallemand and Gronier, 2015). Starting with the seventh user, we found that no additional features were suggested, but there were more user comments to improve the user interface design.

In the **relevance cycle**, we evaluated the final prototype version obtained after implementing the requirements of the 10 users during the design cycle. The evaluation was conducted through a user testing followed by an online survey to assess the effectiveness of CitizenApps for promoting OGD to users and to gather additional features for future versions. Appendix B presents the questions contained in the survey. Apart from the survey, we asked users to give their opinion about their perception of the awareness of open data. This question was asked before and after the presentation of CitizenApps to users to assess whether the use of CitizenApps has an impact on their perception of the awareness of open data. We also asked users four additional questions to get their overall opinion of the implemented features: *what features should be kept, improved, removed, or added in CitizenApps to raise user awareness of OGD?* Using a convenience sampling method, 25 users were recruited through the following communication channels: UNAMUR mailbox and social media (Robinson, 2014; Emerson, 2015). This sampling method was chosen for two reasons. First, due to the COVID situation, it was impossible to recruit participants in public places. Therefore, participants were selected based on ease of access and proximity to the researchers. Second, given that the number of respondents was small, there were no exclusion criteria for participation in the survey, so all subjects were invited to participate in the process. Prior to filling out the survey, the participants were asked to install the application on their android phones and review the various features. The survey was pretested with two users to ensure that all kinds of errors associated with survey research were reduced (Grimm, 2010) and later shared on Facebook and Twitter to recruit the users. The survey included three types of questions: questions with a 7-point Likert scale (from "Strongly Disagree" to "Strongly Agree") based on the Technology Acceptance Model (TAM) (Davis, 1989; Moreno Cegarra et al., 2014) to assess four aspects: attitude, perceived usefulness, perceived ease of use and behavioral intention; free text questions to collect general opinions and suggestions for additional features for future versions and to justify previous ratings; and 4 additional questions to collect demographic data. After collecting user feedback, the median, mean and standard deviation (SD) were calculated for the questions with a 7-point Likert scale to evaluate the four aspects. These statistical measures were chosen because they are the most appropriate for analyzing Likert data and for having a central tendency measure (Boone and Boone, 2012). Verbal thoughts and responses collected from the free text questions were coded using short sentences to retain context and conceptual relations.

In the **rigor cycle**, the requirements collected from the design and relevance cycles will be used to improve the current knowledge base.

5.3.2 Results

In this section, we present the results of the tasks performed during the three phases of DSR described above. First, we present the requirements that need to be implement in a usable tool to raise awareness of OGD among users and also present how the requirements were implemented in a mobile application called "CitizenApps" and evaluate its effectiveness to raise awareness of OGD among users.

5.3.2.1 Requirements Identification and Prototype Description

Based on the results of the interviews with users during the design cycle and the user feedback gathered during the design and relevance cycles, we are able to identify 11 requirements that need to be implemented in an application to raise user awareness of OGD. We are aware of the social, emotional, and motivational aspects of the adoption of OGD by users. However, it is important to note that in this study we focus more of the technical aspect of the problem. Table 5.1 presents the 11 identified requirements.

The requirements identified were implemented in a mobile application that we called "CitizenApps"[3] (available on Google Play Store). We chose to implement the prototype as a mobile application for the following reasons. First, according to a report from Eurostat[4] (2019), 92% young people were using their smartphones to

[3]https://play.google.com/store/apps/details?id=com.ogd.citizenapps
[4]https://rb.gy/ahkzzj/

access the internet outside of their homes or workplaces, compared to 52% who were using a laptop in this way. Second, with smartphones, it is easier to access a user's location in order to automatically offer them appropriate OGD reuses (i.e., the applications or services developed using OGD) in their area (**R2**). Third, with cell phones, it will be easier to send notification to users about new OGD reuses without running additional campaigns (**R10**). Fourth, a mobile application is a great visual support on pocket to show everywhere (e.g., meeting room, bus, train) what can be done with open data without having to go into a browser and type in a link with the problem of forgetting the access link.

Table 5.1: List of requirements to design a tool that can raise awareness of OGD among users. OGD reuse refers to the application or service developed using OGD.

Requirement	**Requirement description**	**Implementation in CitizenApps**	**Added value compared to existing applications mentioned in section 5.2.2.1**
R1. List of OGD reuses	Possibility to see OGD reuses with information about their title, description, associated topic (e.g., transport, finance, etc.).	OGD reuses were collected from “reuses” section of some government portals (e.g., Belgium portal[5], French portal[6], portal for European data[7]) and from reuses posted by developers and citizens on ODEON (Open Data Exchange solutiON) platform, which will be further explored in Section 5.4.3.2.	Compared to OGD portals or other applications (e.g., Datafruit, WallonieEnPoche and SNCB) that only focus on specific regions or topics or offer a specific service, CitizenApps is designed to aggregate OGD reuses for different topics and countries.
R2. Categorize OGD reuses	Possibility to categorize OGD reuses or to search for OGD reuses.	Ability to search for OGD reuses based on the following criteria: country, state, topic and keywords. In addition, since many participants were interested in OGD reuses in their region or by a specific topic, we provided quick access to OGD reuses by country or topic.	CitizenApps differs from other applications in that it allows searching by country and state which is not available in other applications since they focus on a specific country or topic.

[5]https://data.gov.be/
[6]https://www.data.gouv.fr/
[7]https://data.europa.eu/

Table 5.1: List of requirements to design a tool that can raise awareness of OGD among users. OGD reuse refers to the application or service developed using OGD.

Requirement	Requirement description	Implementation in CitizenApps	Added value compared to existing applications mentioned in section 5.2.2.1
R3. List of success stories of OGD reuses	Possibility to present users a list of OGD reuses that have an impact in the daily life of the users.	Attract more users by presenting them with the list of OGD reuses they have commonly used in the daily life or bookmarked in CitizenApps. The list of success stories is by default defined by us based on applications used in the daily life of the users like SNCB which is for example an application for train route schedules in Belgium.	This requirement to highlight key success stories in a specific area does not exist in the existing applications.
R4. List of latest OGD reuses	Possibility to view recently added OGD reuses to the application.	Ability to list recently added OGD reuses in the ODEON platform to users.	-
R5. See details about a specific OGD reuse	Possibility to view details of a specific OGD reuse.	Ability to view details of OGD reuse such as: title, description, associated topic, associated country/state, contact information, list of datasets used and list of user comments.	Details such as the associated country/state, list of datasets used and user comments provide added value over existing applications.
R6. Execute OGD reuse	Possibility to access or execute OGD reuse.	Possibility to access or execute OGD reuse without leaving CitizenApps if it is hosted online.	This requirement is more user-friendly within CitizenApps since the user does not need to exit CitizenApps before running the application thanks to the webview feature of the mobile application.

Table 5.1: List of requirements to design a tool that can raise awareness of OGD among users. OGD reuse refers to the application or service developed using OGD.

Requirement	Requirement description	Implementation in CitizenApps	Added value compared to existing applications mentioned in section 5.2.2.1
R7. Give feedback on OGD reuse	Possibility to provide feedback to the OGD reuse's developer in order to improve the OGD reuse.	Possibility to rate and let comments on OGD reuse.	This feature does not exist in the existing applications.
R8. Share OGD reuse	Possibility to inform friends about the existence of a specific OGD reuse.	Possibility to share by email a specific OGD reuse.	This feature does not exist in the existing applications.
R9. Quick access to favorite OGD reuses	Possibility to facilitate access to the favorite OGD reuses of the user.	In order to achieve this requirement, CitizenApps first allows users to create a free account. Then once logged in, users can then add an OGD reuse to their favorites, so they will not need to go through a search every time before running their favorite reuses.	This feature does not exist in the existing applications.
R10. Receive notification	Possibility to receive notification when new OGD reuses are published.	Ability to receive notification when new OGD reuses are added in ODEON platform.	This feature does not exist in the existing applications.
R11. Easy to use and intuitive	Facilitate the use by users even with low technical skills.	Organize the functionalities into a pleasing layout after collecting feedback of users and an UX expert.	-

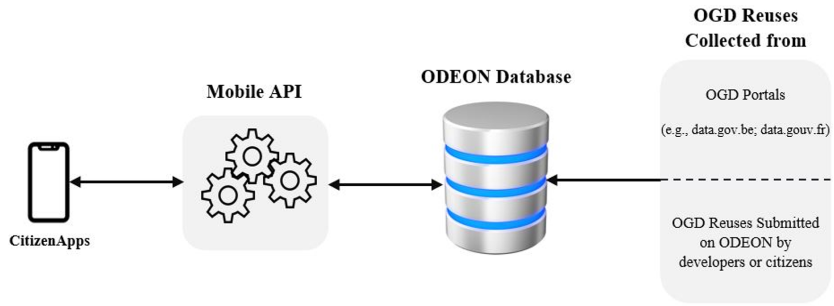


Figure 5.3: Architecture of CitizenApps. OGD reuses collected from governmental portals or reuse submissions are first stored in ODEON Database. Then, when the user launches “CitizenApps”, the system first calls the mobile API to retrieve the necessary information.

The mobile application was built using the following technologies: Java (Programming language) and OneSignal[8] (notification plugin). Figure 5.3 presents the architecture of CitizenApps. The mobile application was supported by an API built on top of the ODEON platform (which will be discussed further in Section 5.4.3.2), a web application that aggregates several OGD reuses from different countries and also allows developers to register their OGD reuses themselves. Thus, once the OGD reuse is added to ODEON, it will also appear directly in the mobile application and a notification will be sent to users to inform them of the newly added OGD reuse. To date, the mobile application has 30 reuses from 4 countries: Belgium, France, Austria and Bulgaria. These reuses were selected based on data of proven interest to users in (Chokki et al., 2021), such as data on transport, finance and health to name a few. However, it is important to note that other countries and reuses can be added into CitizenApps through ODEON. Screenshots of the prototype are shown in Figure 5.4.

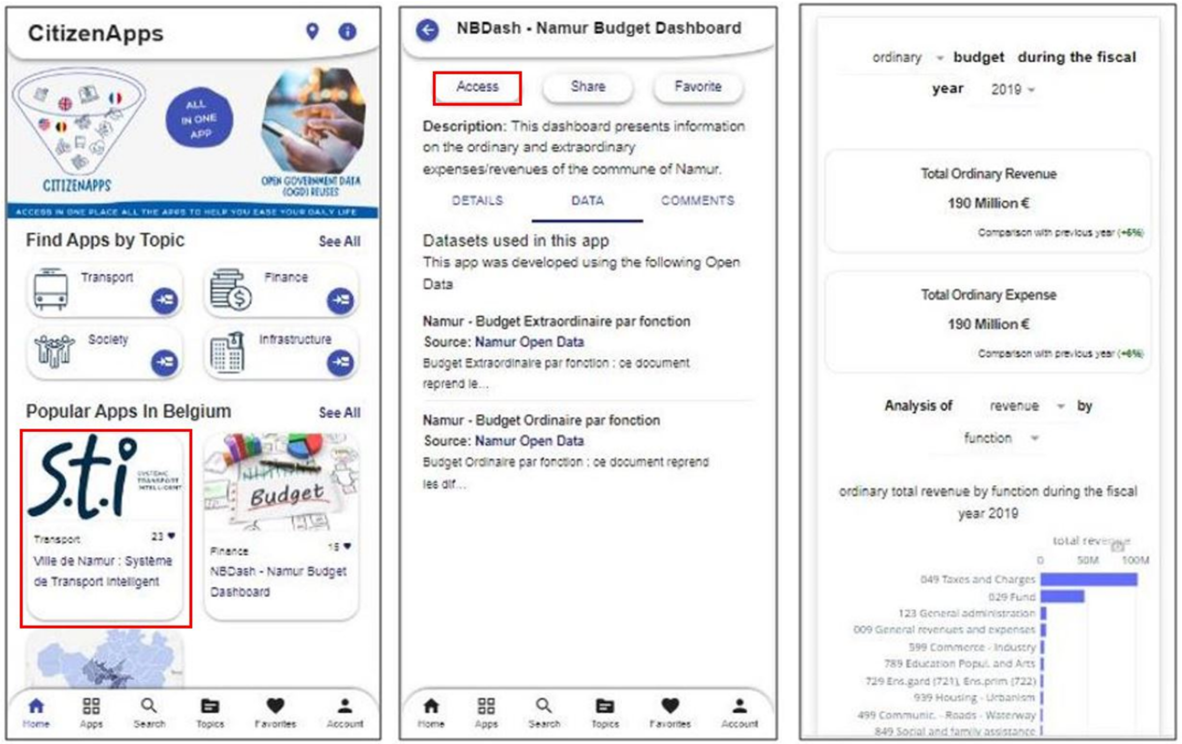
5.3.2.2 Evaluation Results

Through the surveys that participants filled out after exploring the mobile "CitizenApps" application, we were able to gather their opinions. A total of 25 participants completed the surveys. All participants are between the ages of 18 and 50 and have at least a high school degree.

Table 5.2: Median, mean and standard deviation (SD) of survey scores.

	Attitude	**Perceived usefulness**	**Perceived ease of use**	**Behavioral intention**
Median	6	5	6	5
Mean (SD)	5.93 (0.59)	5.47 (0.74)	5.93 (0.59)	4.9 (1.19)

[8] https://onesignal.com/



(1) HomePage (2) Details of an App (3) Run App

Figure 5.4: Screenshots of CitizenApps. (1) shows the homepage where we list the most important topics and some popular applications; (2) shows the details of an application when the user clicks for example on "NBDash – Namur Budget Dashboard" from the homepage; (3) shows the live application when the user clicks on the "Access" button.

Table 5.2 presents the median, mean and standard deviation of the questions with a 7-point Likert scale regarding the 4 aspects (attitude, perceived usefulness, perceived ease of use, behavioral intention) evaluated for the prototype. The following conclusions can be drawn from the results of Table 5.2. Users' attitude towards the use of CitizenApps is favorable in general as evidenced by the median & mean attitude ≥ 5 and the standard deviation is low (SD = 0.59) showing that users' scores tend to be close to the mean of users' scores. Citizens also agree that CitizenApps is useful to them in better understanding the existence of OGD and its usefulness (median & mean ≥ 5 for perceived usefulness and users' scores tend to be close to the mean of users' scores (SD = 0.74)). This is also reinforced by the fact that before showing CitizenApps to users, we asked them "what was their perception of the usefulness of OGD to the public?" Only two participants were able to give us a clear example of how OGD can be used in real life. However, after the CitizenApps presentation, all participants agree on their usefulness and some were surprised by the fact they are already using them without knowing it. Users also find CitizenApps easy to use (median & mean ≥ 5 for perceived ease of use and users' scores tend to be close to the mean of users' scores (SD = 0.59)). This is justified by the fact that many users think that CitizenApps was well designed and user-friendly. However, users' intention to use CitizenApps was somewhat lower than previous aspects but was acceptable (median & mean ≥ 4 and SD = 1.19 for behavioral intention). This may be due to the fact that some users see only a few applications that interest them in CitizenApps, as only 30 applications have been added so far for testing purposes.

Regarding the feedback on implemented features in CitizenApps, all participants agreed that the features implemented should be kept. However, 4 participants suggested removing the login feature that was required to address R9 (quick access to favorite OGD reuses), as its added value is only to help users to add certain applications to their favorites. Apart from that, participants did not suggest any additional features, as many of them think that the existing features are already sufficient and it is better not to add any more to avoid making the application difficult to use. In terms of features that could be improved, participants suggested adding more applications to CitizenApps to increase the interest of the CitizenApps for users.

5.3.3 Conclusion

The purpose of this study was to identify the requirements for an application that raises awareness of OGD to users. To achieve this objective, we conducted interviews with 10 users to collect requirements that need to be implemented in such an application. Next, we implemented the requirements in an application we called "CitizenApps". Finally, we examined whether the identified requirements can raise awareness of OGD through an evaluation conducted with 25 users using the implemented application as proxy.

This study contributes to theory in the following aspects. First, we contribute to the identification of users' requirements for the design of an application to improve OGD awareness. A total of 11 requirements (see Table I) are identified. To the best of our knowledge, this study is the first study to identify users' requirements for the design of an application to improve OGD awareness. These identified requirements can be used by developers or practitioners to implement a usable tool to raise user awareness or used by researchers to evaluate applications whose objective is to raise user awareness. Second, the identified requirements are formulated in a generic manner and thus can be implemented in a variety of ways. Third, the findings of the evaluation provide preliminary evidence that the use of an application implementing the identified requirements can improve user awareness of the existence and usefulness of OGD.

This study also contributes to practice in the following aspects. First, we present through a concrete tool (CitizenApps) how to implement the different requirements (see Table 5.1). Thus, the implemented mobile application as well as the description of the implementation of each requirement can be a source of inspiration for developers or OGD managers on how they can implement the requirements in their existing or new platforms in order to improve OGD awareness. Second, since CitizenApps is a federated application that can include OGD reuses (collected from developers or external platforms such as OGD portals), the implemented mobile application can be used by developers to showcase or promote their OGD reuses or by researchers to explain to users the usefulness of OGD with real-world examples. Third, CitizenApps addresses the shortcomings of OGD portals and topic-based applications (e.g., Datafruit, WallonieEnPoche[9]) by integrating the following requirements: **R2**. Categorize OGD reuses (especially by country), **R5**. See details about OGD reuse (especially the list of datasets used in each OGD reuse), **R7**. Provide feedback on OGD reuse, **R10**. Receive notification (especially when there is a new OGD reuses in the user area) (see Table 5.1).

This study has some limitations that will need to be addressed in future work. The first limitation concerns the representativeness of the participants in the both surveys. From a statistical point of view, we note that there are limits to the conclusion validity. The number of participants may be small, especially for the evaluation survey, but referring to previous studies (Nielsen, 2000; Faulkner, 2003), the use of at least 5 participants for usability tests is a good baseline. However, to increase this representativeness, we suggest using other communication channels or collecting data on-site in universities or public places. In this study, this was not feasible due to the COVID-19 situation. The second limitation is the use of convenience sampling method to recruit participants, this may have a bias on the validity of these findings however we tried to reduce that bias by also recruiting participants through social media. Future work will be to improve the validity of the identified requirements with a large number of participants from different backgrounds by using a random sampling method, which was not possible in our case due to the small number of participants responding to the survey. The third limitation is that this research focuses only on the technical aspects of the issue related to the user awareness of OGD. Other researchers can extend this work by integrating the social, motivational and emotional aspects of the issue, which will be necessary to address the previous limitation.

5.4 Improving Communication and Collaboration between OGD stakeholders

While raising awareness is an essential step in the data reuse process, it does not guarantee extensive usage. In this section, we will address another potential reason for this lack of utilization, which is the lack of communication and collaboration among different stakeholders involved in open government data (i.e., citizens, developers acting as infomediaries, and publishers), with a focus on the technical aspect of this gap. In this study, "communication and collaboration" refers to the interaction between OGD stakeholders in order to create innovative OGD-based services for citizens.

Various methods and platforms have been used in the literature (Gebka et al., 2019; Crusoe et al., 2020; Gil et al., 2022) to address the technical challenges of bridging the gap. However, none of them offer a complete set of functionalities

[9] https://wallonie.enpoche.be/

required to achieve this objective. Furthermore, to the best of our knowledge, no study has been conducted to provide a comprehensive list of requirements for designing a usable tool that facilitates communication and collaboration between OGD stakeholders. This section aims to fill these gaps by initially proposing a list of requirements for designing a usable tool that enables effective communication and collaboration among OGD stakeholders, followed by the implementation of these requirements in a tool and its subsequent evaluation.

The remainder of this section is organized as follows: Section 5.4.1 provides background information on the barriers to OGD use and publication, as well as existing methods and platforms for communication and collaboration among OGD stakeholders. Section 5.4.2 outlines the design science research approach adopted in this study. The developed tool and its evaluation are presented in Section 5.4.3. Finally, Section 5.4.4 summarizes the contributions of this study, highlights its limitations, and suggests future research directions.

5.4.1 Background

In this section, we start by examining earlier research that emphasizes the difficulties stemming from insufficient communication and collaboration among stakeholders involved in open government data. These challenges were subsequently employed to formulate the necessary requirements for developing an effective tool that fosters communication and collaboration among these stakeholders. Following that, we investigate different methods and platforms proposed in existing literature to improve the communication and collaboration among these stakeholders.

5.4.1.1 Impediments to OGD Use and Publication

Several studies (Janssen et al., 2012; Gebka et al., 2020; Crusoe et al., 2019; Zuiderwijk et al., 2012; Beno et al., 2017; Polleres et al., 2017) have focused on the impediments to OGD use and publication. Below, we present the impediments resulting from the technical aspects of the lack of communication and collaboration between OGD stakeholders that demotivate them to use and to publish OGD.

From the citizens' perspective, we identified two main impediments. First, **(IC1)** many citizens are not aware of the existence and usefulness of OGD and the services that use them (Janssen et al., 2012; Gebka et al., 2019). Second, **(IC2)** the services developed with OGD are not used by citizens due to either unawareness of the services or mismatch between the services and their needs. Indeed, most of the time, they are not involved in the service design (Gebka et al., 2019; Crusoe et al., 2019).

From the developers' perspective, we identified six main impediments. First, **(ID1)** developers are not able to come up with a service idea that may be interesting for citizens due to a lack of knowledge of citizens' needs (Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013; Beno et al., 2017; Polleres et al., 2017). Second, **(ID2)** developers are not able to find examples of use case or success stories of OGD use to build on to propose services that may be interesting to citizens (Crusoe et al., 2019). Third, **(ID3)** once developers have an idea of the service to implement, one of the issues is the low quality of the datasets required for the service's realization and the difficulty to communicate with the publisher to solve this issue (Janssen et al., 2012; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013; Beno et al., 2017; Polleres et al., 2017). Fourth, **(ID4)** developers are sometimes unable to find the datasets needed to create the service due to the non-publication of these datasets by OGD providers and also the lack of discussion between the two parties to request the necessary datasets (Janssen et al., 2012; Crusoe et al., 2019; Zuiderwijk et al., 2012; Chokki et al., 2021). Fifth, **(ID5)** developers often do not receive feedback from publishers after requesting datasets or asking some questions (Janssen et al.,

2012; Beno et al., 2017). Sixth, **(ID6)** developers often have no information about the datasets used in a specific project in order to replicate or improve it (Matheus et al., 2020; Sarikaya et al., 2019; Kitchin and Mcardle, 2016).
From the publishers' perspective, we identified three main impediments. First, **(IP1)** publishers are not motivated to publish data because most published datasets are not used, making the added value and economic impact of their publication efforts uncertain (Polleres et al., 2017). Second, **(IP2)** publishers are unaware of the datasets that they need to prioritize during the publication process (Gebka et al., 2019, 2020). Third, **(IP3)** publishers do not know which projects are using their published datasets, as many users do not report their reuses (Polleres et al., 2017).

The impediments, repeatedly reported in the literature, highlight the need for increased communication and collaboration among OGD stakeholders.

5.4.1.2 Methods and Platforms of Communication and Collaboration between OGD Stakeholders

Many methods and platforms have been proposed in the literature to facilitate the communication and collaboration between OGD stakeholders (Simonofski et al., 2020). Reviewing each of them would be beyond the scope of this study. Here, we focused on these popular (i.e. most cited or used in an OGD context) methods and platforms: hackathons, interviews, workshops, OGD portals and citizen participation platforms. Many existing platforms such as ADEQUATE (Neumaier et al., 2018), Github, Gitlab, Stack Overflow, gFoge, Jira, Redmine, Wiki, etc. can enable communication and collaboration between users. However, in this study, we only reviewed platforms that are related to our research question i.e., that have already been used in previous studies to facilitate communication and collaboration between users in the context of the OGD process and that clearly distinguish the role of each user in this process. Table 5.3 presents the methods and platforms reviewed, their strengths, and weaknesses.

All the gaps mentioned in the existing methods show that there is currently no method or tool that supports collaboration between OGD stakeholders adequately. This justifies the need to identify the necessary requirements such a tool should satisfy and to implement one fulfilling them. Therefore, we implemented ODEON, which differs from the methods and platforms reviewed in that it addresses each of their shortcomings. It is important to note that ODEON does not aim to replace existing OGD portals but rather proposes complementary communication features. Similarly, ODEON complements citizen participation platforms by allowing to link each service idea to some datasets available on OGD portals that can help to implement the idea.

5.4.2 Research Methodology

The Design Science Research (DSR) method (Hevner et al., 2004; Dresch et al., 2015) was used to address the research question of this section **(RQ1c)**: **"What elements are crucial for a tool that enhances communication and collaboration among OGD stakeholders?"**. It aims to develop solutions (design cycle) that meet defined objectives, contribute to the scientific knowledge base (rigor cycle) and provide utility in the environment (relevance cycle).

In the **rigor cycle**, from July to September 2021, we conducted a literature review to access existing knowledge on the impediments resulting from the technical aspects of the lack of communication and collaboration between OGD stakeholders. These impediments were extracted from 12 articles returned by a search performed on the databases "Scopus" and "Science Direct" with the keywords ("open government data" OR "open data") AND ("feature" OR "impediment" OR "barrier") AND ("reuse" OR "use" OR "re-use" OR "publication" OR "communication" OR "collaboration"). The results of the literature review are presented in Section 5.4.1.1. Once this step was completed, we used the identified impediments to formulate general requirements that should be implemented in a usable tool to solve these impediments. Next, interviews were conducted with 9 stakeholders (1 publisher, 4 developers and 4 citizens) to validate the suggested requirements and gather additional ones, if any. The interviews were handled in four steps. First, we briefly introduced participants to the utility of OGD with a concrete case of the use of OGD by the city of Namur, which developed an Intelligent Transportation System[10] based on the OGD it published. Second, we explained the context of our project by presenting the benefits and problems related to the lack of communication between OGD stakeholders. Third, we asked participants to suggest some requirements that can address these mentioned problems. Fourth, we showed participants the requirements we extracted from our literature review and asked them, on a scale of 1 (Totally irrelevant) to 5 (Totally relevant), how relevant they were and to justify their choice. Appendix C.1 summarizes the questions asked during the interviews. We took note of the participants' feedback and recorded the discussion with their agreement for transcription and later review. The interviews were later coded using short sentences to retain context and conceptual relations. Finally, we used the results of these interviews to provide a list of requirements that need to be incorporated into a usable tool to facilitate communication and collaboration among OGD stakeholders.

In the **design cycle**, we used the validated requirements to propose a list of features that were then implemented in a tool called ODEON (Open Data Exchange solutiON). Once the features were implemented, we presented the tool to two stakeholders (a developer and a citizen) to get their feedback and integrate it before the evaluation phase.

[10]https://sti.namur.be/

Table 5.3: Strengths and weaknesses of methods and platforms of communication and collaboration between OGD stakeholders.

Methods/Platforms	Strengths	Weaknesses
Hackathons (Gebka et al., 2019; Simonofski et al., 2020)	(1) enable developers to design, implement and present services for a specific issue, beyond the "product idea" level (2) physically reunite developers and publishers for several hours to several days, creating many opportunities for discussion (3) usually very focused on technological output, developers tend to oversee other aspects than code due to competition under tight schedule	(1) little or no consideration of citizens' needs (2) access limited to a certain type and number of participants (3) communication and collaboration between these stakeholders is ephemeral (e.g., after the event, there is no possibility to report issues with the datasets or to request datasets) (4) no archiving option that allows, for example, other citizens or developers to know which reuses were implemented during the event (5) impossible to get feedback from participants who were not at the event on the solutions implemented
Interviews and workshops (Gebka et al., 2019, 2020; Chokki et al., 2021; Crusoe et al., 2020; Barbosa Tavares et al., 2011)	allows different OGD stakeholders to physically discuss to collect their needs or feedback	(1) access limited to a certain type and number of participants (2) communication and collaboration between these stakeholders is ephemeral
OGD Portals (Janssen et al., 2012; Beno et al., 2017)	(1) help local governments to publish and manage open data on the web (2) allow developers to submit their reuses or to see existing reuses	(1) can only handle reuses and datasets from the specific portal (2) citizens' needs are not considered (3) no common space to facilitate discussion among stakeholders (4) no archiving of the data issues and their status (5) developers are not able to collect feedback about the reuses they have submitted (6) publishers are not able to know which datasets to prioritize for publication

Table 5.3: Strengths and weaknesses of methods and platforms of communication and collaboration between OGD stakeholders.

Methods/Platforms	**Strengths**	**Weaknesses**
Citizen participation platforms (e.g., Citizenlab[11], Leuven make it happen[12]) (Gil et al., 2022)	(1) allow local governments to interact with the public (2) allow stakeholders to propose project ideas (3) allow stakeholders to vote on projects they are interested in (4) allow stakeholders to track the status of their projects (e.g., rejected or not)	(1) not focused on promoting the use of open data (2) unable to integrate publishers and developers into the project process (3) unable to track project progress after approval (4) unable to record existing reuses of open datasets (5) unable to link open datasets to projects (6) unable to request data to publishers (7) unable to report issues on open datasets (8) unable for publishers to know which datasets have priority for publication

[11] https://www.citizenlab.co/
[12] https://leuvenmaakhetmee.be/

In the **relevance cycle**, we evaluated the prototype through a user test with 22 stakeholders (5 publishers, 8 developers and 9 citizens) to assess its ease of use and usefulness in addressing the identified impediments, and to gather additional features for future versions. The publishers were recruited through contact forms available on their OGD portals. As for the developers and citizens, they were recruited through the following communication channels: UNamur mailbox and social media. The interviews were handled in four steps. The first two steps are identical to those of the rigor cycle interviews. In the third stage we briefly presented the prototype's features to the participants in 10 minutes, and then gave them 20 minutes to perform scenarios related to their profile with ODEON. Citizens were invited to (1) explore existing projects, (2) suggest a new project with the content of their choosing, and (3) provide feedback on two projects of their choosing. Developers were asked to (1) explore existing projects, (2) suggest a new project with the content of their choosing, (3) request a dataset of their choosing they would be needed to develop the project they suggested in the second scenario, (4) report an issue with an existing dataset of their choosing, and (5) provide feedback on two other projects and two other datasets of their choosing. Finally, publishers were asked to (1) explore existing projects and (2) provide feedback on two projects and two dataset issues of their choosing. We encouraged them to perform think-aloud as they explored to gather qualitative data on their overall feeling and expectations. They were also given the opportunity to ask the interviewer questions if necessary. However, in accordance with user testing guidelines (Lallemand and Gronier, 2015), the sequence of actions to perform the scenarios was not given to participants. Next, we collected participants' feedback through a questionnaire consisting of three types of questions: questions with a 7-point Likert scale (from "Strongly Disagree" to "Strongly Agree") based on the Technology Acceptance Model (TAM) (Davis, 1989) to assess the ease of use of prototype functionalities, the ease of use and usefulness of the prototype in solving the identified impediments, open-ended questions to gather general opinions and suggestions for additional features for future versions and to explain quantitative ratings, and questions on the respondent's profile. The questionnaire was pretested with two people to ensure that all kinds of errors that are associated with survey research are reduced [29]. Appendix C.2 presents the questions contained in the questionnaire according to the stakeholder roles. After collecting participants' feedback, the median, mean and standard deviation were calculated for the Likert questions to evaluate the following aspects: (A1) ease of use of each prototype feature, (A2) overall ease of use of the prototype, (A3) usefulness of each prototype feature and (A4) overall usefulness of the prototype for facilitating communication and collaboration between OGD stakeholders. The A1 (respectively A3) questions were used as a reference to understand the answers to the A2 (respectively A4) questions. These statistical measures were chosen because they are the most appropriate for analyzing Likert data and for having a central tendency measure (Boone and Boone, 2012). Verbal thoughts and responses collected from the free text questions were coded using short sentences to retain context and conceptual relations.

5.4.3 Results

In this section, we first present the requirements that need to be incorporated into a usable tool to facilitate communication and collaboration among OGD stakeholders. Next, we describe how the ODEON prototype was implemented to meet these requirements. Finally, we present the results of the evaluation of ODEON.

5.4.3.1 Requirements Identification for Communication and Collaboration between OGD Stakeholders

Based on the identified impediments, we derived several requirements that were confronted with the participants during the rigor cycle interviews. Table 5.4 presents the requirements along with insights from the literature and interviewees.

Table 5.4: List of requirements to design a tool that can facilitate the communication and collaboration between the OGD stakeholders along with insights from the literature and interviewees. "All" means that all participants were agreed that the requirement is relevant.

Requirements (C = citizens, D = developers, P = publishers)	**Insights from Literature**	**# Participants**
RC1. Inform citizens of existing projects based on OGD	IC1	All
RC2. Allow citizens to be involved in the service development process	IC2	All
RC3. Allow citizens to propose project ideas	IC2, ID1	All
RC4. Allow citizens to register existing projects	/	1
RC5. Allow citizens to define which requested projects should be prioritized	/	4
RD1. Allow developers to register existing projects	ID2	All
RD2. Inform developers about existing projects based on OGD	ID2	All
RD3. Allow developers to report issues related to the use of published datasets	ID3	All
RD4. Allow developers to request datasets that do not exist on the portals	ID4	All
RD5. Allow developers to provide feedback on projects and datasets	ID5	All
RD6. Inform developers about the datasets used in projects	ID6	All
RD7. Allow developers to propose project ideas	/	1
RD8. Allow developers to define which requested projects and data should be prioritized	/	4
RP1. Inform publishers about existing projects based on OGD	IP1	All
RP2. Inform publishers of priority data to be published	IP2	All
RP3. Inform publishers of projects using their datasets	IP3	All

5.4.3.2 ODEON System Description

We designed features that could meet the collected requirements and then implemented them in the prototype ODEON (source code available[13]). The following paragraphs explain how the implemented features meet the different requirements.

[13]https://github.com/chokkipaterne/odeon

Provide a form to register a project (RC4 and RD1). It allows developers and citizens to register an existing project in two steps. First, they fill in the following information: country, city, domain, title, access link, description, image and contact information. Then, they can add the datasets used in the project in three ways: by searching and selecting datasets directly from the OGD portals, by uploading files, or by using external links. To avoid duplicate entries, an auto-completion feature for the title field was added, as suggested by an interviewee. Regarding the addition of datasets related to a project from the OGD portals, ODEON provides a search option that allows users to search and select the desired dataset directly from any CKAN or OpenDataSoft portal using the APIs provided by these two systems.

Provide a form to suggest a project (RC3 and RD7). The form to be filled in is similar to the project registration form, except that the access link is not requested, and developers and citizens can also skip the step of adding or requesting datasets for the project.

Display the list of projects (RC1, RD2 and RP1). It displays all registered (existing and requested) projects.

Search projects (RC1, RD2 and RP1). It allows stakeholders to search for specific projects among the registered projects based on the following criteria: keywords, domain, project type (requested or existing), country and city.

Display details of a specific project (RD6). It allows stakeholders to see all information about a specific project: general information (title, description, country, comments, etc.) and data used or requested in the project.

Display details of a specific dataset (RP3). It allows stakeholders to see all information about a specific dataset: general information (title, description, country, comments, etc.) and projects that use the dataset.

Provide a form to request data (RD4). It allows developers and citizens to request data by filling in the following information: country, city, title, description and contact information.

Provide a form to comment a specific project (RC2 and RD5). It provides two options. The first option allows any user to provide general feedback on the project or respond to an existing comment. The second option allows developers to inform citizens of the progress of the project development.

Provide a form to comment a specific data (RD3). It provides two options. The first option allows any user to provide general feedback on the data or respond to an existing comment. The second option allows stakeholders to report data issues. The form has the following fields: comment type (general comment or report data issue), name, comment and attachment.

Display and search requested data (RP2). It lists all the requested data. There is also a search option that allows filtering the display using the following options: country, state and keywords.

Display and search data issues (RD3). It lists all data issues reported by stakeholders. There is also a search option that allows filtering the display using the following options: country, state and keywords.

Add a voting option for requested projects and data (RC5 and RD8). It consists of adding a “Like” button for each requested project or data so that any user can click on the button to indicate that the requested project or data is relevant.

ODEON is a web application built using the Django framework. Figure 5.5 shows two pages of the prototype. (A) represents a portion of the homepage where the stakeholders can make a quick search of project or can access other features of the prototype based on their profile (citizen, developer or publisher) and (B) represents a portion of the project list page with available search options (keywords, topic, project type and location).

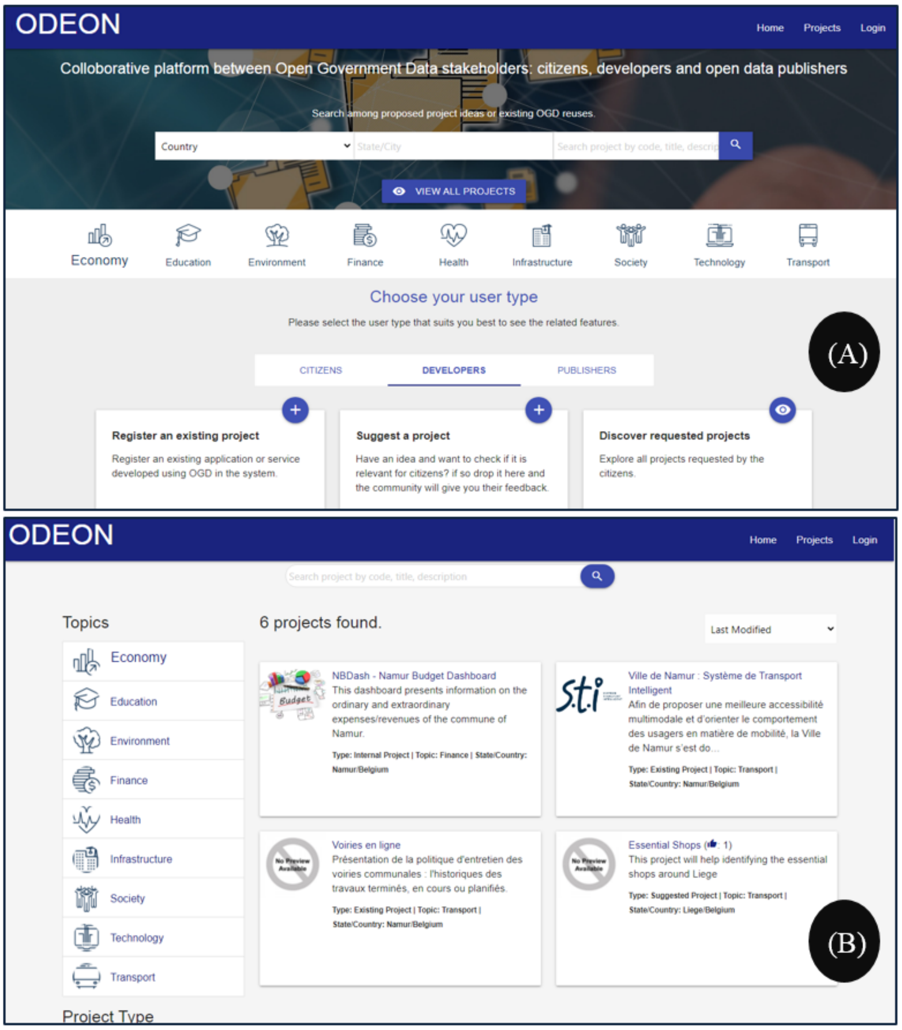


Figure 5.5: Two pages of ODEON. (A) represents a portion of the homepage and (B) represents a portion of the project list page with available search options.

5.4.3.3 Evaluation

In total, 22 participants (9 citizens, 8 developers and 5 publishers) participated in the evaluation of ODEON and completed the questionnaire. Table 5.5 presents the median, mean and standard deviation (SD) for the Likert questions regarding the ease of use and the usefulness of the prototype. The following conclusions can be drawn from the results of Table Table 5.5, for the different stakeholders:

Citizens. Most of the citizens agreed that the proposed prototype was easy to use (median and mean ≥ 5 for A2 with a low standard deviation around 1) and useful for facilitating the communication and collaboration between OGD stakeholders (median and mean ≥ 5 for A4 with a low standard deviation around 1). More specifically, many citizens found that the prototype met their expectations and was user-friendly for the features such as discovering existing projects, registering and suggesting projects, but had a more mixed opinion about the monitoring project development and reporting data issue features. Indeed, the means of participants' scores on the ease of use (A1) and usefulness (A3) questions ranged between 3 and 5 for these features.

Developers. Most of the developers also found that the proposed prototype was

easy to use (median and mean ≥ 5 for A2 with a low standard deviation around 1) and useful for facilitating the communication and collaboration between OGD stakeholders (median and mean ≥ 5 for A4 with a low standard deviation around 1). However, like the citizens, they found it more difficult to use for some features such as updating the status of the project and reporting data issue. The means of participants' scores for these features were less than 5.
Publishers. The publishers were quite satisfied with the proposed prototype and found it easy to use (median and mean ≥ 5 for A2 with a low standard deviation around 1) and useful for facilitating the communication and collaboration between OGD stakeholders (median and mean ≥ 5 for A4 with a low standard deviation around 1). Compared to the other two OGD stakeholders, the publishers are concerned by only three features of ODEON and all these features were easy to use and useful for them (median and mean ≥ 5 for A1 & A3 questions). However, they were concerned that the comments would pile up and that they would get lost in them.

Table 5.5: Median, mean and standard deviation (SD) of ease of use and usefulness questions.

	Citizens (N = 9)		**Developers (N = 8)**		**Publishers (N = 5)**	
	Median	**Mean (SD)**	**Median**	**Mean (SD)**	**Median**	**Mean (SD)**
A2. Ease of use of the prototype	5	5.19 (1.13)	5.5	5.25 (1.18)	6	5.7 (0.82)
A4. Usefulness of the prototype	5	5.36 (1.27)	5	5.67 (1.14)	6	6 (1.41)

In addition to these findings, some new features and suggestions for improving the UI design were gathered from the verbal thoughts and answers to the open-ended questions provided by the participants. The new features are as follows. First, citizens and developers suggested having a feature that helps them to subscribe to a dataset or project to get weekly updates on the data or projects they have subscribed to. Second, publishers suggested having a monthly summary of requested datasets or data issues directly in their inbox. Third, for the discovery functionality of existing and requested projects, participants suggested moving the project type and location filters close to the search field to make them more visible. Fourth, participants suggested adding a status attribute for projects to help to identify the current status of each project: requirements analysis, under development, development complete and abandoned project.

5.4.4 Conclusion

The aim of this study was to address the technical aspects of the lack of communication and collaboration between OGD stakeholders (citizens, developers and publishers). To achieve that goal, we first identified the impediments resulting from the technical aspects of the lack of communication and collaboration between OGD stakeholders that demotivate them to use and to publish OGD through a literature review. Then, through interviews with 9 stakeholders, we deducted and validated, the requirements that should be satisfied by a usable tool that addresses these impediments. Next, we implemented the ODEON prototype based on the collected requirements and used it as proxy to measure the extent to which the requirements were easy to use and useful to facilitate the communication and collaboration between OGD stakeholders through an evaluation conducted with 22 stakeholders from the three different profiles.

This study contributes to theory in the following aspects. First, it extends previous studies related to impediments (Janssen et al., 2012; Gebka et al., 2019, 2020; Crusoe and Melin, 2018; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013; Beno et al., 2017; Polleres et al., 2017), building on the previously identified impediments to propose a list of 16 requirements that should be implemented in a usable tool to facilitate the communication and collaboration between OGD stakeholders. These identified requirements can also be used as reference by open data publishers or developers of citizen participation platforms to help them to know what features need to be added to their existing platforms to fully facilitate communication and collaboration between OGD stakeholders. They can also be used by researchers to evaluate applications whose objective is to facilitate communication and collaboration between OGD stakeholders. Second, unlike previous studies (Janssen et al., 2012; Gebka et al., 2019, 2020; Crusoe and Melin, 2018; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013; Beno et al., 2017; Polleres et al., 2017) that focused only on a specific user profile or only on impediments, this study validates the identified impediments as well as the requirements through interviews with three different stakeholder roles. Third, the results obtained from the evaluation show that ODEON and the implemented features were easy to use and useful to address the technical aspects of the lack of communication and collaboration between OGD stakeholders.

This study also contributes to practice in the following aspects. First, unlike previous studies (Janssen et al., 2012; Gebka et al., 2019, 2020; Crusoe and Melin, 2018; Crusoe et al., 2019; Zuiderwijk et al., 2012; Martin et al., 2013; Beno et al., 2017; Polleres et al., 2017) that identified impediments without providing a list of features of a usable tool to address them, we derived a list of features from the requirements for each OGD stakeholder, and we present it as the use case diagram shown in Figure 5.6. In this diagram, features are subdivided into data features and project features. The data features address the following impediments: data quality and lack of awareness of which data should be released first. The project features resolve other identified impediments, such as unawareness of the existing projects based on OGD, difficulty finding a project idea, lack of citizen involvement in project development, etc. Second, we provide access to the source code of ODEON[13]. This can be used as a starting point by developers to create their own online tool for facilitating the collaboration between OGD stakeholders or to improve the prototype. The use case diagram can be used as starting point as well.

However, this study has some limitations that will need to be addressed in future work. The first limitation concerns the representativeness of the participants in the evaluation. The number of participants may be small, but referring to previous studies (Faulkner, 2003; Nielsen, 2000) 5 participants is a good baseline for usability tests and we also observed that our findings were reaching saturation at that point. However, to increase representativeness, we suggest using other communication channels or collecting data on-site in administrations, universities, workshops, hackathons or public places to recruit participants for the evaluation of the future prototype version. In this study, this was not feasible due to the COVID-19 situation. The second limitation is that we did not consider the discovery step of ODEON. Indeed, a multi-stakeholder collaboration platforms relies heavily on a community that needs to be attracted on the platform. One approach to address this issue will be to communicate about this prototype using social networks and by presenting it at open data workshops and hackathons. The third limitation is the non-generalization of the proposed requirements to other areas. Other researchers can start by investigating whether the existent platforms (e.g., Github, Wiki) can be adapted to the context of open data or whether the proposed requirements can be used or extend to other areas such as open data ecosystems, open government ecosystems, and open source software ecosystems. Future work will also include an implementation of the suggested new features and a field evaluation of ODEON, for example, by offering hackathon organizers to use it as meeting point between the stakeholders involved in the event.

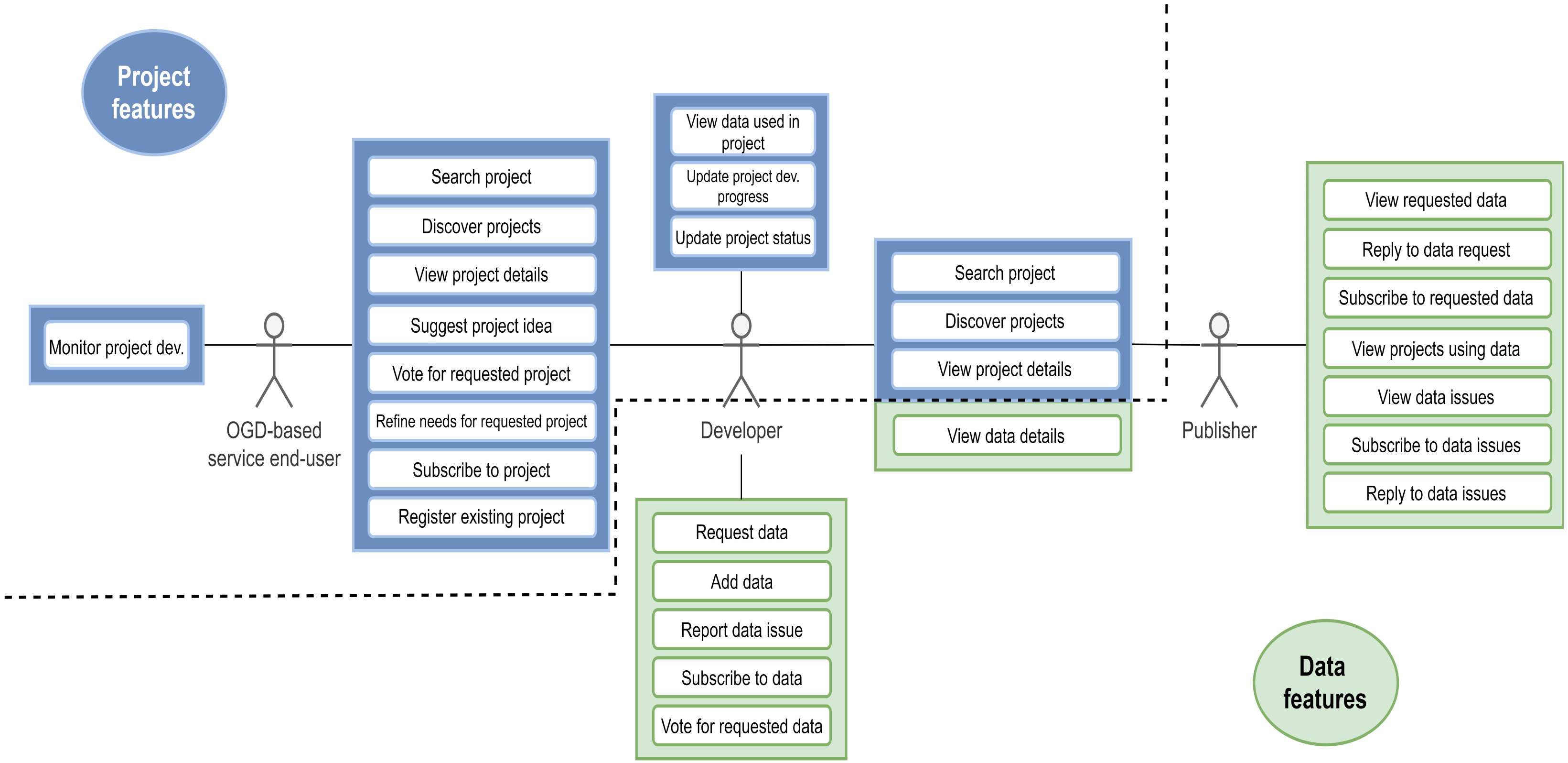


Figure 5.6: Use case diagram of ODEON. The features are subdivided into two groups: the features related to the project (project features) and the features related to the data (data features).

5.5 Wrap up

In this chapter, we investigated various aspects related to raising awareness and enhancing communication and collaboration among OGD stakeholders. We made significant contributions in three sections, each addressing a specific aspect of the overall objective.

In Section 5.2, we focused on identifying effective methods for raising awareness and highlighting the usefulness of OGD to citizens. Through a literature review and an online survey involving 30 citizens, we compared existing methods with citizens' perceptions. The survey results, combined with the literature review, enabled us to provide recommendations for governments. We suggested the use of public outreach campaigns, applications, and leveraging the power of word-of-mouth to inform citizens about the existence and benefits of OGD. We recognized applications as a preferred method and identified a research gap regarding the requirements for a usable tool to promote OGD, which we plan to address in Section 5.3.

In Section 5.3, our objective was to identify the specific requirements for an application that raises awareness of OGD to citizens. To accomplish this, we conducted in-depth interviews with 10 citizens to gather their valuable insights and perspectives. These interviews allowed us to iteratively refine and validate a comprehensive list of 11 key requirements, which we subsequently implemented in an application we named "CitizenApps." Through an evaluation involving 25 citizens, we assessed the efficacy of the implemented requirements in raising awareness of OGD. Notable requirements that emerged include categorizing OGD reuses, providing detailed information on OGD reuse, enabling feedback mechanisms, and facilitating notifications. The findings from this section serve as a fundamental foundation for developers and governments to construct effective tools that enhance citizen awareness of OGD. Additionally, researchers and application reviewers can employ these findings as a framework for evaluating similar applications in the future.

Section 5.4 delved into the technical aspects of addressing the lack of communication and collaboration between OGD stakeholders, namely citizens, developers, and publishers. Through an extensive literature review and interviews with 9 stakeholders, we identified the obstacles and challenges resulting from this communication gap. By meticulously analyzing these impediments, we derived a set of requirements that a usable tool should meet to effectively address these challenges. Building upon these requirements, we developed the ODEON prototype and conducted evaluations involving 22 stakeholders from different profiles. The evaluation results allowed us to propose and validate a comprehensive list of 16 requirements, accompanied by a detailed use case diagram outlining the features that should be integrated into a usable tool. These features encompass both data-related functionalities, such as reporting data issues and responding to data requests, as well as project-related functionalities, such as suggesting and monitoring project development. The evaluation outcomes highlighted that publishers expressed higher satisfaction with the prototype compared to citizens and developers. However, valuable feedback from all stakeholders contributed to future improvement plans, ensuring the continuous enhancement of the tool's usability and effectiveness in facilitating communication and collaboration among OGD stakeholders.

Overall, this chapter provides a comprehensive understanding of the methods for raising awareness, the requirements for an awareness-raising application, and the technical aspects of communication and collaboration in OGD initiatives. The recommendations, requirements, and prototype developed in this chapter contribute to the advancement of OGD initiatives and provide guidance for developers, governments, and researchers alike.

CHAPTER 6

Tackling Barrier 2: Assessing (Meta)data Quality of OGD

6.1 General Introduction

This chapter presents the research conducted to address the second barrier, which pertains to the insufficient (meta)data quality on OGD portals. Indeed, in order to effectively utilize OGD, it is crucial to have access to information regarding their quality. This is because inadequate (meta)data quality can lead to inaccurate or incomplete analyses, decision-making, and research on the part of users (Crusoe et al., 2019; Kubler et al., 2018). Therefore, we emphasize the importance of assessing (meta)data quality of individual datasets before using it in service development.

Many previous works (Bhandari et al., 2021; Chokki et al., 2022a; Neumaier et al., 2016; Li et al., 2018; Nogueras-Iso et al., 2021; Vetrò et al., 2016; Wenige et al., 2021; Raca et al., 2022; Reiche et al., 2014) suggest methods to evaluate the quality of metadata, data, or both. This study focuses specifically on those related to OGD to address several of their shortcomings. To our knowledge no study has explicitly presented a complete list of dimensions applicable separately to OGD and its metadata. The aim of this study is to address this gap by (1) identifying a list of quality dimensions for evaluating metadata and data separately, (2) integrating these dimensions into QualityOGD, a tool that incorporates user preferences, and (3) evaluating the tool's effectiveness. Our research question is **(RQ2) "What supports automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user needs?"**. To answer this question, a literature review was conducted to identify an exhaustive set of quality dimensions for evaluating both metadata and data, which were then incorporated into a novel goal-oriented conversational agent (referred to as "prototype" in this study), meaning that the agent has a programmed structure for the conversation. To evaluate the ease of use and the usefulness of the prototype, we conducted user testing with 14 participants.

6.1.1 Publications

The content of this chapter is based on the following peer-reviewed scientific publications:

Abiola Paterne Chokki, Charalampos Alexopoulos, Stuti Saxena, Benoît Frénay, Benoît Vanderose, and Mohsan Ali. Metadata quality matters in Open Government Data (OGD) evaluation! An empirical investigation of OGD portals of the GCC constituents. **Transforming Government: People, Process and Policy**, pages 1–21, 2022a. doi: 10.1108/TG-09-2022-0118

- This paper undertakes an evaluation of the metadata quality within the GCC countries, aiming to identify the best-performing metric for each country while also investigating potential disparities in metadata quality across the region. To accomplish this, we adopt a quantitative research approach comprising a series of steps: defining the metrics for assessing metadata quality, gathering datasets from the portals of GCC countries, calculating the metadata quality of these portals, and conducting a comparative analysis of metadata quality among the GCC countries.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, and Benoît Vanderose. Evaluating a Conversational Agent for Open Government Data Quality Assessment. In **29th Americas Conference on Information Systems**, pages 1–10, 2023

- This paper presents a comprehensive compilation of dimensions and metrics (sub-dimensions) that are essential for evaluating the quality of metadata and data content independently for each open government dataset. Furthermore, it offers a collection of features that are necessary for the development of a quality assessment tool for (meta)data. These findings are the outcome of an extensive investigation that encompasses a literature review and user testing.

6.1.2 Outline

The remainder of the chapter is organized as follows. We discuss the existing methods for OGD quality assessment and their limitations (Section 6.2). Then, we explain our research methodology (Section 6.3), and present in the results section the list of quality dimensions, our conventional agent and its evaluation (Section 6.4). Contributions, limitations, and future avenues are discussed in Section 6.5.

6.2 Background

Within this section, we begin by explaining the key concepts employed in this study. Subsequently, we provide an overview of existing quality assessment studies, concluding with pertinent information concerning conversational agents designed for OGD.

6.2.1 Relevance of Metadata and Data Quality for Open Government Data

OGD are mainly published on online portals and consist of two parts: the data itself and its metadata. Data refers to the resource or distribution available for access or download in various formats (e.g., CSV, PDF, Microsoft Excel spreadsheet, etc.) (Neumaier et al., 2016). On the other hand, metadata refers to data that describes other data (Duval, 2001), making it easier to retrieve, use, or manage the data (Milic et al., 2021). In the OGD context, it typically includes information such as the title, description, related topic, keywords, source, license, publisher, contact details, creation, and modification date, among others (Milic et al., 2021; Neumaier et al., 2016). One goal of OGD is to drive the creation of social and economic value through the development of OGD-based innovative products and services. This reuse of OGD is a process that can be structured as four steps (Crusoe et al., 2019). One of these steps is the search and evaluation of data. In this step, good (meta)data quality is necessary since it can facilitate easier discovery of the desired information by users (Attard et al., 2015). Later in the reuse process, data is acquired, processed, and transformed into a product or service. In this step, poor data quality can obstruct the efficient reuse of the data, escalate the costs associated with accessing and interpreting the data, and potentially result in incorrect conclusions being drawn (Zhang and Xiao, 2020).

As such, data of good quality can be defined as data that is "fit for use by data consumers [i.e., users]" (Wang and Strong 1996). In other words, it is a measurement of the ability of data to meet the specific needs of users. As for the metadata quality, it can be defined as the fitness to describe the data, supporting the tasks of finding, identifying, selecting, and eventually obtaining the data (Reiche et al., 2014). Previous studies (Kubler et al., 2018; Vetrò et al., 2016) have evaluated (meta)data quality by combining the values of various (meta)data quality dimensions, which are sets of attributes that represent specific aspects of data quality (Batini et al., 2009; Wang and Strong, 1996). Examples of data quality dimensions include completeness, accuracy, consistency, and timeliness (Bhandari et al., 2021). Each dimension is described by one or more metrics (also called sub-dimensions) such as the percentage of complete cells (i.e., cells that are not null nor empty) for the completeness dimension (Vetrò et al., 2016). These metrics aim to offer users insights into the potential root causes of (meta)data quality issues, whether they are high-level or more nuanced in certain cases.

6.2.2 Existing Quality Assessment Studies

Many studies have investigated the assessment of the quality of (meta)data, but this study focuses specifically on those related to OGD, excluding linked open data (Zaveri et al., 2012). We focused on OGD in tabular format for two reasons. First, many users, are more familiar with tabular data than the RDF format which is used in LOD. Second, most of the datasets available on OGD portals are published in plain text (CSV). To make a fair comparison between these studies, the criteria established by Zhang and Xiao, 2020 have been utilized and supplemented with additional criteria (see those underlined). The considered key criteria are: indicator (which quality dimensions were taken into consideration), data type (whether the focus was on metadata, data, or both), clarity of distinction (whether there was a clear separation between dimensions for data and metadata), application (which portals or datasets the study has been applied to), operation (whether the assessment was automated or manual), and user needs (whether user can setup the dimensions, metrics, or data attributes to be taken into account in the quality assessment). Table 6.1 presents the previous studies described by the mentioned criteria.

Based on Table 6.1, we can note that no study was able to fully cover all the criteria mentioned. For example, Raca et al., 2022 considered the coverage of metadata and data metrics as well as automation, but their work is applicable to the portal rather than the dataset, there is no clear distinction between the metrics to be used on metadata or data, and there is no option for users to define their preferences. This study aims to provide a data quality assessment that can cover all the listed criteria. It therefore differs from previous studies in several aspects. It defines a list of dimensions that distinguishes between data and metadata and focuses on the dataset level rather than the portal level. It provides a tool that supports fully automated data and metadata quality assessment, and it allows users to set their preferences at the dimension, metric, and attribute levels. Finally, it is the first work that documents the necessary features to be integrated into a tool for assessing OGD quality.

Table 6.1: Existing OGD quality assessment studies. N/A stands for Non Applicable, A for Automated and SA for Semi-automated

Studies	Indicator	Data type	Clarity of distinction	Application	Opera-tion	user needs
(Chu and Tseng, 2016)	Accessibility, Primary, Timely, Accuracy, Integrity, and Abundance	Metadata & Data	No	OGD platform of the central government of Taiwan (including open data of nine agencies)	Not Mentioned	No
(Reiche and Hofig, 2013)	Completeness, Weighted Completeness, Accuracy, Richness of Information and Accessibility	Metadata	N/A	3 public government data repositories, namely GovData.de, data.gov.uk and publicdata.eu	A	No
(Chokki et al., 2022a)	Existence of following attributes: title, description, language, theme, keywords, license, publisher, references, and release date	Metadata	N/A	6 Gulf Cooperation Council national portals: Bahrain, Kuwait, Oman, Qatar, KSA, and UAE	A	No
(Li et al., 2018)	Completeness, Accuracy, Consistency, Timeliness, Uniqueness, Understandability, Openness	Metadata & Data	No	Beijing, Guangzhou, and Harbin data platforms	SA	No
(Raca et al., 2022)	Openness, Availability, Accessibility, Discoverability, Timeless, Completeness, Uniqueness, Consistency, Validity	Metadata & Data	No	6 Western Balkan National Open Data Portals: Albania, Bosnia and Herzegovina, Kosovo, North Macedonia, Montenegro, and Serbia	A	No
(Bhandari et al. 2021)	Completeness	Data	N/A	OGD of South Korea: National and 3 municipalities (Incheon, Seoul, and Gyeonggi)	A	No

Studies	Indicator	Data type	Clarity of distinction	Application	Opera-tion	user needs
(Kubler et al., 2018)	Existence, Conformance, Retrievability, Accuracy, Open data	Metadata	N/A	250 open data portals, powered by organizations across 43 different countries	A	Yes
(Umbrich et al. 2015)	Retrievability, Usage, Completeness, Accuracy, Openness, Contactability	Metadata	N/A	82 CKAN portals	A	No
(Neumaier et al., 2016)	Existence, Conformance, Retrievability, Accuracy, Open Data	Metadata	N/A	260 Open Data portals	A	No
(Kubler et al., 2016)	Usage, Completeness, Openness, Addressability, Retrievability	Metadata	N/A	146 CKAN portals across 44 countries	A	Yes
(Reiche et al., 2014)	Completeness, Weighted Completeness, Accuracy, Richness of Information, Readability, Availability, Misspelling	Metadata	N/A	10 CKAN OGD portals	A	No
(Vetrò et al., 2016)	Traceability, Currentness, Expiration, Completeness, Compliance, Understandability, Accuracy	Data	N/A	Italian portals: National level and municipality level (Torino, Roma, Milano, Firenze, Bologna)	SA	No
(Data.europa.eu, 2020)	Completeness, Findability, Accessibility, Interoperability (Conformity, Machine Readability, Openness), Reusability (Timeliness, Consistency, Accuracy, Relevance, Understandability, Credibility), Contextuality	Metadata	N/A	Applicable to individual and overall datasets on Data Europa portal	A	No

6.2.3 Conversational Agents for Open Government Data

According to Nuseibeh, 2018, conversational agents are software programs that interpret and respond to natural language statements made by users. Nuseibeh also categorizes conversational agents into two types: chatbots and goal-oriented conversational agents. Chatbots are designed to simulate conversations with human users, with ChatGPT being a well-known example. In the context of OGD, several studies (Cantador et al., 2021; Keyner et al., 2019; Porreca et al., 2018; Wang et al., 2023) have utilized chatbots for OGD search and exploration.

Goal-oriented conversational agents, on the other hand, have a pre-programmed structure for the conversation, controlling the conversation flow by asking questions and ignoring user inputs that do not answer the question. Moreau et al., 2019 propose a tool named SemanticBot, a semi-interactive ontology mapping tool that utilizes goal-oriented conversational agents to provide an easy-to-use interface for mapping ontologies on OpenDataSoft datasets.

Despite the popularity of conversational agents in OGD research, there is a lack of studies that utilize them as interfaces for data quality assessment. This study aims to address this gap by proposing a goal-oriented conversational agent that users (i.e, infomediairies and citizens) can use for assessing (meta)data quality. In addition to filling this gap, the implementation of a conversational agent was preferred for two reasons. Firstly, it offers an intuitive and user-friendly way for inputting user preferences, requiring minimal to no training due to its natural language communication and integration into popular instant messaging apps (Cantador et al., 2021). Secondly, it has been shown to be effective in related OGD studies focused on open data search and exploration (Cantador et al., 2021; Keyner et al., 2019; Moreau et al., 2019; Porreca et al., 2018; Wang et al., 2023).

6.3 Research Methodology

To address the research question **(RQ2) "What supports automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user needs?"**, this study follows a research design comprising a literature review, prototyping, and an evaluation through user testing with follow-up interviews.

The literature review, based on (Webster and Watson, 2002), spanned from October to December 2022 and utilized databases such as "Scopus," "Association for Computing Machinery," and "Google Scholar". It helped to identify dimensions and associated metrics related to (meta)data quality, using as the following search queries: ("metadata quality" OR "data quality") AND (evaluation OR assessment OR assess OR evaluate) AND ("data" OR "open data" OR "open government data" OR "public data" OR "public government data" OR "government data" OR "public sector information"). An automated search returned 311 articles, out of which 13 relevant ones were selected through a three-stage process that evaluated type, domain, and title, examined the abstract, and scanned the content. As the focus was on automating quality assessment, only objective (i.e., computable/assessable without human intervention) dimensions and metrics were considered, which were grouped and categorized into metadata or data quality dimensions/metrics based on conceptual similarities.

After identifying the dimensions/metrics, we utilized a prototyping approach (Budde et al., 1992) to propose QualityOGD, a practical tool for users that can effortlessly integrate them into their quality assessment process. Requirements for the prototype were collected from existing tools from the literature review and integrated along with the dimensions/metrics. The backend of the prototype is developed with

Django and PostgreSQL, while the frontend is built using Angular (code available at https://github.com/chokkipaterne/qualityogd).

After implementing the prototype, we recruited 14 users through an open call for participation in a "data analytics" course at the University of Namur and conducted user testing to evaluate the prototype ease of use and usefulness. Additionally, we administered an online survey to gather further feedback. Appendix D presents the questions contained in the survey. To minimize errors associated with survey research, the survey was pretested with two users (Grimm, 2010). The survey comprised items measured on a 5-point Likert scale (from “strongly disagree” to “strongly agree”) based on the System Usability Scale (SUS) questionnaire (Brooke, 1986) with 5 additional questions to evaluate ease of use and usefulness. In addition to these questions, there were free-text questions to collect general opinions and suggestions for future versions and 3 additional questions to collect demographic data (data manipulation skills, age, and education). Participants were invited to test the prototype with their preferred datasets from the OpenDataSoft (ODS) data portal[1] to assess their (meta)data quality. During the test, we used an exploratory approach (Rubin and Chisnell, 2008) and only provided guidance when participants asked for assistance. After completing the survey, users were asked *what features should be kept, improved, removed, or added* to facilitate (meta)data quality assessment. We calculated the median, mean, and standard deviation for the Likert scale questions, and coded verbal thoughts and responses from the free-text questions using short sentences to retain context and conceptual relations.

6.4 Results

6.4.1 A Conversational Agent for (Meta)data Quality Assessment

Using existing tools features, we gather initial features for the OGD quality assessment tool, which include (F1) an interface that is easy to use and intuitive, (F2) an automated assessment process, (F3) the ability to differentiate between metadata and data quality assessments, (F4) direct dataset selection from portals, (F5) user preference selection for dimensions, metrics, and attributes during quality assessment, (F6) data quality visualization, and (F7) an explanation of the reasons behind each score for each dimension.

We developed a conversational agent to incorporate the identified features. A demonstration video performed by us to evaluate the quality of the data "JCDecaux Bike Stations Data"[2] collected from the ODS portal is accessible at https://rb.gy/89yfpr. As explained in the introduction, conversational agents require minimal to no training due to their natural language communication and integration into popular instant messaging apps. Based on this, the implemented prototype meets requirement (F1). Figure 6.1 portrays the prototype conversation flow on its left side, where user inputs (in rectangle boxes) and prototype outputs (in parallelogram boxes) are represented by intents (i.e., user inputs and outcomes of user inputs). On the right side of the figure, there is a screenshot of the prototype interface.

[1] https://data.opendatasoft.com/

[2] https://data.opendatasoft.com/explore/dataset/jcdecaux_bike_data%40public/table/

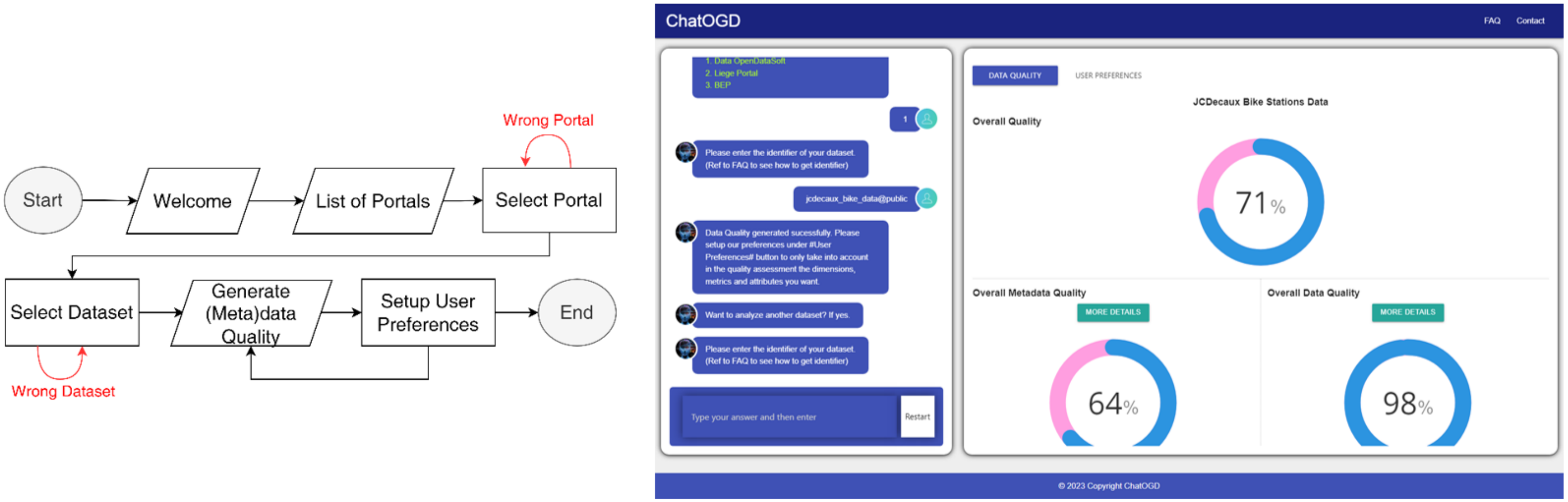


Figure 6.1: Conversation flow of the prototype (parallelogram boxes for prototype outputs and rectangle boxes for user inputs) (Left). Screenshot of prototype interface (Right).

Welcome. This intent is automatically triggered at the start of the conversation, during which the prototype introduces itself and greets the user.
Portals List. This intent is automatically triggered after the welcome intent or when the user clicks the “Restart” button to initiate the process from the beginning. Its purpose is to provide a list of the portals included in the prototype.
Portal Selection. This intent enables users to choose the portal to which the dataset being processed belongs, from the portals list. In the event that the user chooses an incorrect portal identifier, the prototype will alert them with the error message "The information on the dataset appears to be incorrect" and request the user to make an alternative selection.
Dataset Selection. This intent enables the user to input the identifier of the dataset they wish to assess (**F4**). The prototype then sends this identifier to the backend to verify its existence and the distribution format of the dataset, which must be machine-readable. If the identifier exists and the format is supported, the prototype automatically generates the (meta)data quality of the selected dataset using all dimensions, metrics, and attributes. When the identifier is invalid, the prototype communicates this to the user through an error message, stating "Unable to retrieve the data content", and encourages the user to make another attempt.
Generate (Meta)data Quality. This intent enables the user to automatically retrieve the (meta)data quality of the selected dataset (**F2**). The results are presented in three parts with visualizations to facilitate understanding:

- **overall quality**, calculated as the mean of all (meta)data dimensions and displayed using a gauge chart (**F6**).
- **metadata quality (F3),** calculated as the mean of only metadata dimensions and displayed using a gauge chart (**F6**). By clicking on "More details", the gauge chart is replaced with a bar chart (**F6**) where each bar represents one metadata dimension, and its value is computed as the mean of the metrics under that dimension. When the user clicks on a dimension, the prototype displays the bar chart for the metrics under the selected dimension (F7).
- **data quality (F3),** similar to metadata quality but only displays dimensions and metrics related to the data.

Table 6.2 showcases the dimensions and metrics employed in the prototype, providing comprehensive details about these metrics, such as their definitions and corresponding formulas. To ensure that their assessment remains dynamic, we chose to include only dimensions that can be calculated using formulas, based on previous studies (Data.europa.eu, 2020; Reiche et al., 2014; Vetrò et al., 2016).

Table 6.2: List of dimensions and metrics integrated into the conversational agent.

Dimensions	Metrics	Definition	Formula
Metadata			
Completeness	Percentage of complete fields in metadata	Indicates the percentage of complete fields in metadata. A complete field is defined as a value that is not null and is set	Number of complete fields * 100/Total number of fields
Findability	Keywords assigned	Indicates if the keywords field is assigned	100 if yes else 0
	Categories assigned	Indicates if the categories field is assigned	100 if yes else 0
	Title given	Indicates if the dataset title is set	100 if yes else 0
	Description of data given	Indicates if the dataset description is given	100 if yes else 0
	Temporal information given	Indicates if the date/period of the dataset belongs to is given	100 if yes else 0
	Spatial information given	Indicates if the information about the region the dataset belongs to is given	100 if yes else 0
Accessibility	Access URL accessible	Indicates if the access URL of the dataset is accessible	100 if yes else 0
	Download URL given	Indicates if the download URL of the dataset is given	100 if yes else 0
	Download URL accessible without registration	Indicates if the download URL of the dataset is accessible without registration	100 if yes else 0
Conformity	Conformity of URLs	The extent to which the values of access properties (HTTP, URLs) are valid	Number of valid URL fields * 100/Total URL fields
	Conformity of date formats	The extent to which the date fields are using a valid date format	Number of valid date fields * 100/Total number of date fields
	Conformity of email addresses	The extent to which the email fields are valid	Number of valid email fields *100/Total number of email fields
	DCAT-AP compliance of metadata	Indicates if the metadata of the dataset is provided in DCAT format and is compliant with DCAT-AP	100 if yes else 0

Table 6.2: List of dimensions and metrics integrated into the conversational agent.

Dimensions	**Metrics**	**Definition**	**Formula**
Machine readability/ processability	Processability of file format and media type	Indicates if the file format and media type can be handled by automated processes (CSV, JSON, Excel accepted in this study)	100 if yes else 0
Openness	Openness of file format and media type	Indicates if the file format and media type is in a non-proprietary format (CSV, JSON accepted in this study)	100 if yes else 0
	License information given	Indicates if the license information is given	100 if yes else 0
	Openness of license	Indicates if the license corresponds to one of the open Licenses listed in opendefinition.org	100 if yes else 0
Timeliness	Update information given	Indicates if the frequency/periodicity of the dataset update is given	100 if yes else 0
	Creation date given	Indicates if the creation date is given	100 if yes else 0
	Modification date given	Indicates if the last modification date is given	100 if yes else 0
Accuracy	File format accuracy	Indicates if information given about file format can be compared with the actual file format of the resource	100 if yes else 0
	Content size accuracy	Indicates if information given about content size can be compared with the actual content size of the resource	100 if yes else 0
Understand-ability	Percentage of columns with metadata	Indicates the percentage of columns in a dataset that has associated descriptive metadata.	Number of columns with metadata * 100/ Total number of columns
Credibility	Contact point given	Indicates if the contact point of the dataset is given	100 if yes else 0
	Dataset publisher given	Indicates if the dataset publisher information (e.g., name or email) is given	100 if yes else 0
Uniqueness	Title is unique	Indicates if the dataset title is unique	100 if yes else 0
	Description is unique	Indicates if the dataset description is unique	100 if yes else 0
	Identifier is unique	Indicates if the dataset identifier is unique	100 if yes else 0

Table 6.2: List of dimensions and metrics integrated into the conversational agent.

Dimensions	Metrics	Definition	Formula
Data			
Completeness	Percentage of complete cells	A cell which is not empty nor null is considered as complete	Number of complete cells * 100/ Total number of cells
	Percentage of complete rows	A row which all the cells are not empty nor null is considered as complete	Number of complete rows * 100/ Total number of rows
Accuracy	Percentage of accurate cells	Accurate cells are identified by their corresponding column type. Therefore, a cell is deemed accurate if its type matches the column type. For instance, if a column has a type of "email", a cell with the value "abc" in that column would be regarded as inaccurate.	Number of accurate cells * 100/ Total number of cells
Uniqueness	Percentage of unique rows	Indicates the percentage of unique rows in the dataset	Number of unique rows * 100/ Total number of rows
	Percentage of unique columns	Indicates the percentage of unique columns in the dataset	Number of unique columns * 100/ Total number of columns

user needs. This intent allows users to configure their preferences (**F5**). In contrast to other studies (Kubler et al., 2018) that limit user choices to dimensions, this study focuses on three aspects of user needs: dimensions, metrics, and data attributes. Users have the ability to assign weights to each dimension and metric involved in the quality assessment process. The default weight assigned is 1, implying an equal impact of all dimensions and metrics on the assessment. However, users have the flexibility to emphasize the importance of a specific dimension or metric by assigning it a weight greater than 1. Conversely, a weight of 0 can be allocated to exclude a particular dimension or metric from the assessment. This flexible weighting system allows users to highlight those dimensions or metrics that hold more significance over others, based on their requirements. In addition, users can choose which data attributes to include in the data quality assessment by unchecking unwanted attributes using checkboxes.

6.4.2 Evaluation Results

We collected opinions from 14 participants (10 undergraduates between the ages of 18 and 29 and 4 doctoral students between the ages of 30 and 49), regarding the ease of use and usefulness of the conversational agent to assess (meta)data quality. This was done through surveys that the participants completed after exploring the prototype.

Table 6.3 presents the median (MD), mean (M), and standard deviation (SD) of the SUS score (which summarizes questions about usability), and five additional questions used to assess the usefulness of the prototype. The results of Table 6.3 lead to several conclusions. First, based on the interpretability of the SUS score presented in (Lewis and Sauro 2018), the prototype offers excellent usability as both the MD and M scores for the SUS are above 85. Second, participants agree that the prototype is useful for assessing overall quality, metadata quality, and data quality, as the MD and M of these questions are each equal to or greater than 4 with SD < 1. Third, most users agree that the prototype helps them to identify where data quality issues exist (Q14) and set their preferences (Q15), as MD and M are greater than 3.5 with SD ≈ 0.75.

Table 6.3: Median (MD), mean (M) and standard deviation (SD) of survey scores.

	SUS Score (Q1-Q10)	Usefulness for quality assessment				
		Data + Metadata (Q11)	Metadata (Q12)	Data (Q13)	better highlight errors related to a dataset (Q14)	take into account user needs (Q15)
MD	85	4	5	5	4	4
M (SD)	86.07 (6.10)	4.42 (0.53)	4.57 (0.53)	4.57 (0.53)	3.57 (0.78)	3.71 (0.75)

During testing, we found that all participants were able to assess their selected dataset's (meta)data quality with little or no assistance. Participants indicated that the prototype was easy sto use, intuitive, and ergonomic, which explains the excellent SUS score. The user needs section, particularly at the attribute level, was highly valued by most participants as it allowed them to evaluate only the columns that were relevant to them. All participants appreciated the use of visualizations to express the quality of the (meta)data instead of simple numbers. They also appreciated the hierarchical presentation of quality results (starting with an overview, followed by dimension level, and finally metric level), which allowed them to quickly grasp the overall quality of the (meta)data and then delve deeper to understand the underlying values.

All participants agreed that the current features should be kept, but many of them suggested simplifying the presentation of weight setup at the dimension and metric level. As for new features, they proposed adding a dimension to measure the correctness of the (meta)data description in comparison to its title and content, as well as the accuracy of the column descriptions and cell values to ensure that the dataset reflects real-world data accurately. One participant also recommended adding an option to export the (meta)data quality results in PDF or Excel format. Another participant suggested including more details on certain metrics, such as presenting the specific columns that make the rows incomplete in the data for the "Percentage of complete rows" metric.

6.5 Wrap up

The aim of this study is to identify quality dimensions for assessing metadata and data separately in the context of OGD, integrate them into a tool that accommodates user needs, and assess its effectiveness. To accomplish this, we conducted a literature review to identify quality dimensions, metrics, and initial requirements for the prototype. We then incorporated this information into a conversational agent and evaluated its ease of use and usefulness in assisting users to evaluate the quality of a dataset through a user test that involved 14 participants. The user test results show that users found the prototype easy to use and useful to assess the quality of data and metadata, to highlight the errors related to a dataset, and to take into account their preferences.

This study makes theoretical contributions in several aspects. Firstly, it builds upon prior studies (Data.europa.eu, 2020; Reiche et al., 2014; Vetrò et al., 2016) to propose a comprehensive list of dimensions and metrics (sub-dimensions) necessary to evaluate the quality of metadata and data content separately (Table 6.2). Secondly, unlike previous studies, this study explicitly specifies which dimensions are applicable to metadata and data content and provides corresponding metrics to be considered for each dimension. This categorization enables users to be aware of the relevant dimensions and metrics when evaluating metadata or data content. Additionally, it provides an overview of the dimensions and allows for a more in-depth understanding of the errors in the data. Thirdly, through existing tools, we have identified a set of features required to develop a (meta)data quality assessment tool. This list can serve as a reference for infomediaries or researchers to compare or analyze (meta)data quality assessment tools.

This study also carries practical implications in several important areas. Firstly, it is the first known study to propose a quality assessment tool for open government data using a conversational agent based approach. Secondly, unlike the other tools discussed in the Background Section (refer to Table 1), the prototype (QualityOGD) addresses various limitations by providing complete automation, incorporating dimensions and metrics for both metadata and data, clearly distinguishing between these aspects in the interface, and enabling users to personalize their preferences during the quality assessment process. Thirdly, while other tools only allow users to set preferences at the dimension level, the prototype extends this capability to the metric and attribute levels as well. Finally, we provide open access to the prototype source code, which can serve as a valuable resource for developers aiming to create their own tool for assessing (meta)data quality or enhance the existing prototype.

One significant limitation of this study is the potential lack of representativeness of the evaluation sample. While the number of participants may be considered small, previous studies (Faulkner, 2003; Nielsen, 2000) suggest that a minimum of five participants for usability tests is a good starting point, in addition Lallemand and Gronier, 2015 indicate that around 15 users is enough to study general tendency. To improve representativeness, we suggest exploring alternative communication channels or conducting on-site evaluations at universities or with open data users. Another limitation of the study is the small number of dimensions/metrics selected on the data level. However, this decision was made deliberately, as only objective dimensions/metrics were chosen. This was to ensure that their measurements could not differ from one user to another, unlike subjective measures such as column understandability, which can vary depending on the user's perception. In future work, we plan to (re)define (existent) additional dimensions/metrics that are relevant to the (meta)data and explore ways to automate them through machine learning methods, and also assess whether the quality assessment computed by the prototype is consistent with expert judgment. Second, we will assess the impact of a conversational agent's presence on an OGD portal by utilizing the Technology Acceptance Model (Davis, 1989) to measure citizens' intent to use it. Third, we plan to conduct an experimental study to identify which types of users (developers, data scientists, businesses, or citizens) are most likely to find the conversational agent helpful.

CHAPTER 7

TACKLING BARRIER 3: FACILITATING DATA STORYTELLING WITH OGD

7.1 General Introduction

This chapter presents the contributions addressing the third barrier to OGD reuse, that is, the difficulty for users to turn data into stories. Once users have become aware of the existence and usefulness of OGD (Section 5) and have gained knowledge about the quality of these datasets (Section 6), the challenge remains of how to effectively transform these datasets into compelling narratives. The ability to convert data into meaningful stories is essential for informed decision-making, but it can prove challenging for many users, particularly those lacking specialized technical skills. Furthermore, this process can be time-consuming if the necessary skills and tools are not readily available.

Before embarking on the task of creating data-driven stories, it is crucial to determine which data will be of interest to a wide range of users and how to present that data in a manner that facilitates comprehension. To address these considerations, we have formulated two research questions: **(RQ3a) "What datasets are of primary interest to users in the Wallonia context?"** and **(RQ3b) "What visualization optimization approaches are most effective for specific type of datasets?"**. To answer these questions, we conducted an analysis of the usage statistics of the Namur OGD portal and administered an online survey to identify users' specific requirements regarding the types of data they would like to see available on such portals. These endeavors provided valuable insights and allowed us to generate recommendations for OGD publishers and infomediaries regarding the types of data that should be prioritized for publication. Additionally, through interviews, we were able to determine the most suitable visualization methods for different categories of data, taking into account user ratings. This comprehensive approach enabled us to gain a deeper understanding of user preferences and needs when it comes to data presentation and visualization on portals.

Recognizing that individual visualizations may not provide a comprehensive summary of dataset content, we propose the utilization of dashboards consisting of multiple visualizations. The aim is to present data in a manner that enhances user understanding. To address this, we have formulated the following research questions: **(RQ3c) "What are the design principles for effective dashboards in the**

context of OGD?" and **(RQ3d) "In what ways do well-structured dashboards, incorporating best dashboard design principles, enhance citizen engagement with OGD compared to individual visualizations?"**. To answer these questions, we conducted a literature review to gather dashboard principles, which were subsequently validated through user testing using a tool that implemented these principles as a proxy.

Having obtained insights from the aforementioned research questions, our next step was to identify the requirements for data storytelling tools that could facilitate user-driven narrative creation. Consequently, we formulated the research question as follows: **(RQ3e) "What essential features should be incorporated in a comprehensive, end-to-end data storytelling tool tailored for open data?"**. In order to answer this research question, we conducted a literature review to identify the essential features to be integrated into a data storytelling tool, encompassing all stages of the process within the open data context. Once these features were identified, we validated them through user testing using a tool that implemented them as a proxy.

7.1.1 Publications

The content of this chapter is based on the following peer-reviewed scientific publications:

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open Government Data for Non-expert Citizens: Understanding Content and Visualizations' Expectations. **Research Challenges in Information Science**, 415 LNBIP: 602–608, 2021. ISSN 18651356. doi: 10.1007/978-3-030-75018-3_42

- This paper presents the datasets that are of interest to users and the visualizations that they prefer on OGD portals. These insights are valuable for infomediaries who are searching for suitable datasets for their applications, as well as for OGD publishers who need to determine which datasets to make available on their portals. Furthermore, these findings offer guidance on the types of visualizations that enhance user comprehension. These findings were obtained through a multi-method approach, which included an analysis of the usage statistics of the OGD portal of Namur, a complementary online survey to understand the dataset needs of end-users, and interviews to gather their opinions on correctly designed and well-presented visualizations of datasets.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Engaging Citizens with Open Government Data: The Value of Dashboards compared to Individual Visualizations. **Digital Government: Research and Practice**, 2022d. doi: https://doi.org/10.1145/3558099

- This paper presents a compilation of design principles for dashboards in the OGD context, aiming to enhance their usability and user understanding. These principles are identified through a systematic literature review. Furthermore, the paper validates the hypothesis that the use of well-designed dashboards can promote citizen engagement with OGD. This validation is achieved through the implementation of these design principles in a usable tool and the assessment of their impact on citizen engagement using the tool as a proxy.

Abiola Paterne Chokki, Benoit Frenay, and Benoît Vanderose. Open Data Explorer: An End-to-end Tool for Data Storytelling using Open Data. In **Americas Conference on Information Systems 2022 Proceedings. 6**, 2022b

- This paper introduces a collection of essential features required for designing a generic and end-to-end data storytelling tool that enables users with varying levels of data manipulation skills to effortlessly transform OGD into engaging narratives. The identification of these features is based on a comprehensive literature review and interviews conducted with potential users. To validate these features, a functional tool named ODE (Open Data Explorer) is utilized as a proxy, incorporating the identified features.

Abiola Paterne Chokki and Benoît Vanderose. From Conventional Open Government Data Portals to Storytelling Portals : The StoryOGD Prototype. In **24th Annual International Conference on Digital Government Research**, pages 1–2, 2023

- This paper introduces StoryOGD, an enhanced iteration of the ODE tool, which is explicitly tailored to aid publishers in effectively presenting their portals in a user-friendly fashion. Additionally, it outlines the requirements necessary for implementing such a tool, which were identified through an exploratory study of websites that aggregated open datasets and presented them in a user-friendly manner.

7.1.2 Outline

The organization of this chapter is as follows: Section 7.2 provides insights into users' perspectives on content and visualization expectations. Section 7.3 delves into the principles underlying effective dashboards. Furthermore, Section 7.4 addresses the requirements necessary for designing a tool that facilitates data storytelling. Finally, Section 7.5 concludes the chapter by summarizing its key contributions.

7.2 Understanding Content and Visualizations' Expectations from the Users' Perspectives

In this section, we present the datasets of interest to users of Wallonia and preferred visualizations as expressed by users on OGD portals. These insights are valuable for infomediaries seeking suitable datasets for applications and for OGD publishers in determining the datasets to make available on portals. Additionally, they provide guidance on the types of visualizations that enhance user comprehension.

The methodology employed to identify users' content and visualizations expectations is outlined in Section 7.2.1. The findings related to content and visualizations expectations are presented in Section 7.2.2. Finally, Section 7.2.3 summarizes the contributions of this section, acknowledges the study's limitations, and proposes potential avenues for future research.

7.2.1 Research Methodology

In order to address **(RQ3a) "What datasets are of primary interest to users in the Wallonia context?"**, we combined three resources. First, we used the OGD portal of the city of Namur (Belgium)[1] as use case to study the actual consultation statistics of the datasets on the portal. We chose this portal as it is the most advanced portal in Wallonia (Belgium) and access with key stakeholders of this portal was possible. This information was collected through a file sent by the OGD manager of the city to the researchers, as we could not directly access the information on the portal. Second, we initiated the survey using the High-Value Datasets[2] (HVDs) from the Dutch government's data portal, given the historical and cultural proximity between the Netherlands and Belgium. The list was also used to verify if it matches with the real expectations of non-expert users. Third, to complement the findings from

[1] https://data.namur.be/pages/accueil/
[2] https://data.overheid.nl/community/maatschappij/high-value/gemeenten

these statistics, we issued a survey to the users of Namur asking them what datasets they expect to find on portals. Appendix E presents the questions contained in the survey. The survey was pretested by two users to ensure all kinds of errors that are associated with survey research are reduced (Zuiderwijk et al., 2012). The survey was later shared on Facebook groups and was filled in by 43 users. This low participation rate can be explained by three reasons: (i) we focused here on participants interested on OGD (i.e., those who may or may not be familiar with OGD but are keen to learn more) which represent a very specific subset of the population, (ii) we only used online channels due to the COVID situation and (iii) the survey was conducted as a complement to the usage statistics. Figure 7.1 summarizes the methodology used to address RQ3a.

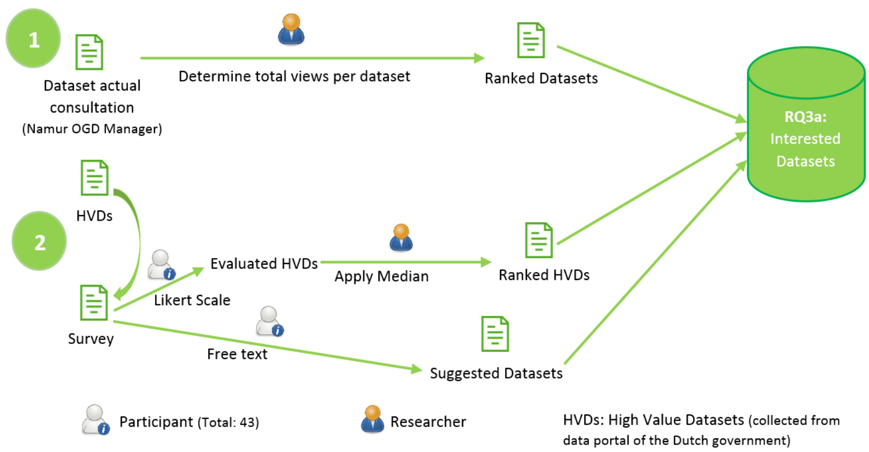


Figure 7.1: Methodology used to identify the datasets of interest to users of Wallonia.

To answer **(RQ3b) "What visualization optimization approaches are most effective for specific type of datasets?"**, we conducted interviews with 10 users of Namur, interested to know more about OGD. These 10 users were recruited on voluntary basis based on their answers to the previous survey. The reasons for low participation are the same as above, with the exception of the third. Figure 7.1 summarizes the methodology used to address RQ3b.

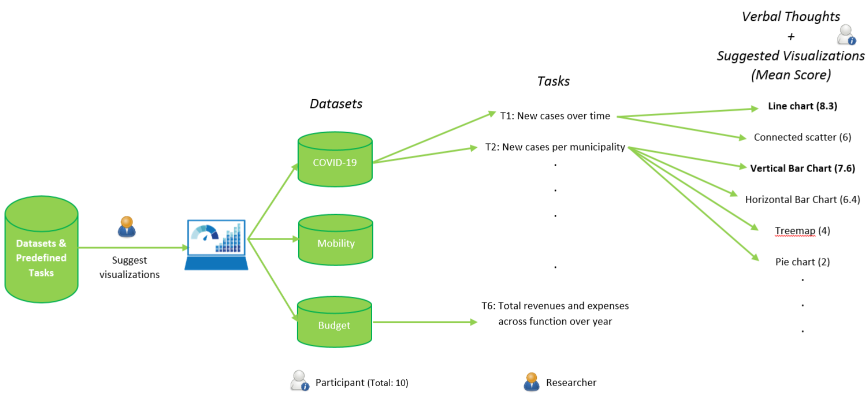


Figure 7.2: Methodology used to collect visualizations' expectations from users.

Datasets & predefined tasks. We selected three datasets from the open data portal of Namur for the interviews. These datasets were chosen because they are easy to understand by participants and also are among the most visited datasets according to

Table 7.1: List of datasets and predefined tasks for interviews.

Datasets	Predefined Tasks
COVID-19 Pandemic - Province of Namur - New contaminations by commune Link: https://rb.gy/4r0ht7	(T1) Total new cases over date (T2) Total new cases per municipality
Namur - Mobility - Parking Link : https://rb.gy/b840w5	(T3) Total places per parking type (T4) Total places per parking type and per municipality
Namur - Ordinary budget by function Link : https://rb.gy/1q79nd	(T5) Total revenues and total expenses across function (T6) Total revenues and total expenses across function over year

the usage statistics file collected from the OGD manager of Namur. For each dataset, we have defined 2 tasks that the participants need to do in order to record their feedback. The predefined tasks were also well selected in order to cover different use cases of data visualization. Table 7.1 summarizes the information about the datasets (name, link for more details and predefined tasks).
Data collection. For each predefined task, we have generated at least 2 different visualizations. Participants were then asked to give a score between 1 (very inappropriate) to 10 (very appropriate) to each generated visualization. In addition, participants were asked to verbalize their thoughts during the study about why they gave a certain score for a specific visualization and also how they would like to represent the visualization to facilitate understanding. These thoughts were recorded so that nothing was missed from their feedback. Each subject spent approximately 30min to note in total 22 visualizations for all the predefined tasks.
Data analysis. After collecting user feedback, the final score of each proposed visualization type for each predefined task was calculated using the average of user ratings. The different scores were then used to find the best visualization type (visualization with the highest score) for each predefined task.

7.2.2 Results

In this section, we present the results obtained after applying the different methodologies presented earlier.

7.2.2.1 Content expectations: usage statistics and survey results

The file collected from the OGD manager of Namur portal concerns the consultation of the portal's data from January to December 2020. This file contains 902 573 rows and 34 columns such as timestamp, user_ip_addr, dataset_id, exec_time and so on. Based on this file, we determine how many times each dataset was visited between January and December 2020. Table 7.2 shows the top 10 datasets consulted on the OGD portal of Namur.

Referring to Table 7.2, the order of the datasets from most to least visited, is as follows: COVID datasets, datasets on cemeteries, data on communities, localities, addresses, and buildings, mobility data and population data. Another observation is that some expected datasets, such as budget data to achieve transparency, were less visited but were among the 100 most visited datasets. Also, many datasets in the list of HVDs are not found in the list of datasets visited on the Namur portal.

Regarding the survey, a total of 43 users completed it. 63% of users had heard about open data, 53% had used an open data portal and 70% had general computer

Table 7.2: Top 10 datasets visited between January and December 2020 on the OGD portal of Namur.

No	Dataset	Dataset Category in Survey	Number of records	% records
D1	Number of confirmed COVID-19 cases by municipality	COVID	298498	33.1
D2	Number of new confirmed COVID-19 cases per municipality per day	COVID	57130	18.4
D3	Number of new hospitalizations of COVID-19 per province per day	COVID	51232	6.33
D4	List of Deceased Related to Cemetery Locations	Non Present	24260	5.68
D5	Administrative boundaries - Municipalities of the Province of Namur	Non Present	20108	2.69
D6	Polygons of 26 localities of the commune of Namur	Non Present	17776	2.23
D7	Boundaries of districts of Namur	Non Present	16345	1.97
D8	Location of Public Cemeteries	Non Present	14588	1.81
D9	Photos and geolocalized old postcards	Non Present	10497	1.62
D10	List of the deceased linked to the cemetery sites in the commune of Namur	Non Present	8756	1.16

knowledge. First, we asked participants to quantify the importance of the predefined datasets (coming from the list of HVDs) using a scale from “Not important at all” to “Very Important”. The importance was calculated as the median response of the 43 respondents. The survey results show that most of the datasets are important (median=3) for users except the datasets about street lighting, places to walk dogs, information on trees and spreading routes, which have a median less than 3 (not important). Second, we asked the following question to participants: “What data (other than those listed) would you like to see on an Open Data site?”. 13 participants answered it. The list of suggested data included: nurseries libraries, road work schedule, local business statistics, position of the refugee centers and their age pyramid, collection and use of tax and information on essential shops.

Based on these findings, we suggest publishers to highlight on the portal (respectively infomediaries to offer services based on) the high-value datasets, COVID-Related Data (or, more generally, data relevant to analyze a current crisis and/or societal debate in an objective manner), administrative boundaries and population data, a list of buildings, mobility data and old photos from the city. On the other hand, publishers should also provide, in addition to the current data, datasets about nurseries libraries, road work schedule, local business statistics, position of the refugee centers and their age pyramid, collection and use of tax and information on essential shops. Figure 7.3 summarizes the most visited datasets and those suggested by users. They should also have a feedback feature which can help collect user expectations in terms of the datasets to be published on the portal. Moreover,

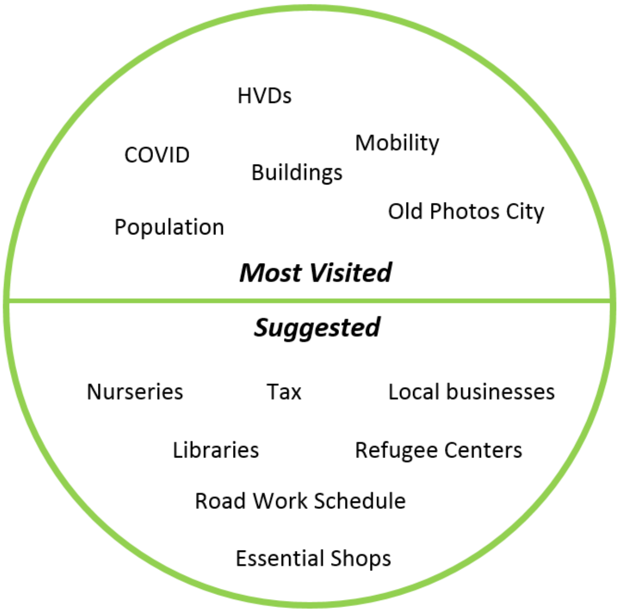


Figure 7.3: Datasets most visited and those suggested by users of Wallonia

.

they should implement a system for tracking the usage statistics of each published dataset, employing a traffic analysis tool to assess their relevance and importance.

7.2.2.2 Visualization expectations: results from interviews

In total, 10 users participated to the interviews. All participants had average to low computer skills and had not previously analyzed the studied datasets. Table 7.3 presents the adequate visualization type for each predefined task based on the user feedback. Note that the best visualization type is determined by taking the visualization type that has the highest final score calculated using the average of user ratings.

Referring to Table 7.3, we can note that the best visualization type for visualizing geographic data is the bubble map, for comparing categorical data is the bar graph, and for seeing the evolution over time is the line graph. In addition, we find that the design of the visualization types is very important for users to help them understand them easily. Thus, based on user feedback on suggested visualizations, we propose that infomediaries and publishers take the following actions to incorporate these user expectations. First, they should have a visualization review feature that allows users to provide suggestions on how to improve visualizations. Second, they should allow users to access the low-level visual encodings such as graph orientation, axis labels, order of data in graph and color, in order to change them if necessary. Third, they should provide filter functionality for each visualisation to allow only the desired data to be displayed rather than all data.

7.2.3 Conclusion

The aim of this study was to understand the content and visualization expectations of users towards OGD. To achieve that goal, we used a multi-method approach including an analysis of the usage statistics of the OGD portal of Namur, a complementary online survey of 43 participants to find out the needs of the end-users

Table 7.3: Best visualization type for each predefined task.

Datasets	Tasks Predefined	Best Visualization Type
COVID-19 Pandemic - Province of Namur - New contaminations by commune	(T1) Total new cases over date	Line chart
	(T2) Total new cases per municipality	Bubble map
Namur - Mobility - Parking	(T3) Total places per parking type	Bar chart with horizontal orientation & Doughnut & Pie chart
	(T4) Total places per parking type and per municipality	Grouped bar chart with horizontal orientation
Namur - Ordinary budget by function	(T5) Total revenues and total expenses across function	Grouped bar chart with vertical orientation
	(T6) Total revenues and total expenses across function over year	Multiple line charts

in terms of datasets and interviews with 10 participants to get their opinion on the correct and well-designed visualizations of datasets. Using this multi-method approach, we identify users' expectations for content in the context of Wallonia (see Figure 7.3) and visualizations (use bubble map for geographic data, bar graph for categorical data, line graph for temporal data and more in (Holtz and Conor, 2018)), and then provide useful recommendations to infomediaries and publishers (see Table 7.4). The results concerning the optimal visualization type for specific data are consistent with the conclusions of the previous study (Holtz and Conor, 2018).

Table 7.4: Data and visualization recommendations for infomediaries and publishers.

Data Recommendations
(a) Publish useful datasets, including HVD datasets and additional suggested datasets (see figure 7.3)
(b) Highlight on the portal (publishers) and offer services (infomediairies) based on the most visited datasets (see figure 7.3)
(c) Incorporate a feedback feature on the portal to capture user expectations for desired datasets
(d) Monitor usage statistics of each published dataset to assess their relevance and importance
Visualization Recommendations
(e) Use adequate visualization based on data types. The figures 7.14, 7.15, 7.16, and 7.17 illustrate the optimal visualization types for particular data, further supported by findings in the prior study (Holtz and Conor, 2018).
(f) Implement a visualization review feature that allows users to provide suggestions on how to improve visualizations
(g) Provide access to low-level visual encodings such as graph orientation, axis labels, order of data in graph and color, in order to change them if necessary
(h) Offer filter functionality for each visualization to allow only the desired data to be displayed rather than all data

This study differs from existing literature in two aspects. First, to our knowledge, this study is the first attempt to use the usage statistics of portal combined with a survey to understand content expectations. Second, in previous research such as (Ornig et al., 2017) only three visualization types (treemap, stacked area chart and map) and uses cases with a single attribute for measure and dimension were suggested for participants, compared to our study, where we provided more than ten visualization types (grouped or stacked bar chart, line chart, connected scatter, map, bubble map, sun-burst, treemap, multiple charts, pie chart and doughnut) and more uses cases with multiple attributes for dimensions and measures. In (Khan and Shah Khan, 2011), the authors have presented the use cases of many visualization types and proposed some methods of interactivity in general contrarily to here where we based on tasks to provide the best visualization type and also suggested the interactivity for each of these visualization types.

However, this study has some limitations that need to be addressed in future work. First, the number of participants was low, but this did not introduce fundamental bias in the study, as the survey was a complementary method to assess user expectations. Furthermore, in-depth usability tests to evaluate a data visualization tool were successful in previous studies with less than ten participants (Faulkner, 2003). To increase the number of participants, we suggest using multiple channels of communication or collecting data on-site in administrations. As in this study, with the COVID, this was not possible. Second, in this study, we only used the usage statistics on Namur portal but this did not also impact the study because most of the portals in the Wallonia contain the same categories of datasets. However, using more portals can be helpful to confirm these findings about the user expectations in terms of datasets in the context of Wallonia. In fact, an international comparison of usage statistics to compare the needs of population from different countries would constitute an interesting lead for further research. Third, the study covered few visualization types, focusing on the most used and consulted on portals by non-experts. Other visualization types concerning distribution, relation and part of whole should be also covered to address this gap by conducting interviews with a wider audience and by using more categories of datasets. Fourth, the results regarding the most visited datasets could be biased. The usage statistics from the Namur portal cover a brief period and might not account for instances where developers shared data directly with peers, bypassing the portal. For instance, while COVID-related datasets are currently in high demand, this trend might change post-pandemic. Future research should consider usage statistics spanning a more extended period and encompass different regions to mitigate such biases.

7.3 Dashboard Design Principles

The use of individual visualizations on open data portals has been shown to lack efficiency in reducing the information asymmetry between governments and citizens (Mellouli et al., 2014; Purwanto et al., 2020). Dashboards may be a promising way to address this problem, as individual visualizations only cover few information contained in the dataset, unlike dashboards that often incorporate more details in a single screen. The objective of this study is to identify the design principles of dashboards in the OGD context that facilitate their use and to investigate whether the use of well-designed dashboards can help citizens to engage with OGD. The contribution of this study is threefold. First, it provides dashboard designers with a list of best practices for deploying dashboards in the OGD context. Second, it applies the design principles to build a usable tool which can be used as a basis to develop other dashboards. Third, it suggests that citizens prefer to use well-designed dashboards over individual visualizations to engage with (i.e., use or explore) OGD.

The remainder of this section is organized as follows. Section 7.3.1 introduces the systematic literature review approach and presents its outputs. Section 7.3.2 offers an overview of NBDash (Namur Budget Dashboard), which incorporates the identified design principles, along with the methodology employed to assess its impact on citizen engagement. The implementation process of NBDash and the results of its evaluation are presented in Section 7.3.3. Finally, Section 7.3.4 summarizes the contributions of this section, acknowledges the study's limitations, and proposes potential avenues for future research.

7.3.1 Systematic Literature Review

In this section, we first describe the Systematic Literature Review (SLR) approach used to identify the design principles of dashboards. Then, we present the outputs of the SLR.

7.3.1.1 Systematic Literature Review Approach

From April to June 2021, we conducted a Systematic Literature Review following the procedure outlined by (Petersen et al., 2008), aiming to understand the design principles of general, city, and OGD dashboards. We do not limit the SLR to the OGD context only, as we believe that the design principles of general and city dashboards can be applied to OGD dashboards, since they also sometimes use open data. In order to cover as many relevant publications as possible, we searched the following databases: Scopus[3], Science Direct[4] and Association for Computing Machinery (ACM)[5]. With the aim of automating the search in the selected databases, the following search string was constructed using the combination of keywords from our research question **(RQ3c) "What are the design principles for effective dashboards in the context of OGD?"**: ("*dashboard" OR "*visual*") AND ("design") AND ("principle" OR "practice" OR "guideline") AND ("open government data" OR "open data" OR "government data" OR "public government data" OR "public data" OR "public sector information"). The search string was later customized based on the requirements of each database. The term "visual*", which can cover, for example, "visualization" or "visual", was added to the search terms because a dashboard is a collection of visualizations and, therefore, visualization best practices can also be applied to dashboards. Based on the automated search, we obtained 274 articles. We then identified relevant articles in three stages: first, we evaluated the type, domain and title; second, we examined the abstract; and finally, we scanned the content (see Figure 7.4).

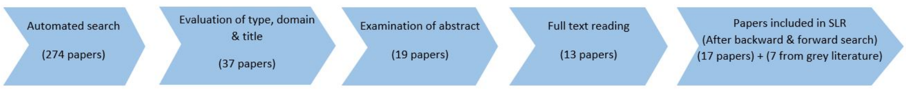


Figure 7.4: Filter processes applied in the systematic literature review (SLR) accompanied with the number of remaining papers for each step.

In the first stage, 237 publications meeting one of the following criteria were excluded from the review: duplicate papers, studies with titles that are not relevant to the keywords from RQ3c, studies published in the health sector or mathematics and studies not written in English. In the second stage, we excluded 18 publications that we deemed irrelevant to dashboard or visualization design principles because none

[3] https://www.scopus.com/
[4] https://www.sciencedirect.com/
[5] https://dl.acm.org/

of these terms were mentioned in their abstracts. In the third stage, we excluded 6 irrelevant publications because they did not provide any design principles and were more focused on describing a proposed platform (or) used for a specific domain (e.g., health, survey, learning analytic dashboards). We also applied the forward and backward search (Webster and Watson, 2002) by examining the references and citations of selected articles and added 4 additional relevant publications. In the end, we retained 17 articles that empirically explored visualization or dashboard design principles. In addition to these 17 articles, we identified 7 articles from the grey literature using the Google search engine. We incorporated these grey literature articles because they predominantly offered comprehensive summaries of best practices in dashboard design, further enriching our understanding and offering practical insights often missing from academic papers. The retained articles are listed in Table 7.5.

Table 7.5: List of 17 scientific publications and 7 grey literature contributions retained in the SLR.

Dashboard design principles	**Category**
(Kitchin and Mcardle, 2016; Matheus et al., 2020; Maheshwari and Janssen, 2014; Ganapati, 2011; Durcevic, 2020; Young and Kitchin, 2020; Young et al., 2021; Tableau, 2021)	Public sector
(Few, 2006a; Smith, 2013; Brath and Peters, 2004; Sarikaya et al., 2019; Janes et al., 2013)	General
Visualization design principles	**Category**
(Graves and Hendler, 2014; Chokki et al., 2021)	Public sector
(Shneiderman, 1996; Heer and Shneiderman, 2012; Holtz and Conor, 2018; Munzner, 2014; Ribecca, 2015; Wilke, 2019)	General
UX and UI design principles	**Category**
(Usability.gov, 2021; UXPin, 2021; Davies, 2020)	General

7.3.1.2 Systematic Literature Review Outputs: Dashboard design principles

Based on the review of the selected papers, we were able to identify dashboard design principles. In this section, we first briefly explain the different stages of the dashboard data cycle and then present dashboard design principles collected from the literature review and applied in the OGD context.

Dashboard data cycle. Figure 7.5 illustrates the different steps considered when deploying dashboards (Matheus et al., 2020; Smith, 2013). There are seven steps in the process of deploying dashboards: metrics choice (defining the metrics to include in dashboards), data collection (collecting datasets), data processing (cleaning and transforming datasets), data analysis (analyzing datasets to get clear details about their content), build the dashboard (creating adequate visualizations for each defined metric, build the dashboard layout and integrate visualizations in the layout) and deployment (sharing the created dashboard to citizens). After the deployment, the citizens can then use the dashboard and provide feedback or the dashboards can be customized by adding more datasets or changing metrics.

Table 7.6 provides a summary of the dashboard design principles from the literature review with their application in the OGD context.

Table 7.6: Dashboard design principles in the OGD context.

Design principle	**Application in the OGD context**
P1. Pick meaningful metrics (Kitchin and Mcardle, 2016; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Ganapati, 2011; Sarikaya et al., 2019; Durcevic, 2020)	Using meaningful and understandable indicators that are relevant to citizens help them to understand the utility of the implemented government dashboards, creates added value and increases their usability.
P2. Collect accurate and precise data (Matheus et al., 2020; Ganapati, 2011; Young and Kitchin, 2020; Janes et al., 2013)	Data quality and lack of understanding of published data are among the reasons why open data are not used (Crusoe et al., 2019). Government dashboards should therefore avoid using ambiguous or unreliable data and metadata in order to make the data easier to understand and to reassure users of its veracity.
P3. Ensure your data makes sense (Matheus et al., 2020; Kitchin and Mcardle, 2016)	A lot of open data is made available online and integrated into dashboards after aggregating some initial data. It is therefore very important to ensure that the data is consistent before and after the transformation process in order to assure the quality and the veracity of data.
P4. Consider audience (Brath and Peters, 2004; Sarikaya et al., 2019; Durcevic, 2020; Young and Kitchin, 2020; Young et al., 2021; Tableau, 2021)	Referring to previous studies (Young et al., 2021; Lassinantti et al., 2019; Graves and Hendler, 2014; Nikiforova and Lnenicka, 2021), there are users from different backgrounds who are interested in using Open data. Thus, since not all users have the same technical skills to understand data and visualization, it is important to consider the target audience before implementing government dashboard or, if possible, to propose different dashboards according to the type of user, as was done in Dublin dashboard (Young et al., 2021).
P5. Use best visualization practices (Few, 2006a; Matheus et al., 2020; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Ganapati, 2011; Sarikaya et al., 2019; Durcevic, 2020; Young and Kitchin, 2020; Janes et al., 2013)	According to survey results on visualizations for OGD (Chokki et al., 2021), citizens like to have simple and attractive visualizations that can help them to understand the data without much effort. For example, simply arranging the data in descending order in a bar chart can help users quickly grasp the highest and lowest values of the data presented. More details on the design principles of visualizations can be found in (Graves and Hendler, 2014; Chokki et al., 2021; Shneiderman, 1996; Heer and Shneiderman, 2012).

Design principle	Application in the OGD context
P6. Use the right type of chart (Few, 2006a; Matheus et al., 2020; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Sarikaya et al., 2019; Durcevic, 2020; Young and Kitchin, 2020; Janes et al., 2013)	According to previous studies (Chokki et al., 2021; Holtz and Conor, 2018; Munzner, 2014; Ribecca, 2015; Wilke, 2019), visualizations are very useful for understanding the information in the data, but they should be carefully chosen according to the types of data and the target audience in order to be easily understandable. So, for example, using a treemap to show population distribution in a city can be very useful for journalists, but can be difficult to understand for ordinary citizens who are not familiar with visualization. Thus, in government dashboard, these parameters should be taken into account before choosing the visualization for a specific metric.
P7. Provide easy to use tools (Kitchin and Mcardle, 2016; Matheus et al., 2020; Young and Kitchin, 2020)	Only a few cities and open data portals provide dashboards for their data and the reasons are that they do not have sufficient technical skills, time and financial resources to build and update these dashboards (Purwanto et al., 2020). In order to overcome that issue, one solution would be to provide free and easy to use tools that could help them to create their dashboards and thus facilitate the understanding of published data by citizens and improve their reuse.

Design principle	**Application in the OGD context**
P8. Clear presentation (Kitchin and Mcardle, 2016; Few, 2006b; Matheus et al., 2020; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Ganapati, 2011; Durcevic, 2020; Young and Kitchin, 2020; Tableau, 2021)	• The layout and the presentation of the dashboard is very important for its success and should be also adapted to the device screen size (e.g., phone, tablet, PC) (Tableau, 2021). • When presenting government dashboards, it is recommended to display key information at the top of the screen (using for example single value charts or a single value with indicator) followed by more details (using advanced visualizations: e.g., line chart, bar chart) (Matheus et al., 2020). • It is also recommended to group together items within the same domain in case the dashboard covers different topics, as is the case in government dashboards. More common mistakes to avoid a messy and unclear presentation are presented in (Few, 2006a; Sarikaya et al., 2019; Tableau, 2021). • In addition, some tips about UX and UI design principles such as: avoid unnecessary elements and be clear in the language used on labels and in messaging, consider the spatial relationships between items on the page and structure the page based on importance, use of different sizes, fonts and arrangement of the text to help to increase scalability, legibility and readability, etc. are presented in (Usability.gov, 2021; UXPin, 2021; Davies, 2020).
P9. Provide context and data interpretation support (Maheshwari and Janssen, 2014; Durcevic, 2020; Young and Kitchin, 2020; Young et al., 2021)	Not all citizens have the ability to easily understand visualization, so it is important to have additional contextual information or metadata to clarify the meaning of data and to avoid misinterpretation for each visualization presented on dashboard governments. In addition, each visualization should have a title and titled axes.
P10. Think about data literacy levels (Sarikaya et al., 2019; Young and Kitchin, 2020; Young et al., 2021)	As mentioned earlier, one of the reasons for the lack of use of open data is the lack of clarity in the metadata provided on open data portals. Therefore, dashboard governments need to use clear and consistent terminology, familiar words, phrases and concepts to explain their purpose and then allow citizens from different backgrounds to understand the dashboards.

Design principle	Application in the OGD context
P11. Ensure data is up to date (Kitchin and Mcardle, 2016; Matheus et al., 2020; Young and Kitchin, 2020)	The government dashboards are mainly used to verify some information and then make decisions. Therefore, it is very important to let the citizens know that the data is current or when it was extracted, which can really help the citizens to make a correct decision.
P12. Allow access to data source (Kitchin and Mcardle, 2016; Matheus et al., 2020; Sarikaya et al., 2019)	Many government dashboards sometimes use data that is collected from external organizations and thus not accessible and sometimes also provide no link to access the data used. Allowing access to data sources will enable customization and increase user confidence in the implemented dashboard.
P13. Check for personal data/outliers (Kitchin and Mcardle, 2016; Matheus et al., 2020; Sarikaya et al., 2019)	Since much of governments' data is obtained by collecting citizen data, they need to ensure the confidentiality of the data when using it in dashboards, for example by aggregating data and avoid the risk of the data becoming identifiable in any way. This will then help to solve the data privacy issue observed in dashboards.
P14. Interaction support (Matheus et al., 2020; Maheshwari and Janssen, 2014; Sarikaya et al., 2019; Durcevic, 2020)	Government dashboards should offer interaction features such as the ability to hover over an item in the visualizations to get more details, the ability to use filters to update data in the dashboards and the ability to dig deep into certain trends, metrics, or insights with ease, instead of displaying static visualizations in dashboards that can be good for novice users but not for some less advanced and advanced users such as journalists, or developers, who request more interaction. Thus, it is important to add interaction to support different types of users.
P15. Ensure feedback support (Matheus et al., 2020; Durcevic, 2020; Tableau, 2021)	One factor influencing citizen engagement with OGD is the availability of OGD feedback mechanisms. Therefore, government dashboards should never stop evolving and allow users to provide feedback through user testing during development or after deployment that can be later used to improve the layout, functionality, look and feel of the dashboard to ensure optimal value at all times. In addition, feedback feature that can help citizens to report suspected fraud or corruption to an independent and trusted agency can be a good element to encourage citizens to use dashboards.

Design principle	Application in the OGD context
P16. Customization (Sarikaya et al., 2019; Young and Kitchin, 2020; Young et al., 2021)	Citizen reuse is one of the open data initiatives. Government dashboards should provide citizens with all the information needed to customize existing dashboards in order to facilitate reproducibility that can improve trust and also allow citizens to create an improved version of dashboards without contacting the dashboard manager.

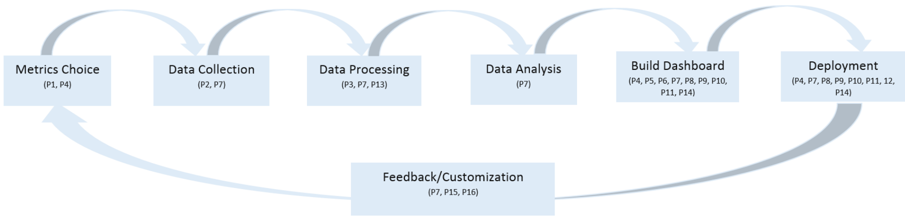


Figure 7.5: Dashboard data cycle with associated design principles.

7.3.2 NBDash Overview and Evaluation Method

After gathering the list of dashboard design principles, we applied them to build the Namur Budget Dashboard (NBDash) in order to have an example of a well-designed dashboard for the participants to use during the evaluation. NBDash is a web application built using two budget datasets (Namur-Ordinary Budget by function[6] & Namur-Extraordinary Budget by function[7]) available on the Namur open data portal[8]. It is the official open data portal of the city of Namur (Belgium) created in 2018 with several objectives of information, statistics, creation of useful applications for citizens and transparency. As of August 2021, the portal contains 174 datasets across 13 topics such as urbanism, population, administration, transport, culture, environment, health, sport, energy, economy, education, internal data and closed data (data that can be accessed by a specific group of users). These datasets are either statistics, which can be consulted online, or batches of data which can be used directly for the creation of applications. We selected the budget datasets for two reasons. First, transparency-related datasets in general (and budget datasets in particular) are very interesting for citizens (Chokki et al., 2021; Araújo et al., 2016; Corrêa et al., 2014). Second, creating a dashboard with these datasets can allow citizens to see how the budget of the municipality is dispatched and can then increase transparency. In addition, we chose the portal of Namur as there is no budget dashboard available for this municipality and access with key stakeholders of this portal was possible. More details on the development of NBDash will be presented in Section 7.3.3.1.

An evaluation was later conducted to determine whether citizens prefer to use well-designed dashboards rather than the individual visualizations offered on a traditional OGD portal, that do not fully incorporate some of the identified design principles (such as feedback support, data literacy levels and data interpretation) to engage with OGD. We adopted an experimental design based on the static-group comparison model and thus divided the participants into two groups (Campbell and Stanley, 1969) (p12) (the profile of participants in each group is presented in Section 7.3.3.2 and further details on participants' recruitment are provided below). The first group (control group) evaluated only the individual visualizations presented on the Namur portal for both budget datasets (see Figure 7.6). The second group (treatment group) evaluated only the NBDash dashboard. The participants' feedback was collected through a survey consisting of 16 questions with a 5-point Likert scale (from "Totally Disagree" to "Totally Agree") accompanied with a free text to justify their ratings and 3 additional questions to collect demographic data (age, gender, education). At the beginning of the questionnaire where the context of the survey was presented, we mentioned that these two datasets were used as illustrations and

[6]https://rb.gy/61r8dk
[7]https://rb.gy/dpayws
[8]https://data.namur.be/

the participants' evaluation should not be only based on these specific datasets but on dashboards and individual visualizations in general.

The first fifteen questions were carefully constructed to correspond to the following constructs (conditions and factors) that have been proven to impact citizen engagement in (Purwanto et al., 2020) and also to verify the implementation of the defined design principles: C1) citizens perceived ease of use (refers to effort expectancy or citizens' perceived ease of engagement in (Purwanto et al., 2020)), C2) diversity of citizens' skills and capabilities, C3) citizens perceived data veracity and quality (e.g., accuracy, completeness, timeliness (Dekkers et al., 2014; Nikiforova, 2020). See example in section 3 of Figure 7.7a). The last question Q16 was added to gather the general opinion on RQ3d and thus to verify whether the use of dashboards or visualizations encourages citizens to engage more with OGD. We select these 3 constructs among the 15 constructs identified from (Purwanto et al., 2020) because these constructs are independent of the political, financial or social conditions and factors (e.g., citizen motivation and citizens/government resources) and can therefore be evaluated using tools. The list of questions of the surveys are presented in Appendix 7.7. Table 7.8 summarizes the corresponding questions for each factor, and also explains why we chose them and provides references where appropriate. Two surveys (one for visualizations and one for dashboards) were distributed to collect feedback from participants using dragnsurvey[9] and were pretested with two citizens to ensure all kinds of errors associated with survey research are reduced (Grimm, 2010). Next, we shared the surveys on social networks such as Facebook groups and Twitter to recruit participants. We later used Amazon Mechanical Turk[10] to recruit other participants. We do not set any conditions that participants must meet in order to complete the survey, except that they must properly justify their choices. We chose this option because we want to have different profiles among the participants. The Amazon Mechanical Turk was used because after one month of posting the surveys on social networks, we observed that only 3 participants responded to the surveys and this tool has proven effective and reliable in previous studies (Crowston, 2012; Berinsky et al., 2012). Each participant received a $1.5 compensation for completing the survey, which takes about 15 minutes.

7.3.3 Results: The impact of NBDash on citizen engagement

In this section, we first describe the development of NBDash. Then we present the results of the evaluation.

7.3.3.1 NBDash: System Description

NBDash is a web application (source code available at https://github.com/chokkipaterne/nbdash) built using three technologies: Python as the programming language, Pandas[11] as the data processing library and Dash Plotly[12] to create and display the visualizations on the web page. The deployment of NBDash was done after following each step of the deployment process shown in Figure 7.5 and applying the associated design principles for each step.

[9] https://www.dragnsurvey.com
[10] https://www.mturk.com
[11] https://pandas.pydata.org/
[12] https://plotly.com/dash/

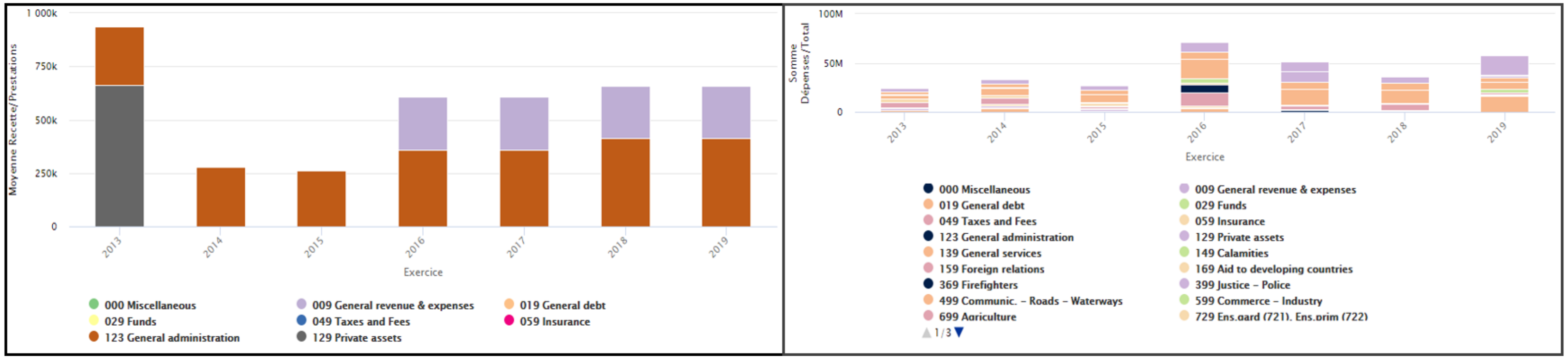


Figure 7.6: Individual visualizations proposed on Namur portal for the ordinary and extraordinary budget datasets (translated to English using Google translate). (Left): Average ordinary service revenue by function over time and (Right): Total extraordinary expense by function over time.

Table 7.7: Survey questions for the evaluation.

Questions for [visualizations/dashboard] to address citizen engagement
Q1. I think that I would like to use [these visualizations/this dashboard] frequently
Q2. I found [these visualizations/this dashboard] unnecessarily complex
Q3. I thought [these visualizations/this dashboard] were easy to use
Q4. I think that I would need the support of a technical person to be able to use [these visualizations/this dashboard]
Q5. I found the various functions in [these visualizations/this dashboard] were well integrated
Q6. I thought there was too much inconsistency in [these visualizations/this dashboard]
Q7. I would imagine that most people would learn to use [these visualizations/this dashboard] very quickly

Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
+ Free text to justify their choice

Table 7.7: Survey questions for the evaluation.

Questions for [visualizations/dashboard] to address citizen engagement	
Q8. I found [these visualizations/this dashboard] very difficult to use	
Q9. I felt very confident using [these visualizations/this dashboard]	
Q10. I needed to learn a lot of things before I could get going with [these visualizations/this dashboard]	
Q11. I can easily tell what we can learn from the datasets based on [these visualizations/this dashboard]	
Q12. I can easily draw conclusions based on [these visualizations/this dashboard]	
Q13. I can easily understand [these visualizations/this dashboard]	
Q14. I can easily modify or customize [these visualizations/this dashboard] to see other aspects of the datasets	
Q15. [these visualizations/this dashboard] provide(s) me necessary information to verify that the data used are accessible, accurate and up to date and to easily access the quality of the datasets	
Q16. Using this type of [visualization/dashboard] to present data makes me want to engage with (i.e., explore, understand and use) more data on a portal	
Demographic questions	
Q17. How old are you?	[18 - 29 / 30 - 49 / 50 +]
Q18. What is your gender?	[Female / Male / Other]
Q19. What is your level of education?	[None / Primary / High School / High Education / PhD]

Table 7.8: Correspondence between citizen engagement constructs and survey questions.

Constructs	Questions	Comments
C1) citizens perceived ease of use	10 questions of the SUS questionnaire (Q1 to Q10)	SUS was used as it is suitable to measure the usability of a system in a standalone and also to compare the usability of multiple systems (Brooke, 1986, 2013)
C2) diversity of citizens' skills and capabilities	Q11. I can easily tell what we can learn from the datasets based on [these visualizations/this dashboard] (Refer to P8, P9) Q12. I can easily draw conclusions based on [these visualizations/this dashboard] (Refer to P8, P9) Q13. I can easily understand [these visualizations/this dashboard] (Refer to P10) Q14. I can easily modify or customize [these visualizations/this dashboard] to see other aspects of the datasets (Refer to P16)	These questions were constructed with reference to the design principles that need to be covered to consider that a system takes into account the skills of end-users (P4. Consider audience). We have mainly based on the following design principles to formulate these questions: P8. Clear presentation, P9. Provide context and data interpretation support, P10. Think about data literacy levels, and P16. Customization
C3) citizens perceived data veracity and quality (citizens' confidence in the veracity of data and ease of access to data quality)	Q15. [these visualizations/this dashboard] provide(s) me necessary information to verify that the data used are accessible, accurate and up to date and to easily access the quality of the datasets (Refer to P2, P12)	This question was constructed with reference to design principles regarding data quality and veracity (P2. Collect accurate and precise data, P12. Allow access to data source)

Metrics choice & Data Collection. In this step, we defined metrics with reference to existing metrics in the London budget dashboard[13] and validated these metrics by gathering feedback from two citizens (P1). The metrics were also categorized based on the users' skills in order to display only necessary and understandable metrics to the users (see Table 7.9) (P4). Later, we collected the latest update of the budget datasets related to the selected metrics on the reliable open data portal of Namur (P2 and P11). We also provided access links to the datasets to users for possible reuse (P12) and calculated the data quality of each dataset based on the basic features: missing values, data information and metadata information (column titles and descriptions) (Vetrò et al., 2016), to indicate users the quality of the data (see section 3 of Figure 7.7a) (P2).

Table 7.9: Metrics in NBDash based on the different display types. [ordinary/extraordinary] and [revenues/expenses] are used as filters.

Metrics	Display Type
M1. Total [ordinary/extraordinary] [revenue/expense] for specific year with possibility to compare to the previous year	All (Simple, Less advanced, Advanced)
M2. Analysis of [ordinary/extraordinary] [revenue/expense] by function for specific year	All
M3. Analysis of [ordinary/extraordinary] [revenue/expense] by [revenue/expense] type for specific year	All
M4. Analysis of [ordinary/extraordinary] [revenue/expense] by function over time	Less advanced, Advanced

Data Processing & Data Analysis. In this step, we used Excel formulas in parallel to Pandas to aggregate and filter data according to each metric in order to ensure that we get the same results for both options (P3). Data aggregation was also used to ensure that personal data is not disclosed (P13).

Build Dashboard & Deployment. In this step, we implemented three display types in order to accommodate for the skills of the audience which vary from novice (low knowledge in visualizations) to advanced users (high knowledge in visualizations) (Young et al., 2021). The display types are as follows: "simple" display for novice users on visualizations (see Figure 7.7c), "less advanced" display for users who need more control over the data and visualizations displayed on dashboard (see Figure 7.7b) and "advanced" for users who want to customize the dashboard (see Figure 7.7a) (P4). For each display type, we organized the metrics that represent the big picture of the data on the top of the dashboard, followed by the metrics that provide more details (see sections 7 & 8 of Figure 7.7a) (P8). Three main visualization types were used to represent the selected metrics: a bar chart with descending sorting and single color for the representation of categorical data (see section 8 of Figure 7.7a), a pie chart for the representation of proportions and a line chart for the temporal data (P5 and P6). In order to help users to understand the graphs and avoid misinterpretation, we provided the chart title, axes titles and a small interpretation for the graphs (see Figure 7.7c) (P9). We noticed that datasets contained many financial terms. To help users understand key aspects of the graphs, we created a terminology section to explain key budget terms (see section 2 of Figure 7.7a) and also used easy to understand words to represent title, axes, interpretation of graphs and to design the

[13] http://openbudget.lacity.org/#!/year/default

layout of the dashboard (P10). In addition, we also added filters to the dashboard (e.g., fiscal year, function, budget type) to allow users to update the data displayed and change the visualization type (see sections 6 & 9 of Figure 7.7a) in order to support interaction (P14).
Feedback/Customization. In this step, we provided a feedback form to users and allowed them to track the status of their feedback (see section 4 of Figure 7.7a) (P15). We also added an “Edit chart” button to allow users to directly edit the represented graph in the chart studio of Plotly[14] and also provided access to source code that can be used to enhance NBDash or create a new dashboard (see sections 5 & 8 of Figure 7.7a) (P16).

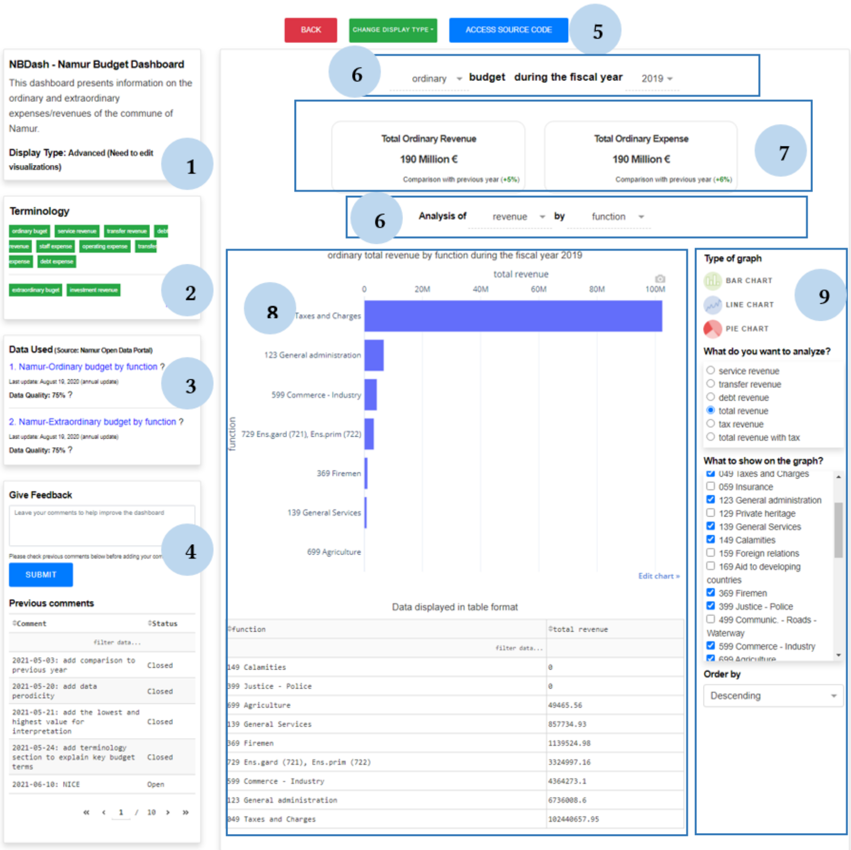
Figure 7.7 shows the interface of NBDash for the three different display types. The layout is nearly the same for all display types, with the exception of sections 5, 8 and 9 which are slightly modified depending on the display type. Table 7.10 details the difference between these three display types.
Figure 7.7(a) NBDash Interface for advanced display. (1): Description of the dashboard. (2): Explanation of the key budget terms used. (3): Information about the data used. (4): Possibility to give feedback and track status (5): Possibility to choose the display type (simple, less advanced and advanced) and to download source code. (6): Filters used to update the visualizations. (7): Display of the data overview. (8): Visualization for more details with editing option followed by data presented in table form. (9): Advanced filters used to update visualizations.
Figure 7.7(b) NBDash Interface for “simple” display. Adding a small interpretation of the graph, removing advanced filters and deleting “Edit chart”& “Access source code” buttons compared to “advanced” display.
Figure 7.7(c) NBDash Interface for “less advanced” display. No “Edit chart” and “Access source code” buttons compared to “advanced” display.

7.3.3.2 Insights from the Experimental Study

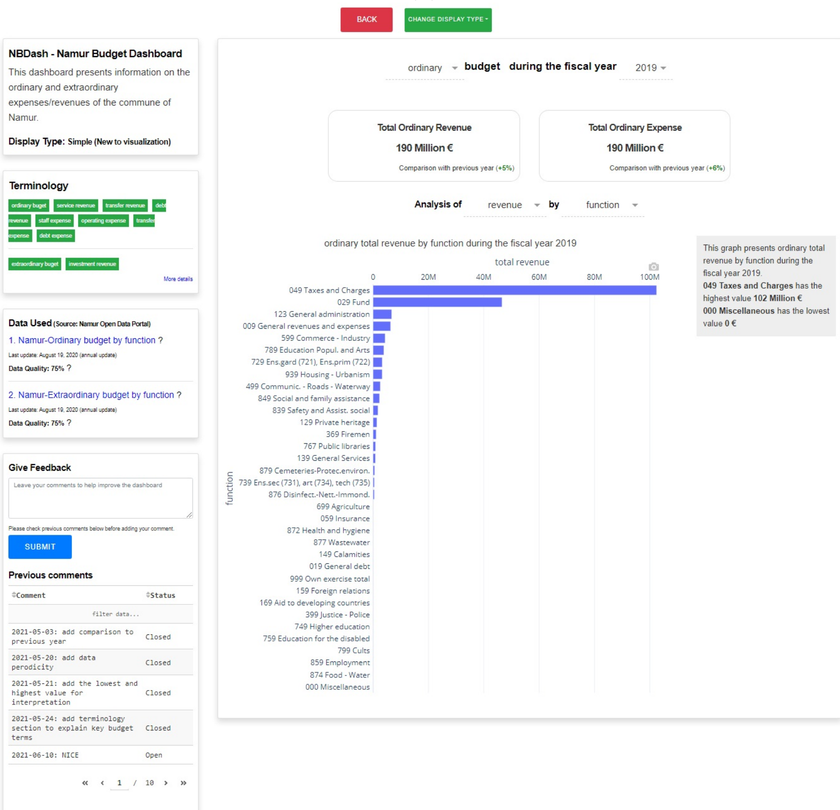
Through questionnaires that participants completed after exploring the visualizations on the Namur Portal and NBDash, we were able to gather their opinions on **(RQ3d) "In what ways do well-structured dashboards, incorporating best dashboard design principles, enhance citizen engagement with OGD compared to individual visualizations?"**. A total of 50 participants completed the survey on visualizations and 58 completed the survey on NBDash. A minimum of 50 participants were recruited for each group because when referring to previous studies (Faulkner, 2003; Nielsen, 2000; Six and Macefield, 2016; Alroobaea and Mayhew, 2014), using 5-50 participants for comparison or usability tests is a good baseline. Table 7.11 presents the demographic representation of participants for both surveys. We did not observe a significant relationship between this demographic information and participants' choices. We will therefore not discuss it further in this study.

[14]https://chart-studio.plotly.com/

(a)


BACK
CHANGE DISPLAY TYPE
ACCESS SOURCE CODE
NBDash - Namur Budget Dashboard
This dashboard presents information on the ordinary and extraordinary expenses/revenues of the commune of Namur.
Display Type: Advanced (Need to edit visualizations)
Terminology
Data Used
1. Namur-Ordinary budget by function
2. Namur-Extraordinary budget by function
Give Feedback
SUBMIT
Previous comments
ordinary budget during the fiscal year 2019
Total Ordinary Revenue
190 Million €
Total Ordinary Expense
190 Million €
Analysis of revenue by function
ordinary total revenue by function during the fiscal year 2019
total revenue
Data displayed in table format
Type of graph
BAR CHART
LINE CHART
PIE CHART
What do you want to analyze?
What to show on the graph?
Order by
Descending
1
2
3
4
5
6
6
7
8
9

(b)


BACK
CHANGE DISPLAY TYPE
NBDash - Namur Budget Dashboard
This dashboard presents information on the ordinary and extraordinary expenses/revenues of the commune of Namur.
Display Type: Simple (New to visualization)
Terminology
Data Used
1. Namur-Ordinary budget by function
2. Namur-Extraordinary budget by function
Give Feedback
SUBMIT
Previous comments
ordinary budget during the fiscal year 2019
Total Ordinary Revenue
190 Million €
Total Ordinary Expense
190 Million €
Analysis of revenue by function
ordinary total revenue by function during the fiscal year 2019
total revenue
This graph presents ordinary total revenue by function during the fiscal year 2019.
049 Taxes and Charges has the highest value 102 Million €
000 Miscellaneous has the lowest value 0 €

(c)

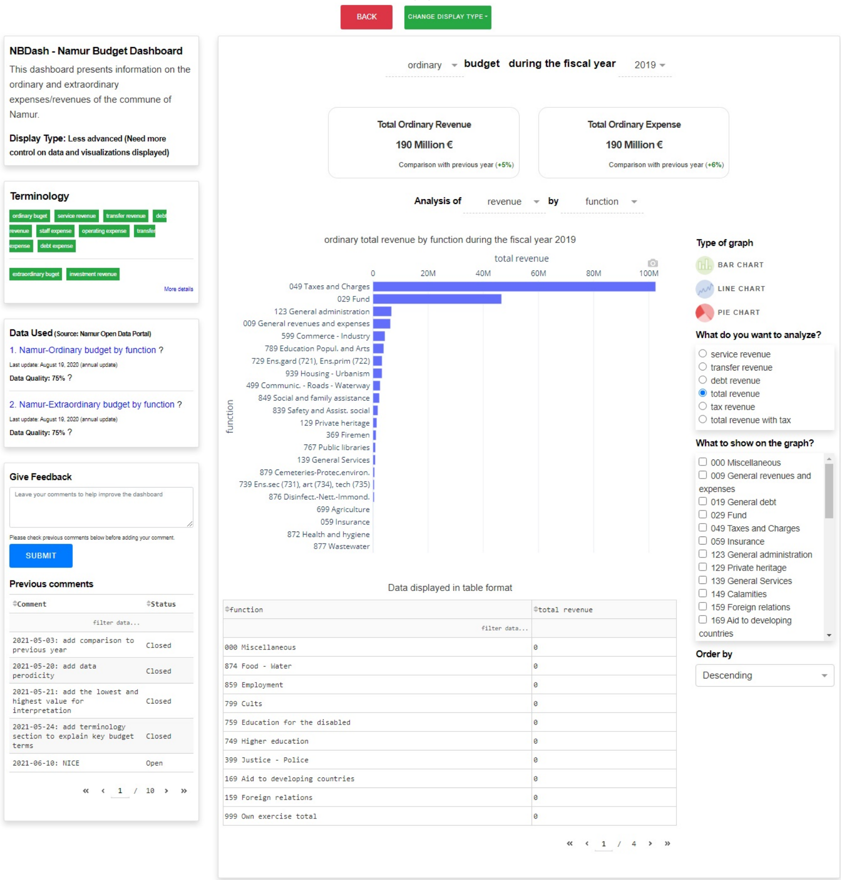


Figure 7.7: NBDash Interface for different display types.

Table 7.10: Difference between the different display types in NBDash.

Criteria	**Display types in NBDash**		
	Novice	**Less Advanced**	**Advanced**
User type	Used by users who have a low level of visualization knowledge such as the general public.	Used by users who have a middle level of visualization knowledge such as public servants using visualization often in their work.	Used by users who have a high level of visualization knowledge such as developers.
Section 5	The “access source code” button is not visible.	The “access source code” button is not visible.	The “access source code” button is visible.
Section 8	We show only a bar chart. The “edit chart” menu is not visible.	We show the selected chart (bar, line or pie) followed by the data in table format. The “edit chart” menu is not visible.	We show the selected chart (bar, line or pie) followed by the data in table format. The “edit chart” menu is visible.
Section 9	We hide the advanced filters. We also display only a small interpretation of the chart.	We show the advanced filters (e.g., fiscal year, function, budget type) to allow users to update the data displayed and change the visualization type. We hide the chart interpretation.	We show the advanced filters (e.g., fiscal year, function, budget type) to allow users to update the data displayed and change the visualization type. We hide the chart interpretation.
Other sections (1, 2, 3, 4, 6, 7)	Other sections remain the same except for the text about the display type in section 1 which is updated according to the chosen display type.		

Table 7.11: Demographic data of surveys.

		Group 1 (Visualizations) 50 participants	**Group 2 (NBDash)** 58 participants
Sex	Male	31 participants (62%)	39 participants (67.24%)
	Female	19 participants (38%)	19 participants (32.76%)
	Other	0 participants (0%)	0 participants (0%)
Age	18 - 29	21 participants (42%)	22 participants (37.93%)
	30 - 49	25 participants (50%)	32 participants (55.17%)
	50 +	4 participants (8%)	4 participants (6.9%)
Education	None	0 participants (0%)	0 participants (0%)
	Primary	0 participants (0%)	0 participants (0%)
	High School	3 participants (6%)	3 participants (5.17%)
	High Education	39 participants (78%)	50 participants (86.2%)
	PhD	8 participants (16%)	5 participants (8.63%)

In order to analyze the gathered responses, we computed the average (avg) along with the standard deviation (σ). The average served to ascertain the central tendency of the responses, providing a summary of the central point of our dataset. The standard deviation was calculated to measure the amount of variation or dispersion in the responses, offering insights into the diversity of views among our respondents. Additionally, we carried out a one way ANOVA test (especially the p-values) to verify the statistical significance of the answers between the two groups. Table 7.12 presents the corresponding p-values, averages, as well as the standard deviations of the responses for each group. The visualization group is referred to as group 1 and the dashboard group is referred to as group 2. The results of Table 7.12 show that NBDash offers greater usability than individual visualizations, as its average SUS score (76.85) is higher than 68, which is the minimum required (Brooke, 1986) and also greater than the average SUS score of individual visualizations. The results also show that there is significant variation between the SUS scores of participants in the two groups ($\sigma \geq 11$). On the basis of this first part of the results, we can therefore deduct two things: (1) citizens perceived that the well-designed dashboard is easy to use and (2) citizens perceived that the well-designed dashboard is easier to use than individual visualizations which have a usability score below the threshold. Regarding statements Q11, Q12 and Q13, the results show that the averages of the scores collected on these three questions by the group 2 participants (avg resp. 4.21, 4.31, 4.5) are higher than those collected by group 1 participants (avg resp. 3.98, 4.08, 4.14) and there is no significant variation between the scores ($\sigma \leq 1$ for Q11 to Q13). For statement Q14, more participants in group 1 found easy to modify or customize the individual visualizations (avg = 4) than participants in group 2 (avg = 3.97). However, there is a smaller difference between these Avg scores (diff = 0.03) compared to the previous differences and there is also no significant variation between the scores ($\sigma \leq 1$ for Q14). Based on the scores of these four statements (Q11 to Q14), we can therefore infer that participants agree that well-designed dashboards take into account the diversity of skills and capabilities rather than individual visualizations. Indeed, participants with different levels of education in group 2 were more likely to be able to easily understand, draw conclusions and modify (with a smaller difference between the two groups in this aspect) than those in group 1. Regarding the statement about the evaluation of the data quality (Q15), there are more participants in group 2 (avg = 4.23) compared to group 1 (avg = 4.02) who agreed that they have necessary information to evaluate the quality of the datasets. We can therefore deduct that citizens more easily perceived data veracity and quality in using well-designed dashboard compared to individual visualizations. The results on the statement Q16 show that there are more participants in group 2 (avg = 4.43 ($\sigma \leq 1$)) compared to group 1 (avg = 4.02 ($\sigma \leq 1$)) that agreed that they will be interested to explore, understand and use (engage with) more data on a portal if the data is presented with this type of dashboard compared to the individual visualizations. We can therefore deduct that citizens would be more interested to engage with OGD if there were represented using well-designed dashboard than individual visualizations. In addition, the results of the p-values show that the differences between NBDash and individual visualizations were statistically significant (p-value$\leq$0.05) for the answers to questions Q11, Q13, and Q16. Based on all these conclusions, we can respond to RQ3d by saying that the use of well-designed dashboards can help citizens to engage more with OGD than individual visualizations.

These observations can be justified by the following reasons. First, according to participants' comments, the visualizations used on NBDash are easier to understand than those on the Namur portal. On the Namur portal, some participants found that the visualizations contained too much data and that the visualization technique used to represent data was not easy to understand. Second, many participants found NBDash to be user-friendly and well-organized and therefore easy to use and understand. However, two participants of group 2 disagreed with this statement. They thought that there was a lot of information in the dashboard and too much text. For example, they suggested hiding the terminology, feedback and data used sections and only displaying them when the user requests so. Third, participants found that on the Namur portal, they can easily modify the visualizations compared to NBDash. Five participants of group 2 cannot figure out how to modify or customize the visualization in NBDash, probably because they were on the “Simple” display while the option to modify the chart is available on the “Advanced” display. Fourth, in NBDash participants perceived that they can more easily access data quality information than in individual visualizations because in NBDash, we clearly specified the data used accompanied with their sources and the last update time and also evaluated the data quality to help users to have an idea about it without accessing them.

Table 7.12: Average (avg), standard deviation (σ) and p-value of survey scores.

	Visualizations on Namur Portal		**NBDash**		**p-value**
	avg	σ	**avg**	σ	
SUS Score (Q1 to Q10)	66.3	12.81	**76.85**	11.18	-
Q11	3.98	0.71	**4.21**	0.59	**0.04**
Q12	4.08	0.9	**4.31**	0.79	0.16
Q13	4.14	0.81	**4.5**	0.54	**0.006**
Q14	**4**	0.86	3.97	0.88	0.83
Q15	4.02	0.77	**4.23**	0.83	0.13
Q16	4.02	0.89	**4.43**	0.57	**0.004**

Another finding from the participants' comments is that the most important design principles for the participants are the selection of meaningful metrics (P1), the use of appropriate visualization (P5 and P6) and a clear presentation and design layout (P8). This is supported by the fact that many participants justified their ratings on the basis of these four design principles. Other design principles were also important to the participants, as without them we would not achieve a sufficient SUS score. However, the design principle of providing context and assistance in interpreting the data (P9) was somewhat criticized by some participants, as applying this design principle added more text to the dashboard. Therefore, we need to find a compromise between providing more details and providing little details in order to not clutter the dashboard.

7.3.4 Conclusion

The objective of this study is to identify the design principles of dashboards in the OGD context (RQ3c) that facilitate their use and to investigate whether the use of well-designed dashboards can help citizens to engage with OGD (RQ3d). To address RQ3c, a systematic literature review was conducted, which allowed us to provide 16 design principles applicable to OGD dashboards. To address RQ3d, we developed the Namur Budget Dashboard (NBDash) that implements the mentioned design principles and then compared it to the budget visualizations on the Namur portal in terms of ease of use, diversity of citizens' skills and capabilities, data veracity and quality and citizens' intention to engage.

This study contributes to theory and practice in the following aspects. First, this study extends this recent work (Purwanto et al., 2020) by using its recommended conditions for the emergence of OGD citizen engagement to propose some design principles that can be incorporated into dashboards to fulfill these conditions. Second, unlike previous studies (Kitchin and Mcardle, 2016; Few, 2006a; Matheus et al., 2020; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Ganapati, 2011; Durcevic, 2020; Young and Kitchin, 2020; Young et al., 2021; Tableau, 2021) that have focused on design principles for general or city dashboards, it contributes to the knowledge base by proposing 16 design principles with a clear application in the OGD context (see Table 7.6). Therefore, the design principles provided can be used by dashboards designers and OGD managers to implement usable and understandable dashboards that can then improve citizen engagement with OGD. Third, different from previous studies (Kitchin and Mcardle, 2016; Few, 2006a; Matheus et al., 2020; Maheshwari and Janssen, 2014; Brath and Peters, 2004; Ganapati, 2011; Young and Kitchin, 2020; Young et al., 2021), we show through a concrete case study how to apply each design principle. Thus, this case study can be a source of inspiration for dashboards designers and OGD managers to create their own dashboards using OGD. Fourth, we provide access to the source code of the case study. This can be used as a starting point by dashboards designers and OGD managers to create their own dashboards or improve the Namur budget dashboard. Fifth, the usefulness of the dashboards in helping citizens to engage with OGD on the portals was proven on the basis of the evaluation results. We suggest OGD managers to provide more dashboards on their portals and also follow the design principles to make them easy to use and understand. The evaluation results also showed that while all design principles are important for citizen engagement with OGD through dashboards, choosing meaningful measures (P1), using appropriate visualization (P5, P6) and clear presentation and layout (P8) are the most important.

However, this study has some limitations that will need to be addressed in future work. The first limitation concerns the representativeness of the participants in the evaluation. The use of Amazon Mechanical Turk can be a bias (Dupuis et al., 2013; Ehrich, 2020) but in our research we tried to minimize this by following best practices [60], (Cobanoglu et al., 2021) such as using strict criteria to select relevant participants and also checking the consistency of the participants' comments with their ratings before validating their submission because we noticed that some participants were not fluent in English and also sometimes their ratings and their comments did not match. However to solve this issue, we suggest using other channels of communication or collecting data on-site in administrations, universities or public places. In this study, this was not feasible due to the COVID-19 situation. The second limitation resides in the use of three of the factors mentioned in (Purwanto et al., 2020; Kitchin and Mcardle, 2016) that impact citizen engagement, to define our design principles. Other factors that were excluded in this study concern for example citizen motivation and citizens/government resources, which we believe communication and financial resources can be used to address. However, other researchers can build on our study and investigate whether there are design principles that can address these remaining factors. The third limitation is the non-implementation of generic tool for OGD dashboards. Other researchers and programmers can build on the implemented dashboard as well as the proposed design principles to implement a usable tool that can be generic and can help OGD managers to easily create dashboards that follow all of these best design principles. Further research may also involve collecting additional data following the Unified Theory Of Acceptance And Use Of Technology (UTAUT) (Venkatesh et al., 2003; Momani, 2020) model to check whether demographic and social factors actually impact citizen engagement, as we were unable to cover this aspect in this study due to the sample size and distribution. The UTAUT model is suggested because it includes four main constructs, namely: performance expectancy, effort expectancy, social influence and facilitating conditions while accommodating four moderators: age, gender, voluntariness and experience. Therefore, it is the most appropriate model for assessing the impact of demographic and social factors compared to other models (e.g., TAM, TOE) that do not incorporate these factors.

7.4 Identifying Requirements for Data Storytelling Tools

In this section, we introduce the essential features required in the design of a generic data storytelling tool that can help users with low and high data manipulation skills to easily turn OGD into stories. These features are later validated through the use of a functional tool called ODE (Open Data Explorer), which serves as a proxy and incorporates these features. Additionally, we present an extended version of the ODE tool called StoryOGD, specifically designed to assist publishers in presenting their portals in a user-friendly manner.

The remainder of this section is structured as follows. We present the background related to data storytelling (Section 7.4.1.1), and existing data storytelling tools (Section 7.4.1.2), explain the methodology used to address the research questions (Section 7.4.2), present the proposed ODE tool (Section 7.4.3.2), its extended version StoryOGD (Section 7.4.3.3), and its evaluation, discuss the findings and limitations of this study (Section 7.4.3.4) as well as avenues for future work, and conclude with a summary of the contributions (Section 7.4.4).

7.4.1 Background

In this section, we first clarify the concept of data storytelling. Then, we present its stages in the OGD context. Finally, we present tools used in the literature to facilitate data storytelling.

7.4.1.1 Data Storytelling

Data storytelling can be explained as a process of translating data analysis into simple, logical stories that can be understood by a non-technical audience (Brolcháin et al., 2017). It can also be seen as a process that consists of using graphs that make sense and weaving them into compelling, action-inspiring stories[1]. A well-known subfield of data storytelling in the literature is the data journalism, where journalists make use of large databases to produce stories (Gray et al., 2012; Kalatzi et al., 2018). In the context of open data, in addition to turning data into stories for data exploration or development of digital services, data storytelling is also about users using datasets published on open data portals for the following purposes: to better understand governments actions and to enable deeper and easier monitoring of government work (Brolcháin et al., 2017).

The data storytelling process is subdivided into 6 stages, as shown in Figure 7.8 (adapted from (Aanderud et al., 2020; Brolcháin et al., 2017)). The process begins by looking for answers to an identified need or question (**seeking answers** stage). Next, the user tries to identify and collect the datasets needed to answer the question (**data collection** stage), followed by the stage where the user processes the collected data, for example by grouping it or deleting certain rows or columns, to keep only the relevant information (**data processing** stage). Then, the user can create visualizations from the processed data to facilitate understanding (**data visualization** stage). Finally, the user can accompany the different visualizations with an interpretation or a small description and share them with other users to present their findings about the studied data (**story creation** stage). Once the story is shared, the user can receive feedback to improve the story or to engage in a discussion with other users (**feedback collection** stage).

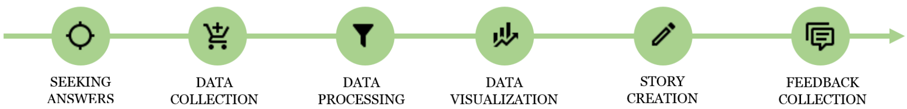


Figure 7.8: Data storytelling process.

7.4.1.2 Data Storytelling Tools

Many tools have been proposed in the literature to help users to turn data into stories in the context of open data (Ansari et al., 2022; Eberhardt and Silveira, 2018). Reviewing each of them would be beyond the scope of this study. We have focused here only on the popular and generic tools that have been used in the context of open data. We are aware of Linked Open Data (LOD) tools in the literature (Ansari et al., 2022; Böhm et al., 2012; Dadzie and Rowe, 2010; Eberhardt and Silveira, 2018) but we focused on data storytelling tools for open data in tabular format for two reasons. First, many users, especially those with low technical skills, are more familiar with tabular data than the RDF format which is used in LOD. Second, most of the datasets available on open data portals are published in plain text (CSV) (see e.g., data.gov (US), data.europa.eu (EU)). We grouped these tools into two categories: non-open-data oriented tools (e.g., Tableau, D3) and open-data oriented tools (e.g., ChartViz, SPOD). Non-open-data oriented tools are tools whose primary purpose was not designed for the open data context, but which can be used in that context. The second category includes tools that are initially designed to be used for open data. Table 7.13 presents the tools reviewed including a short description, their strengths (data storytelling features implemented), and weaknesses (data storytelling features not implemented).

In summary, although many tools have been proposed for data storytelling, they still lack some features to fully support users to process open data across the different stages of the data storytelling process (see Table 7.13). To address this gap, we proposed a generic and end-to-end tool called ODE, which will be described in the following sections. ODE differs from existing tools in that it addresses all of their mentioned shortcomings by providing a generic tool to turn any open dataset into stories, providing an end-to-end tool to cover all data storytelling stages, allowing a direct connection to open data, providing a quick overview of the data, providing an estimation of the data quality, allowing multiple datasets to be used in a single story, providing visualization recommendations on the entire data and an easy way to create embeddable and interactive visualization from scratch by simply selecting attributes, integrating up to 17 visualization types, and allowing users to customize existing stories even they are not the owners.

Table 7.13: Strengths and weaknesses of data storytelling tools.

Tools (including a short description)	Strengths (+) & Weaknesses (-)
Non-open-data oriented tools	
BI Tools (e.g., Tableau[15], Power BI[16], Qlik[17]) are mainly used by businesses and help them to understand trends and deriving insights from their data so that they can make tactical and strategic business decisions	(+) powerful features from data processing until story creation (+) handle millions of rows of data (-) no direct connection to datasets available on portals (-) no overview of the content of each data column (-) no evaluation of the quality of data and metadata (-) need for technical skills (e.g., SQL) before merging data (-) steep learning curve (-) no information about how a visualization was created and what data was used (-) no customization of others' stories (-) no direct feedback on the story
Web Tools (e.g., Datawrapper[18], Google Data Studio[19]) are mainly designed as web applications and their target users are mainly people with low technical skills	(+) no need to install software before using them (+) basic and easy to use features from data processing until story creation (+) no collaboration on story creation (-) same shortcomings as the BI tools, except that these tools are easier to use than the previous tools
The Gamma (Petricek 2017) provides a scripting language for working with data and produce reproducible source code making the analysis fully transparent and open	(+) same strengths as Web tools (+) information about how a visualization was created (+) customization of others' stories (-) no direct connection to datasets available on portals (-) no overview of the content of each data column (-) no evaluation of the quality of data and metadata (-) need for writing code to create visualizations or perform advanced functions such as merging datasets (-) no collaboration on story creation (-) no recommendation visualizations from selected data (-) no multiple visualizations in a single story (-) no filters on the story (-) no direct feedback on the story
Open-data oriented tools	

[15]https://www.tableau.com/
[16]https://powerbi.microsoft.com/
[17]https://www.qlik.com/
[18]https://www.datawrapper.de/
[19]https://datastudio.google.com/

Tools (including a short description)	**Strengths (+) & Weaknesses (-)**
Open data portals (e.g., data.gov (US), data.europa.eu (EU)) are mainly used by public organizations to publish their data to the public but also provide visualizations	(+) direct interaction with open datasets on the specific portal (+) easy-to-use features for data visualization (-) no interaction with open datasets from other portals (-) only the data visualization stage is supported
YourDataStories (YDS) (Brolcháin et al., 2017) is a European Horizon 2020 project which provides a list of features to integrate data storytelling into open data platforms	(+) predefined stories about public projects and contracts in European countries using maps and graphs (-) as of now, only focused on public projects and contracts in EU countries (-) need to register or log in before accessing the main features (-) no generic tool to allow any users to create their stories using open dataset on portals
OpenDataVis (Graves and Hendler, 2013) & (Graves and Bustos-Jiménez, 2015) are complementary and allow to interact with the data of any portals by providing a data URL. OpenDataVis provides an easy way to visualize data in less than 5 clicks. On the other hand, (Graves and Bustos-Jiménez, 2015) produces a series of visualizations describing the variables of the selected dataset	(+) interaction with a dataset from any portals by providing a data URL (+) provide an easy way to visualize data (+) information about how a visualization was created (+) customization of others' visualizations (+) overview of the data distribution of open dataset (-) no direct connection to open datasets from portals (users need to enter the data URL manually) (-) no use of multiple datasets in a single visualization (-) only handle data overview and data visualization
ChartViz (Pirozzi and Scarano, 2016) creates charts from a remote csv open dataset based on a decision tree algorithm	(+) interaction with a dataset from any portals by providing data URL (+) quantitative measure of data quality (homogeneity and completeness) of open dataset (+) provide an easy way to visualize data (-) same shortcoming as OpenDataVis (Graves and Bustos-Jiménez, 2015)
DEEP (De Donato et al., 2017) & SPOD (Cordasco et al., 2017) are complementary and enable the creation of interactive, reusable, and shareable visualizations	(+) direct connection to open data portals (+) creation of interactive, reusable and shareable visualizations (+) column completeness information (-) no estimation of data quality (-) no overview of the selected open dataset (-) no use of multiple datasets in a single story (-) no multiple visualizations in a single story (-) no recommendation visualizations from selected data

7.4.2 Research Methodology

The research questions of this study were addressed using the Design Research Science (DSR) methodology (Baskerville, 2008; Dresch et al., 2015; Hevner et al., 2004; Hivon and Titah, 2017; Peffers et al., 2007; Vaishnavi and Kuechler, 2007), as it aims to develop solutions (artefacts, design cycle) that meet defined objectives, contribute to the scientific knowledge base (rigor cycle), and provide utility in the environment (relevance cycle). The methodology of (Hevner et al., 2004) was adjusted in this study because we found that users were less proactive when we simply came to them to ask about their needs rather than presenting them with a prototype built on the basis of a literature review to obtain additional features and feedback.

Within the **rigor cycle**, we initially undertook a literature review from January to April 2022, following the methodology outlined by (Kitchenham, Barbara Brereton et al., 2009). Our goal was to access existing knowledge on the features needed to be integrated into a data storytelling tool to cover all its stages in the open data context and also on the barriers encountered by users. The literature review was conducted using the databases "Google Scholar" and "Science Direct" with the keywords ("open government data" OR "open data") AND ("technical feature" OR "technical barrier") AND ("re-use" OR "reuse" OR "use") AND ("visualization" OR "dashboard" OR "data storytelling"). Based on the automated search, we obtained 107 articles. We then identified relevant articles in three stages: first, we evaluated the type, domain and title; second, we examined the abstract; and finally, we scanned the content. In the end, we retained 13 articles that were relevant to our research. Features were then collected directly from the retained articles or inferred from the barriers identified in the articles. The literature review along with the feedback collected from users on the design and relevance cycles will be used to improve the current knowledge base. This is detailed in Section 7.4.4 by positioning the contributions to the current literature.

In the **design cycle**, we implemented the features gathered in the **rigor** cycle into a generic and end-to-end tool called Open Data Explorer (ODE). ODE is built using three technologies: Python as the programming language, Pandas as the data processing library, and Plotly to create and display the visualizations. An incremental approach based on the agile methodology (Fowler and Highsmith 2001) was used during the implementation of ODE. Once we implemented 2-3 features in ODE, we presented them to 2 users (one with high data manipulation skills and another with low data manipulation skills) to collect additional features, get their feedback and to improve the user interface.

In the **relevance cycle**, we evaluated the overall ease of use and usefulness of ODE as well as each of the implemented features collected during the **rigor** and **design** cycles. The evaluation was conducted through interviews followed by an online survey to assess the usability and usefulness of ODE and to gather additional features for future versions. Appendix F presents the questions contained in the survey. The survey was pretested with two users to ensure that all kinds of errors associated with survey research were reduced (Grimm, 2010). The survey included three types of questions: questions with a 7-point Likert scale (from "Strongly Disagree" to "Strongly Agree") based on the Technology Acceptance Model (TAM) (Davis, 1989; Moreno Cegarra et al., 2014) to evaluate two aspects: ease of use and usefulness; free-text questions to collect general opinions and suggestions for additional features for future versions and to justify previous ratings; and 3 additional questions to collect demographic data (level of data manipulation skills, age and education level). To recruit participants, a recruitment survey was sent through UNAMUR mailbox and social media and 11 participants (6 with high data manipulation skills and 5 with low data manipulation skills) were selected from that. Before completing the survey, participants were invited to test ODE with their preferred datasets on the Namur or Liege (Belgium) portals. Their tasks were to analyze their selected datasets and created stories from them. During the ODE test, we adopted an exploratory approach (Rubin and Chisnell, 2008), i.e., we let participants did what they considered to be the right action and guided them only when they felt confused and asked for our assistance. Once users completed the survey, they were asked four questions to get their overall opinion of the implemented features: what features should be **kept, improved, removed, or added** to facilitate data storytelling? After collecting user feedback, the median, mean, and standard deviation (SD) were calculated for the questions with a 7-point Likert scale. These statistical measures were chosen because they are the most appropriate for analyzing Likert data and for having a central tendency measure (Boone and Boone, 2012). Verbal thoughts and responses collected from the free-text questions were coded using short sentences to retain context and conceptual relations.

7.4.3 Results

In this section, we first presented the features identified from the literature review. Then, we presented how they were implemented. Finally, we reported on the results of the evaluation of ODE.

7.4.3.1 Features of a Data Storytelling Tool in the Context of Open Data

Table 7.14 presents the 15 features that we identified by conducting the above literature review. They were either proposed by previous studies or inferred from barriers identified therein. These features are grouped with respect to the different stages of the data storytelling process. No feature has been proposed for the **seeking answers** step, as this step is left to the users to decide if they just want to explore the data or if they already have a specific goal in mind.

Table 7.14: Features needed in the design of a data storytelling tool in the open data context

Data Collection (DC)
DC1. Direct access to open datasets from portals (Cordasco et al., 2017; De Donato et al., 2017; Graves and Bustos-Jiménez, 2015; Graves and Hendler, 2013; Pirozzi and Scarano, 2016)
Data Processing (DP)
DP1. Get a quick overview of the data content (Crusoe et al., 2019; Graves and Bustos-Jiménez, 2015)
DP2. Evaluate data quality (Brugger et al., 2016; Crusoe et al., 2019; Janssen et al., 2012; Pirozzi and Scarano, 2016; Zuiderwijk et al., 2012; Zuiderwijk and Janssen, 2014)
DP3. Filter useful data (Brolcháin et al., 2017; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)
DP4. Combine multiple data (Brolcháin et al., 2017; Crusoe et al., 2019; Graves and Hendler, 2013)
Data Visualization (DV)
DV1. Facilitate the creation of interactive visualization and provide instant visualizations (Brolcháin et al., 2017; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013; Pirozzi and Scarano, 2016)
DV2. Download or embed visualization (Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)

Table 7.14: Features needed in the design of a data storytelling tool in the open data context

Story Creation (SC)
SC1. Facilitate the creation of story that is easy to understand, use, and trust by potential users (Chokki et al., 2022d; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)
SC2. Share story (Brolcháin et al., 2017; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)
SC3. Embed story (Brolcháin et al., 2017; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)
SC4. Get information about the story (learning tool) (Chokki et al., 2022d; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013)
SC5. Customize story (Chokki et al., 2022d; Graves and Hendler, 2013)
Feedback Collection (FC)
FC1. Give feedback (Chokki et al., 2022d; Cordasco et al., 2017)
Other Features (OF)
OF1. Collaborate on story (Cordasco et al., 2017; De Donato et al., 2017)
OF2. Ease of use and shallow learning curve (Chokki et al., 2022c; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013; Pirozzi and Scarano, 2016)

7.4.3.2 Open Data Explorer (ODE)

ODE is a web application available at https://rb.gy/olmekk. A Video showing the steps performed by us to create a story from COVID19 hospitalizations in Belgium collected from the Namur (Belgium) portal[20] is available at https://rb.gy/cor6qt. The figures 7.9, 7.10, 7.11, 7.12, and 7.13 showcase several screenshots of the ODE tool. However, for a more comprehensive collection of screenshots, please visit https://rb.gy/atxowj.

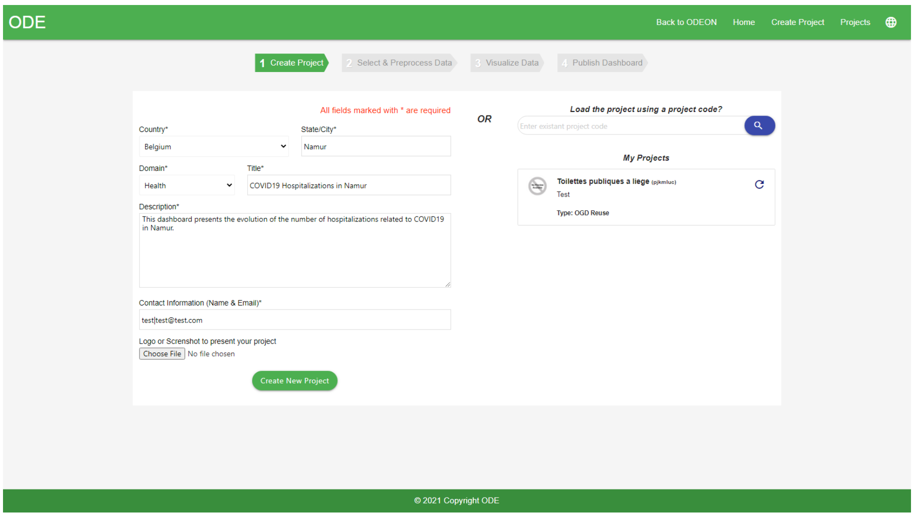


Figure 7.9: Create project.

Project Creation. Before moving to the data collection stage, ODE has implemented a workplace concept that helps users to collaborate on a single story **(OF1)**. Thus, before users begin collecting their data, they need to create a project by providing

[20]https://data.namur.be/

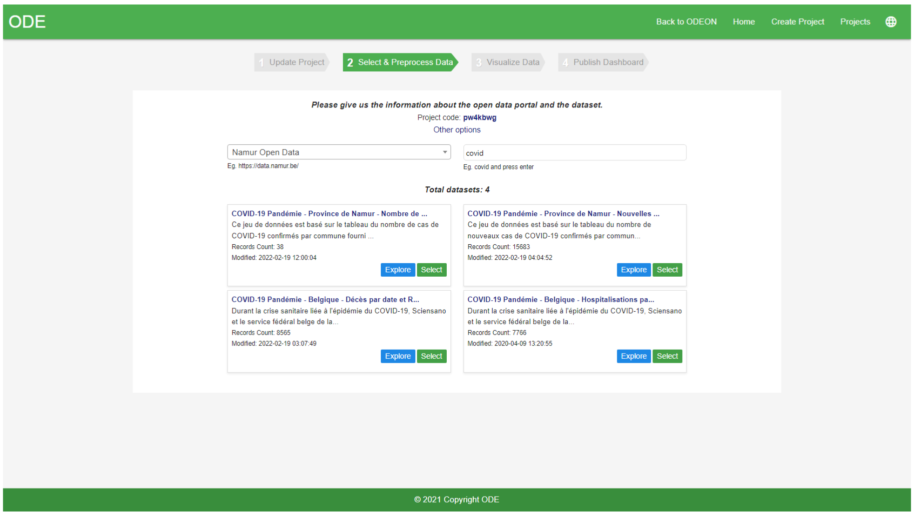


Figure 7.10: Select open dataset from Namur Portal.

information about its name, description, topic, country/state, and contact information (see Figure 7.9). After filling this information, the system then generates the project with a unique code that can be used to later modify the project or collaborate **(OF1)**.

Data Collection. In this stage, users can search for the data they want that is connected to any CKAN or OpenDataSoft portals **(DC1)**. ODE relies on the APIs provided by these two open data management systems (ODMS) to allow users to collect their data directly from a portal (see Figure 7.10). For now, we have integrated these two ODMS because they are among of the most widely used in the European countries (Berends et al. 2020).

Data Processing. In this stage, once users have selected a dataset, the system presents in table form the content of the selected data, a quick overview of each column of the dataset **(DP1)**, a data quality **(DP2)**, a correlation between the numerical columns and an auto-detection of the data type of each column with the possibility of adjusting the proposed data type. For the data type detection feature, the system relies on regular expressions and data type auto-detection in Pandas library. Currently, ODE covers four data types: numeric (integer, float), categorical (nominal, text, boolean), temporal (date), and geographic (latitude, longitude, geo point). For the quick overview of each column **(DP1)**, ODE proposes graphs with the information about the data type, the percent of missed values in the column and the list of columns that are correlated with the column in case that the column is numeric. ODE offers two types of graphs: a histogram for a numeric column and a word cloud for a categorical column. ODE incorporates four data quality metrics: completeness of cells, completeness of column labels, completeness of column descriptions and completeness of data information such as title or description or timeliness (last modified). The average of these four metrics is then calculated and corresponds to the data quality of the selected dataset. For the correlation feature, ODE shows a heatmap where each cell corresponds to the correlation between two numerical columns. After the users have the important information on the dataset, ODE suggests them four functionalities to process their data before using it **(DP3)**: drop columns, aggregate data, combine data **(DP4)** in case they have selected more than one dataset, search and replace some values.

Data Visualization. In this stage, ODE offers two features: create a visualization or get visualization recommendations **(DV1)** (see Figure 7.12). For the feature of

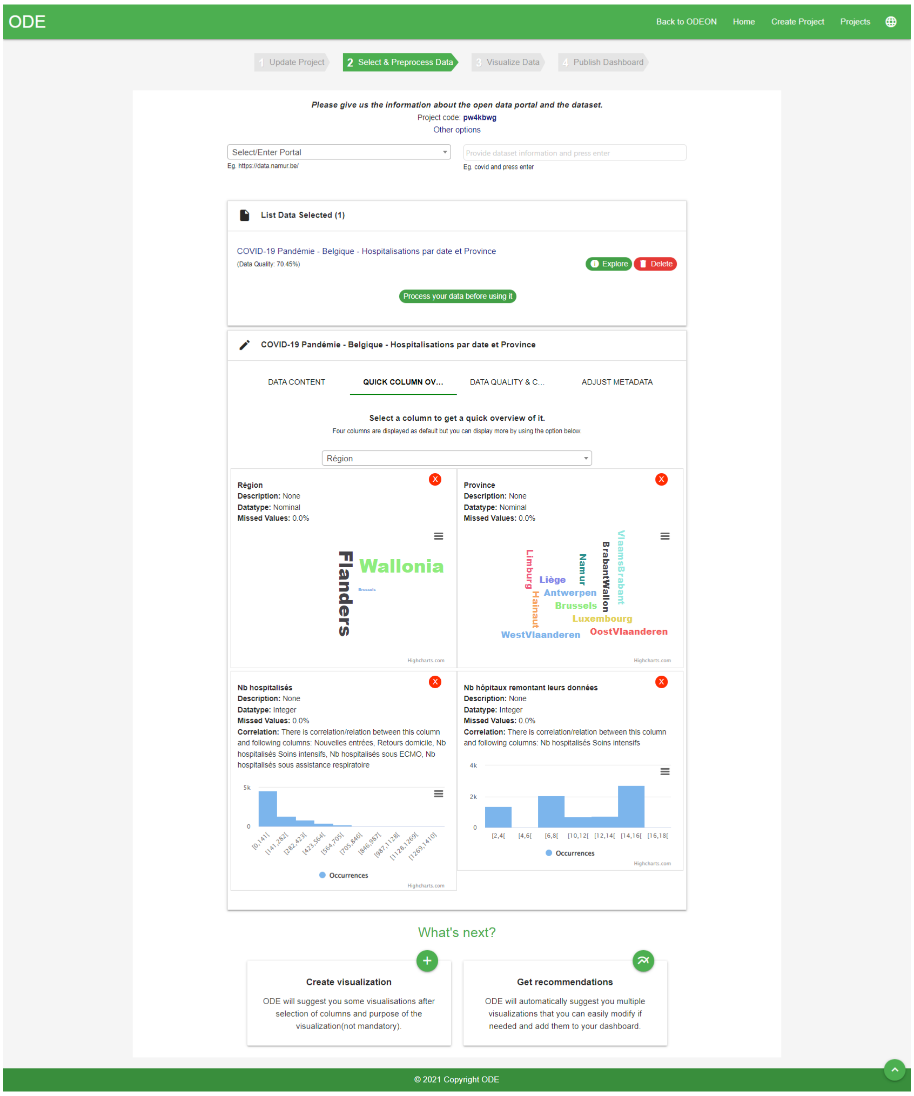


Figure 7.11: Present a data overview of the selected dataset.

visualization creation **(DV1)**, ODE offers a drag-and-drop option similar to Tableau for users with the difference that in ODE, users do not need to map each selected column to the shelves (rows, columns in the case of Tableau) themselves. They simply drag and drop their desired columns into a single shelf and ODE can automatically suggest multiple visualizations that best suits their selected columns based on the decision tree produced by this previous study (Holtz and Conor 2018) and additional rules provided by (Chokki et al. 2021; Munzner 2014; Wilke 2019) (more details in "Recommendation Module" section). For example, if users selected a nominal column and a numerical column with a maximum function applied on the numerical column, then ODE will generate visualizations using the following visualization types: bar chart, pie chart, doughnut, and treemap. Contrary to other tools, the decision tree in ODE can be also improved by integrating the user feedback on the suggested visualizations. For the visualization recommendations feature **(DV1)**, once users have selected their dataset, ODE can suggest several useful visualizations. ODE first detects significant columns. For numerical columns, significant ones are correlated with more numerical columns. For categorical columns, significant ones have few distinct values. Once significant columns are detected, we combine

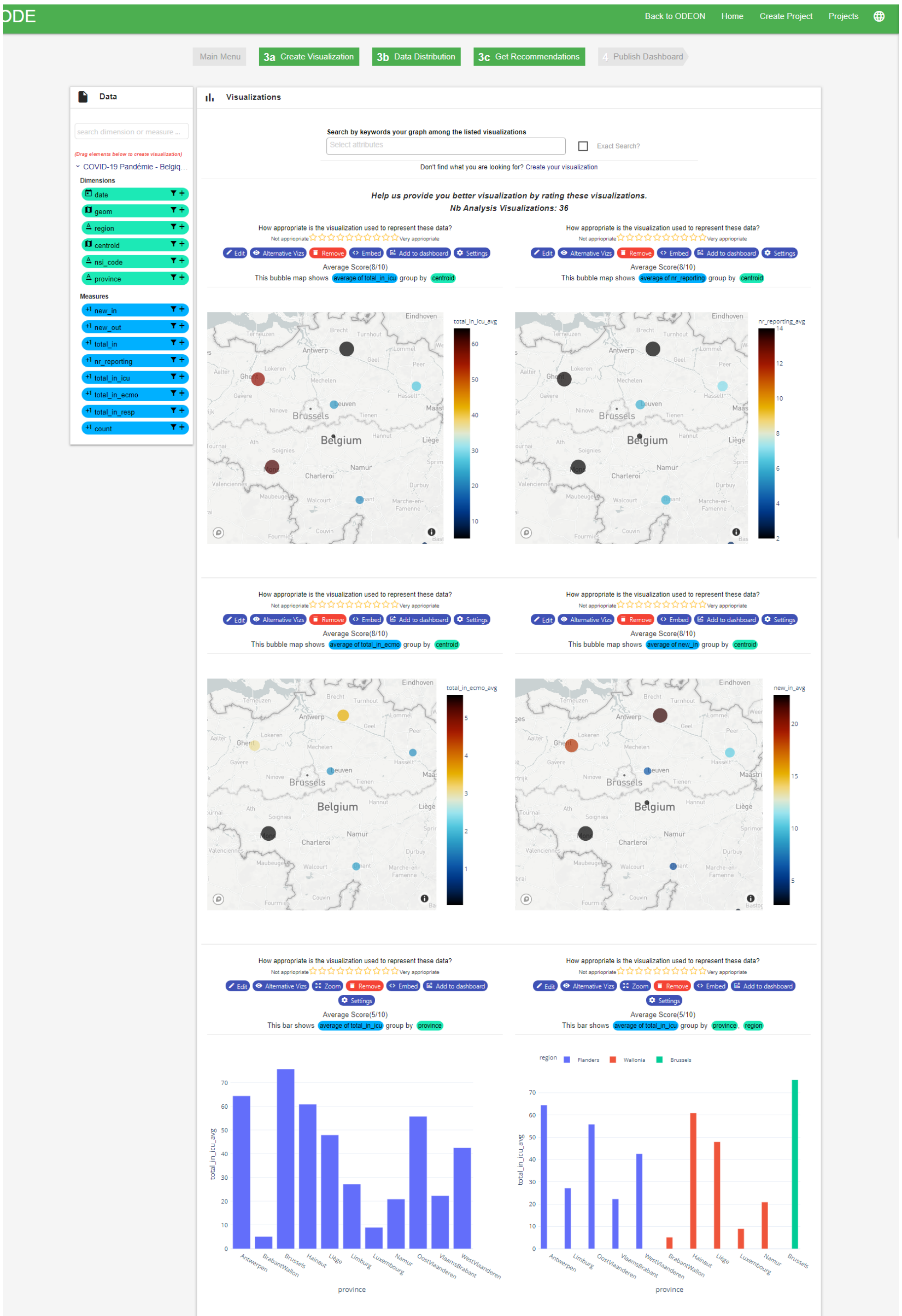


Figure 7.12: Suggest multiple visualizations based on the selected dataset.

one to two significant categorical columns with one to four numerical columns and apply an average transformation to the numerical columns. After obtaining these combinations, we used the decision tree to generate a visualization for each combination. Each generated visualization is interactive, has a button at the top that allows users to easily embed them into other web pages **(DV2)** and also has a rating option to collect user feedback on the visualization and use it to improve the decision tree. In addition to suggesting visualizations, we introduced a basic automated narrative system for every generated visualization, utilizing the template: "This {{graph_name}} displays {{measures}} categorized by {{dimensions}} where

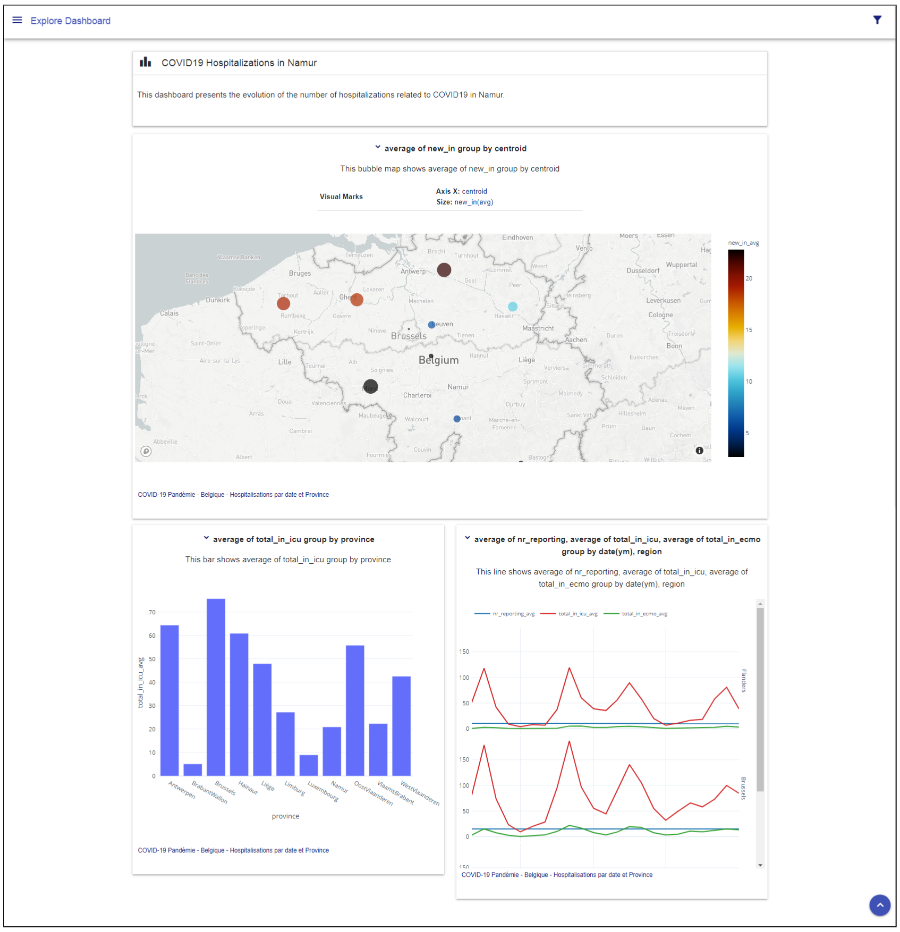


Figure 7.13: Dashboard generated.

{{filter}}". This structure, while straightforward, adjusts based on the context and the specific type of graph. For example, in the absence of a user-defined filter, the "where {{filter}}" part is excluded from the narrative. Once generated, this narrative can be edited by users, allowing them to add their unique insights and perspectives.

Recommendation Module. This module connects with the preceding stage, "Data Visualization," and it serves to automatically pinpoint the most suitable visualization marks correlating with the attributes users desire to display. In this study, we use "visualization marks" to describe three components: visualization type, design marks, and transformation. When discussing visualization types, ODE currently supports 17 distinct types: boxplot, violin, histogram, scatter, bubble plot, correlogram, bar, line plot, area Plot, connected scatter, bubble map, map, doughnut, pie, sunburst, treemap, and parallel coordinates. Moreover, it offers the ability for the ODE administrator to incorporate additional visualization types in the future. To deliver on the capability to generate visualizations based on specific visualization objectives, ODE relies on the principles laid out in previous studies (Munzner, 2014; Holtz and Conor, 2018). Each visualization type is associated with at least one of the following objectives: distribution, correlation, comparison, part to whole, trend over time, and geographic data observation. Regarding design marks, ODE currently suggests 12 (x, y, color, shape, size, label, orientation, grouping, facet row, facet column, dimensions, animation), enabling users to visually represent up to 5 attributes in dimensions and up to 5 attributes in measures. Table 7.15 illustrates some of ODE's design marks. Utilizing the guidelines defined in Table 7.15, such

Table 7.15: Examples of design marks in ODE

Design marks	**Description**	**Applicable to which data type**	**Applicable to which visualization type**
Color	Used to define the attribute to distinguish different categories in the plot	Categorical	all visualization types except sunburst, treemap parallel coordinates, correlogram
Size	Used to define the attribute to represent the size of bubble in bubble plot or bubble map	Numerical	bubble plot, bubble map
Facet row	Used to specify the attribute to put on the opposite of the axis Y	Categorical, Temporal	multiple graphs

as: "Applicable to which data" and "Applicable for which visualization type", we can assist users in identifying incongruities between attributes and visualization marks when crafting a visualization manually. In terms of transformations, ODE encompasses transformations like *Bin, Sum, Avg, Min, Max* for numerical data and *yyyy(year), yyyy-mm(year-month),yyyy-mm-dd(date), hh:ii (hour)* for temporal data. Categorical and geographic data types do not necessitate any transformation and thus, none will be applied.

The recommendation module is mainly based on the decision tree that is interpreted in ODE using the features and the feedback of different types of users: experts (users with high technical skills) and non-experts (users with low technical skills). The features are used here to determine the characteristics of data used to represent a specific type of visualization. Table 7.16 summarizes the list of features which are currently taken into account in ODE. Note that compared to the existing solutions, ODE saves the features in a database and it is therefore possible to add new features at any time without modifying the implementation. This allows our approach to be dynamically scalable. The recommendation module in ODE consists of the following steps:

- *Initialization:* This phase kickstarts the model with the visualization rules established in (Munzner, 2014; Holtz and Conor, 2018). Figures 7.14, 7.15, 7.16, and 7.17 showcase the revised version of the visualization rules identified in (Munzner, 2014; Holtz and Conor, 2018). Each rule is mapped to the pre-existing features detailed earlier (considered as inputs), and for each rule, we associate the visualization type, the score, and the visualization marks (considered as outputs). Table 7.17 gives some examples of inputs and outputs from various users for the pie chart which can be interpreted in the following way:
  - Various users assign a score between 5 and 9 to the pie chart when the maximum number of observations per dimensions (F1) equals 1, the number of numerical (F2) equals 1, the number of categorical (F3) equals 1, the data has no negative values (F10), we do not use “Avg” as a transformation (F9), and the number of elements to plot (F13) is less than 7.
  - One user assigns a score of 0 (not recommended) for a pie chart when the visualization employs “Avg” as a transformation (F9). This can be justified by the fact that using a pie chart to represent the average is generally discouraged in the literature.

- *Recommendation:* Post-initial rules, users can obtain recommended visualizations in four ways: by choosing the attributes they want to visualize, by selecting the visualization objectives, by specifying the design marks themselves, or by picking an existing visualization to get alternative visualizations. Depending on the option selected, the system automatically computes the value associated with each of the existing features, and then uses this value to find the list of rules defined by the users that match the computed features. After this step, the system groups the list of rules by visualization type and user type to calculate the average score of each visualization type. The scores are then utilized to find the visualization marks that best fit the user's request (refer Equation 7.1). The higher the score, the better the visualization.
- *Users feedback:* After ODE has provided some recommendation visualizations, users can then assign a score between 0 and 10 (10 meaning the visualization is the best and 0 the worst) to each visualization and submit it. For each rated visualization, ODE will generate the features linked to the visualization and associate the user type (by default it is non-expert if the user does not log in or we automatically retrieve from his profile if he is already registered in the system) and the score then add all this information in the list rules that will later enhance the knowledge base of our recommendation module. To control the impact of the feedback on the score calculation , we assign some weights to the scores of the different types of users (refer Equation 7.1).

The calculation of the score for each visualization type V_i is described by

$$\begin{gathered} S(V_i) = W_{ne} * AvgS_{ne}(V_i) + W_{ex} * AvgS_{ex}(V_i) \\ W_{ne} + W_{ex} = 1 \\ V_{opt} = Max(S(V_i)) \\ V_{rec} = List(V_i)\ where\ S(V_i) \geq S_{rec} \end{gathered} \tag{7.1}$$

where

- V_i means the ith visualization type;
- W_{ne} (default value: 0.2), W_{ex} (default value: 0.8) mean respectively the weights assigned to the scores of non-experts and expert users. The values of W_{ex} and W_{ne} are configurable in the system. The default values of these parameters are automatically used to get the optimal visualization and recommendations but can also be changed by the end users;
- Avg$S_{ne}(V_i)$, Avg$S_{ex}(V_i)$ mean respectively the average of the scores given by non-experts, expert users for the visualization type V_i;
- V_{opt} means the optimal visualization, -V_{rec} regroups the list of recommended visualizations, -S_{rec} means the minimum score used to decide which visualization type can be suggested to the users.

Story Creation. In this stage, whenever users create the visualization of their choice, they can add it to a story by simply clicking to a button "Add to dashboard" situated on the top of the visualization **(SC1)**. After adding all their desired visualizations to their story, they can then configure some dashboard settings (optional as default settings are already defined) **(OF2)** such as adding filters for the dashboard to allow users to interact with the dashboard, adding the title, description and width of each visualization. Once these parameters are defined, they can save them and the dashboard is generated. The 16 dashboard design principles (eg., use best visualization practices, use the right type of chart, integrate feedback support, allow customization) summarized in (Chokki et al., 2022d) were incorporated into ODE to ensure that the dashboard generated by ODE follows the best practices and thus is easy to use and understandable by the end users. They also help to propose a presentable design of the dashboard without the need for users to do many settings before having an attractive and interactive dashboard **(SC1, OF2)**. In the generated story (see Figure 7.13), users have the following options: view dashboard information (e.g., see open data used in the story, see previous user comments), share dashboard (share the story with others by mail) **(SC2)**, embed dashboard (integrate the dashboard into other web pages), give feedback (submit comments on the story) **(SC3)**, customize dashboard (ODE will automatically duplicate the current story into another story and let users edit it as they wish) **(SC5)**, visualization details (e.g., title, description, visualization marks, and open data used to create the visualization) **(SC4)**.

Feedback Collection. In this stage, users can view or give the feedback on any stories through two other applications: ODEON (Chokki et al., 2022c) and CitizenApps (Chokki et al., 2022e) which are linked to ODE (FC1). Once users generate their story using ODE, their story is automatically published on ODEON and CitizenApps which are platforms that allow users to get feedback of their submitted projects. So, instead of implementing the feedback option in ODE, we simply used the application APIs to directly post the stories published on ODE.

Table 7.16: List of features in ODE

No	**Feature**	**Remarks**
F1	Maximum number of observations per dimensions	This feature helps to decide if the task to visualize is for exploration (distribution or relation) or for data analysis (Comparing, change over time, etc)
F2	Number of numerical data type	This feature helps to determine the number of attributes of numerical type
F3	Number of categorical data type	This feature helps to determine the number of attributes of categorical type
F4	Number of temporal data type	This feature helps to determine the number of attributes of temporal type
F5	Number of geographical data type	This feature helps to determine the number of attributes of geographical type
F6	Number of dimensions	This feature helps to determine the number of attribues used as dimensions
F7	Number of measures	This feature helps to determine the number of attribues used as measures
F8	Number of elements to group per category	This feature helps to decide about the readability of the graph by the human' eyes and whether it is better to use small graphs rather than grouped/stacked graphs. Based on (Munzner, 2014; Holtz and Conor, 2018), for example it is better to use small graphs when F8 ≤ 5
F9	Use Average as transformation?	This feature is specially designed for pie chart, doughnut, treemap, sunburst and helps to avoid using the mentioned charts when the transformation of the measure to plot is average
F10	Number of negative values per measure is equal 0	This feature is specially designed for pie chart, doughnut, treemap, sunburst and helps to avoid using the mentioned charts when there are negative values to plot
F11	Number of distinct values per measure is equal to number of observations per measure	This feature is specially designed for pie chart, doughnut. Based on (Munzner, 2014; Holtz and Conor, 2018), when some observations have the same values, it is difficult for the human' eyes to distinguish them in the mentioned charts so another chart should be more appropriate in this case
F12	Absolute value of correlation between 2 numerical attributes	This feature is specially designed for distribution plots such as scatter plot, bubble plot, boxplot and violin and helps to see if there is a relation between two numerical attributes
F13	Number of elements in measures to plot	This feature helps to decide about the readability of the graph by the human' eyes. Based on (Munzner, 2014; Holtz and Conor, 2018), for example, in pie chart it is better to use it when F13 (number of angles to represent) ≤ 7 or in bar chart when F13 (number of bars to represent) ≤ 20

Table 7.17: Example of rules defined in ODE. (Cat = categorical, Num = Numerical)

Inputs													**Outputs**				
F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**	**F12**	**F13**	**Viz Type**	**Score**	**Viz Marks**	**User**	**User Type**
Initialization: Rules from literature																	
=1	=1	=1	=0	=0	=1	=1	1	False	True	False		≤7	Pie	5	x:Cat y:Num	Admin	Expert
								True					Pie	0		Admin	Expert
Users' feedback																	
=1	=1	=1	=0	=0	=1	=1	1	False	True	False		≤7	Pie	9	x:Cat y:Num	A	Non-expert
=1	=1	=1	=0	=0	=1	=1	1	False	True	False		≤7	Pie	7	x:Cat y:Num	B	Non-expert
=1	=1	=1	=0	=0	=1	=1	1	False	True	False		≤7	Pie	6	x:Cat y:Num	C	Expert

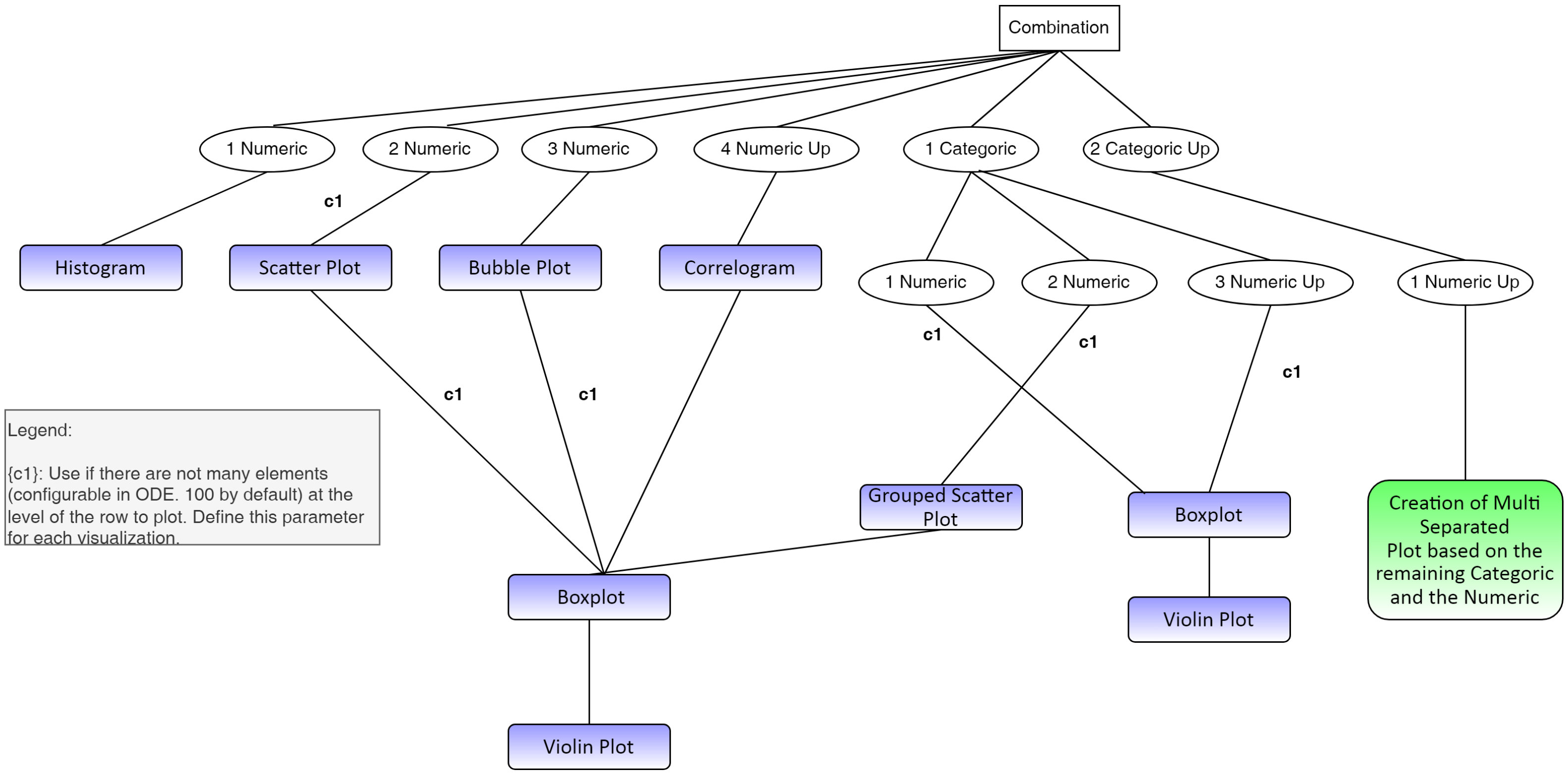


Figure 7.14: Decision tree to find the best visualization for many observations per row.

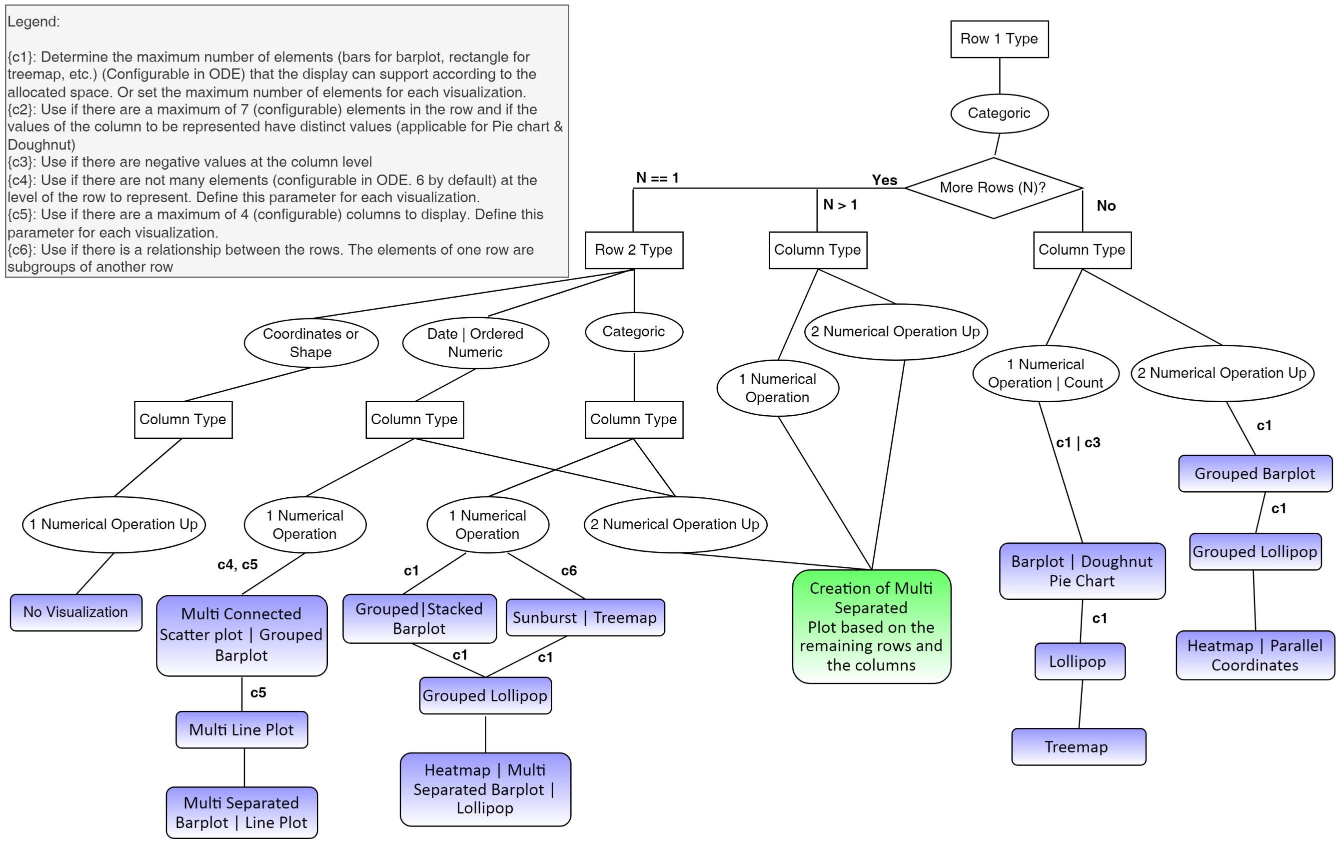


Figure 7.15: Decision tree to find the best visualization for one observation per row: Categoric case.

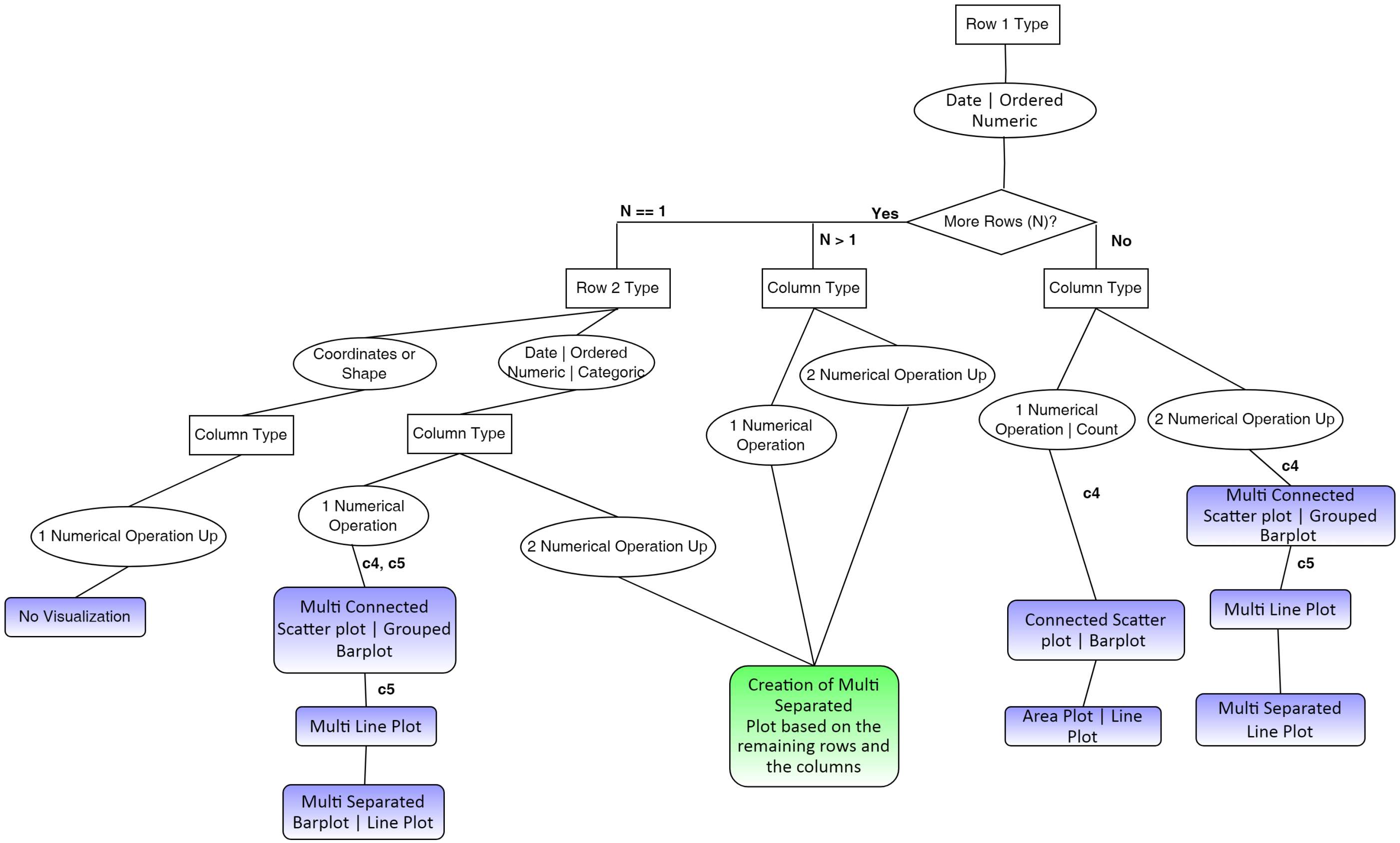


Figure 7.16: Decision tree to find the best visualization for one observation per row: Date or Ordered Numeric case.

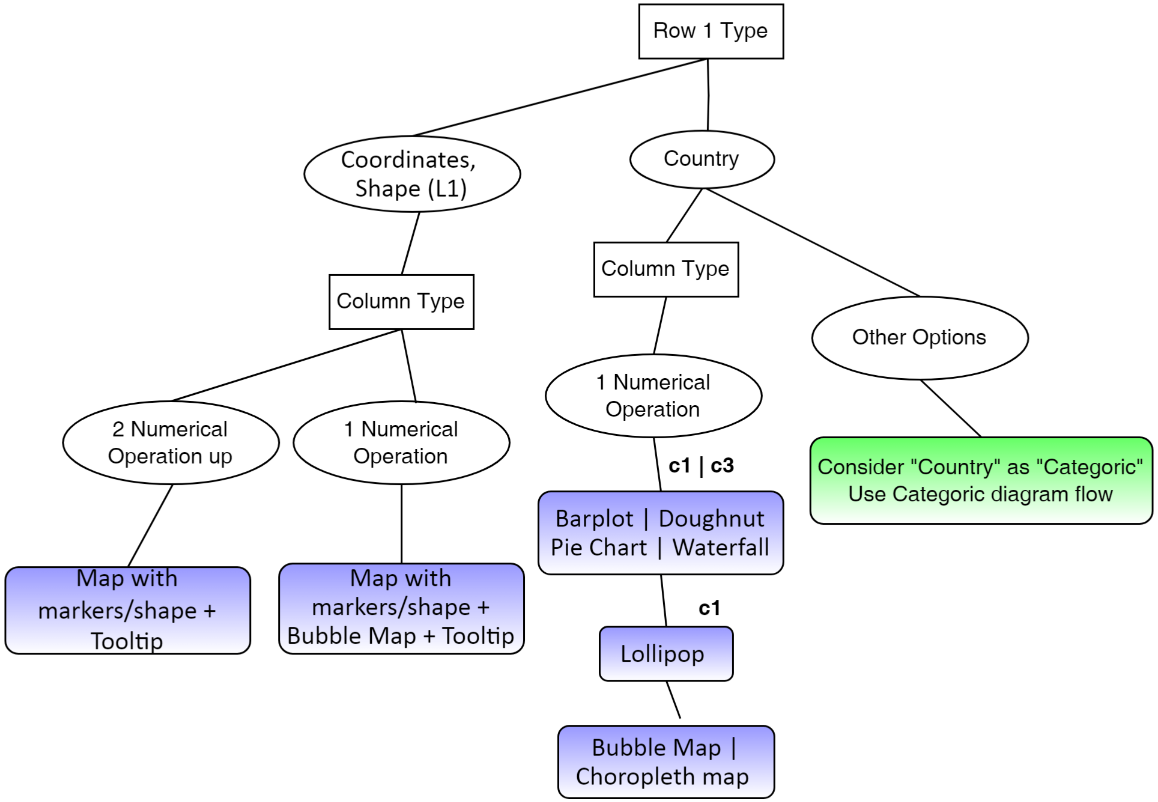


Figure 7.17: Decision tree to find the best visualization for one observation per row: others cases.

7.4.3.3 An Extension of ODE for Publishers: StoryOGD Prototype

The previous section introduced the ODE prototype, which assists users in creating a data dashboard. In this section, we take it a step further by introducing an extended version of ODE called StoryOGD, specifically designed to help publishers present their portals in a user-friendly manner.

To accomplish this objective, we conducted an exploratory study of websites (accessible at https://rb.gy/oqeipl and https://rb.gy/gglgl7) that aggregate open datasets and present them in a user-friendly way.

Through the exploratory study of the aforementioned websites, we successfully identified the supplementary requirements essential for extending ODE to assist publishers in transforming their portals into storytelling portals (see Table 7.18).

Table 7.18: Additional requirements for enhancing ODE to support publishers in converting their portals into storytelling portals.

(R1) Data aggregation by topic
(R2) Creation of topic-specific dashboards
(R3) Inclusion of filtering options for each dashboard
(R4) Provision for sharing or embedding the dashboard
(R5) Presentation of information (e.g., data sources) on each dashboard
(R6) Incorporation of a feedback mechanism to enhance the dashboard

These requirements, derived from the exploratory study, were integrated into StoryOGD, accessible at http://79.143.180.14:7001/home-trans. StoryOGD consists of a backend and frontend developed using Django. The backend matches the dashboards created by the ODE data storytelling tool (see Figure 7.18) with the relevant topics (see Figure 7.19), while the frontend displays a list of topics (see Figure 7.20) and their respective dashboards, offering various options such as filters, sharing, embedding, and obtaining information (see Figure 7.21).

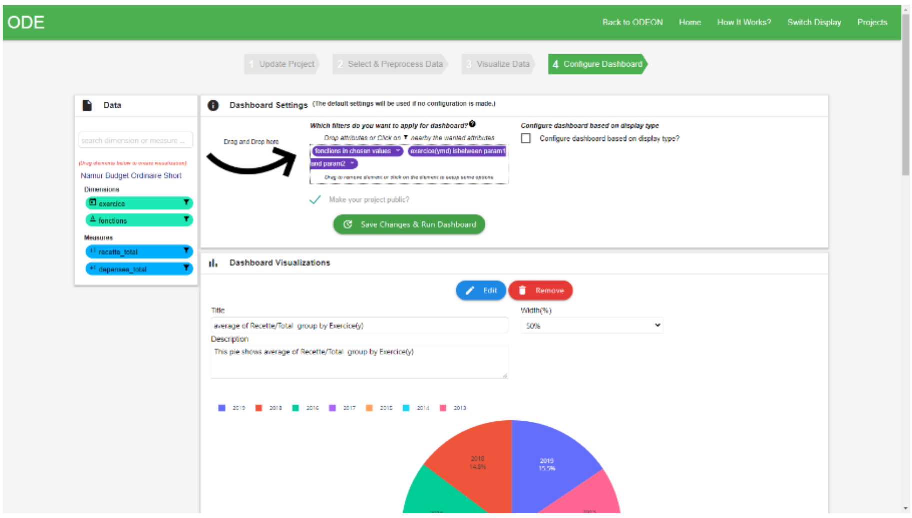


Figure 7.18: Step 1 - Creation of the Namur budget dashboard using ODE (Backend).

7.4.3.4 Evaluation Results and Analysis

Through the surveys that participants completed after exploring the ODE prototype, we were able to collect their opinions related about the ease of use and usefulness

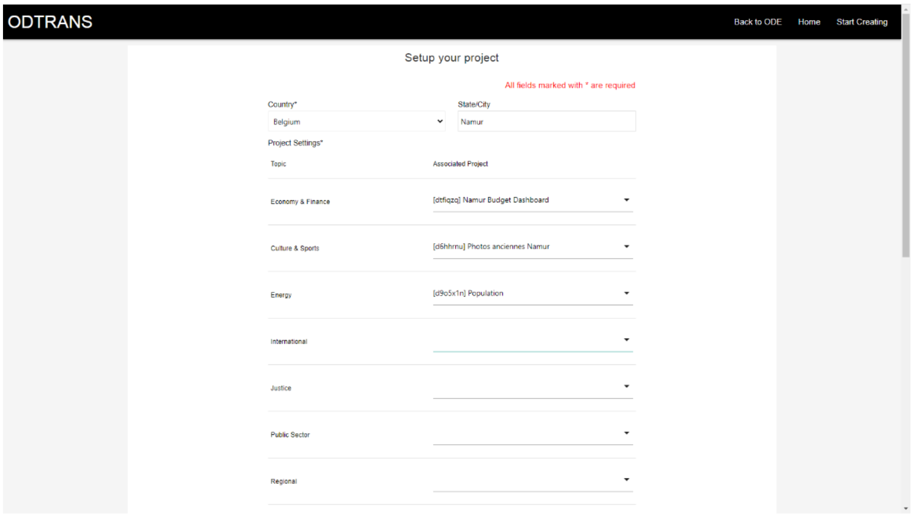


Figure 7.19: Step 2 - Setup of the storytelling portal. Link created dashboard to "Economy & Finance" topic (Backend).

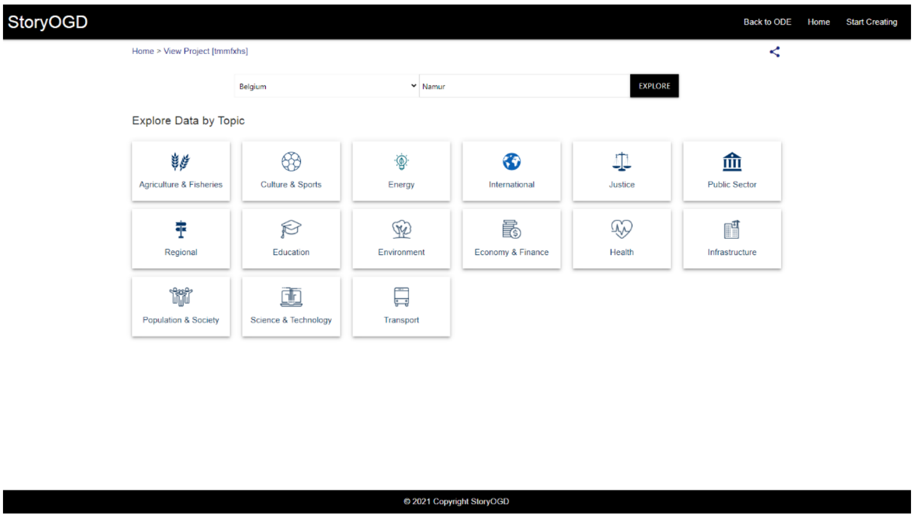


Figure 7.20: Step 3 - Main page of the storytelling portal (Frontend).

of ODE and also of each feature implemented in ODE to turn data into stories. A total of 11 participants (5 with low data manipulation skills and 6 with high data manipulation skills) completed the surveys. All participants are between the ages of 18 and 50 and have at least a high school degree. Each participant's evaluation section lasted a maximum of 1 hour.

Table 7.19 presents the median, mean and standard deviation of the questions with a 7-point Likert scale regarding the 2 aspects (perceived ease of use and perceived usefulness) evaluated for the prototype. The following conclusions can be drawn from the results of Table 7.19. First, most of the participants with low data manipulation skills agree that the proposed prototype is easy to use, as evidenced by the median and mean of perceived ease of use ≥ 5 and the low standard deviation (SD = 0.93) showing that there is no high significant difference between users' scores. They also agree that the prototype is useful to them to better explore open data and turn them into stories (median & mean ≥ 5 for perceived usefulness and there is no high significant difference between users' scores (SD = 1.02)). Second, participants

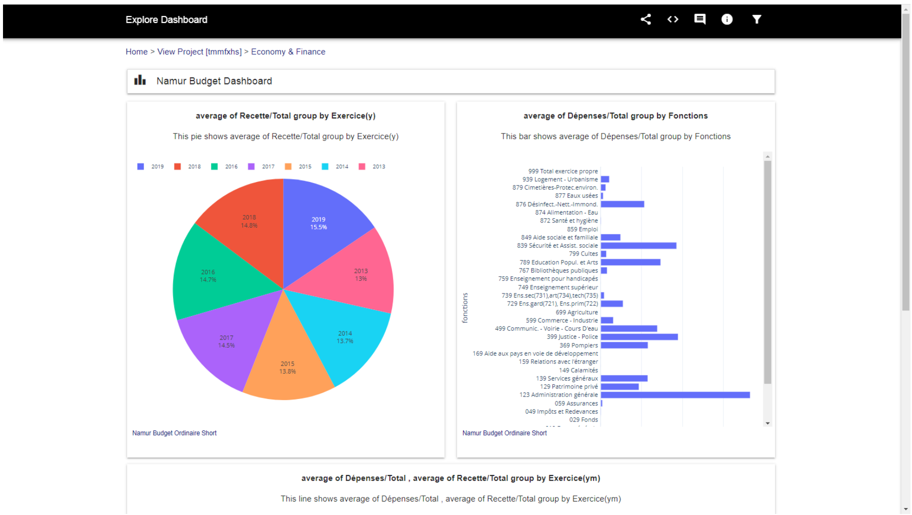


Figure 7.21: Step 4 - Displaying the dashboard after the user clicks on “Economy & Finance” topic (Frontend).

with high data manipulation skills find it easier to use the prototype than the participants with low data manipulation skills as evidenced by the median equaling 6. They also agree that the prototype is useful to them in better exploring open data and turning them into stories (median & mean ≥ 6 for perceived usefulness and there is no high significant difference between citizens' scores (SD = 1.02)).

Table 7.19: Median, mean and standard deviation (SD) of survey scores

	Participants with low data manipulation skills		**Participants with high data manipulation skills**	
	Median	**Mean (SD)**	**Median**	**Mean (SD)**
Perceived ease of use	5	5.4 (0.93)	6	5.39 (1.10)
Perceived usefulness	6	5.87 (1.02)	6.5 (1.10)	6.1 (1.16)

These observations can be justified as follows. First, many participants, even those with high data manipulation skills, found the prototype's interface not too intuitive, but they all agree that they could quickly become proficient if they had more time or if we had done a tutorial at the beginning of the evaluation to show the basic features. Second, all participants were able to create a dashboard from their selected dataset during the evaluation section.

Regarding the ease of use and usefulness of implemented features, all participants found that the implemented features presented in 7.14 are useful in turning their selected dataset into stories, but made the following suggestions. Users with low data manipulation skills suggested that the features related to data processing (**DP1.** Get overview to **DP4.** Combine datasets) to be removed because they are difficult for them to understand. On the other hand, users with high data manipulation skills think that these features are necessary. All of them also found the functionality of recommending visualizations from the selected dataset (**DV1**) useful, as many of them were able to find the visualizations they wanted to create their dashboards from this feature. In addition, they suggested making the interface design more intuitive. However, participants did not suggest any additional features, as many felt that existing features were sufficient and that it was best to avoid making the application more difficult to use.

7.4.4 Conclusion

The goal of this study was to identify a list of features needed in the design of a separate and generic data storytelling tool. To achieve this goal, we first conducted a literature review and discussed with potential users to gather features needed in the design of a data storytelling tool in the open data context. Next, we implemented the ODE prototype based on the collected features, and then examined through an evaluation conducted with 11 participants, whether the prototype and each of its features were easy to use and useful in helping users with low or high data manipulation skills to turn their data into stories.

This study contributes to theory in several significant aspects. Firstly, it builds upon prior studies and tools (Brolcháin et al., 2017; Cordasco et al., 2017; De Donato et al., 2017; Graves and Hendler, 2013; Pirozzi and Scarano, 2016) to propose a comprehensive list of 15 essential features required for the design of a data storytelling tool in the context of open data. These features are detailed in Table 7.14 and are discussed in Section 7.4.3.2. Additionally, an extended version of ODE (StoryOGD) led to the identification of six requirements for developing a tool that assists publishers in transforming conventional portals into storytelling portals and are discussed in Section 7.4.3.3. This extension was based on an exploratory study of two existing storytelling portals. Secondly, unlike the YDS project (Brolcháin et al., 2017), which primarily focused on four aspects (discovery, assistance, insight, and leverage) to enhance open data platforms with data storytelling features, our research concentrates on the technical features necessary for a generic tool to facilitate data storytelling with open data. Moreover, we categorized these features based on different stages of data storytelling. This categorization proves valuable for designers and developers, as it helps them identify which features to implement at specific stages of data storytelling, thus catering to users' needs more effectively. Lastly, the evaluation results demonstrate the usefulness of all 15 implemented features in ODE, except for users with limited data manipulation skills, who suggested removing the features related to data processing (**DP1.** Get overview to **DP4.** Combine datasets). Furthermore, the prototype proved highly valuable in transforming open data into engaging stories, as most participants agreed that it met their expectations and enabled them to create their stories using their selected data.

This study also makes significant practical contributions in the following aspects. Firstly, unlike the previous tools presented in Section Background (refer to Table 7.13), ODE addresses the shortcomings of each of these tools, providing users with an end-to-end tool to seamlessly turn their data into compelling stories. Users no longer need to rely on separate tools for data processing or getting an overview of data content (see Section Open Data Explorer (ODE)). Secondly, in contrast to other tools that offer static rules for generating visualizations, ODE allows users to provide feedback on the visualizations, which is then leveraged to improve the initial visualization rules. Thirdly, through StoryOGD, developers (i.e., infomediairies) and especially publishers can easily convert existing portals into captivating storytelling portals, making it easier for them to engage their audiences with open data. Lastly, we provide access to the source codes of ODE and StoryOGD. This serves as a valuable starting point for developers looking to create their own tools for facilitating open data storytelling or seeking to enhance the existing prototypes.

An important limitation of this study pertains to the representativeness of the participants in the evaluation. While the number of participants may appear small, it aligns with established guidelines from previous studies (Faulkner, 2003; Nielsen, 2000), which recommend using at least 5 participants for usability tests as a suitable baseline. To enhance the representativeness of future evaluations, we suggest exploring alternative communication channels or conducting on-site data collection at universities or public spaces. Unfortunately, such endeavors were not feasible during the research period due to the prevailing COVID-19 situation. For future work, it is imperative to evaluate StoryOGD with a broader range of citizens to gauge how the use of storytelling portals contributes to enhancing citizen engagement with OGD portals. While the current research places a heavier emphasis on the visualization aspect, an intriguing avenue for future exploration would be to enhance the automated narrative generation process in tandem with the visual representation. This would offer users a richer and more nuanced understanding of the data being presented. Additionally, conducting an evaluation of StoryOGD with end-users, specifically publishers, is essential to assess the prototype's ease of use and overall usefulness, validate the identified requirements, and gather any additional requirements that may arise. Such assessments will further enrich the effectiveness and practicality of the tool.

7.5 Wrap up

In this chapter, we investigated various aspects related to facilitating data storytelling with Open Government Data. We made significant contributions in three sections, each addressing a specific aspect of the overall objective.

In the first section, we provided valuable insights into the datasets and visualizations preferred by users on OGD portals. These findings are instrumental for infomediaries seeking appropriate datasets for their applications and for OGD publishers aiming to determine the datasets to make available on their portals. Moreover, they offer guidance on the types of visualizations that enhance user comprehension. Through a multi-method approach encompassing the analysis of usage statistics, an online survey, and interviews, we obtained these insightful findings.

Moving on to the second section, we compiled a comprehensive set of design principles for dashboards in the OGD context. These principles were identified through a systematic literature review. By implementing these principles in a usable tool (NBDash) and evaluating their impact on citizen engagement, we validated the hypothesis that well-designed dashboards can effectively promote citizen engagement with OGD. This validation was achieved by utilizing the tool as a proxy for assessing citizen engagement.

Lastly, in the third section, we introduced essential features necessary for the design of a generic and end-to-end data storytelling tool. These features were identified through a combination of thorough literature review and interviews with potential users. To validate these features, we employed the ODE tool as a proxy, integrating the identified features. Additionally, we introduced an enhanced version of the ODE tool named StoryOGD, which specifically caters to the needs of publishers in presenting their portals in a user-friendly manner. The requirements for implementing such a tool were identified through an exploratory study of websites that successfully presented open datasets in a user-friendly manner.

Collectively, this chapter provides a comprehensive understanding of users' preferences for datasets and visualizations on OGD portals, design principles for dashboards to enhance citizen engagement, and features required for a generic and end-to-end data storytelling tool. These findings have significant implications for infomediaries, publishers, and citizens alike, ultimately facilitating the seamless transformation of OGD into compelling narratives.

Part IV

Closing Comments

CHAPTER 8

DISCUSSION AND FUTURE DIRECTIONS

In this chapter, we provide a concise overview of the contributions made in this thesis. We then examine their implications for both research and practice. Following that, we acknowledge the limitations under which the research in this thesis was conducted and suggest potential avenues for future research.

8.1 Contributions

This thesis encompasses various contributions, categorized into theoretical and practical aspects. Within the realm of theoretical contributions, a total of eleven significant findings are presented, outlined as follows.

Communication Methods for OGD Awareness: This thesis compiles an **inventory of communication methods** (C1) that are used to promote OGD to users. Subsequently, it conducts an online survey to evaluate these communication methods and derives **recommendations regarding users' preferred communication methods** (C2).

Requirements for OGD Awareness Applications: Following the preference of users to be informed of OGD from applications, this thesis identifies and validates with users **11 requirements for designing an application to improve OGD awareness** (C3).

Requirements for Communication and Collaboration Applications between OGD Stakeholders: Building upon previous impediment studies, this thesis proposes and validates a list of **16 requirements necessary to facilitate communication and collaboration between OGD stakeholders** (C4).

Meta(data) Quality Assessment: This thesis presents a comprehensive **list of dimensions and metrics for evaluating the quality of metadata and data content separately** (C5). Additionally, it identifies a set of **features required to develop a (meta)data quality assessment tool** (C6) through existing tools.

Data and Visualization Expectations for OGD: This thesis combines usage statistics and a survey to provide **recommendations about users' data expectations** (C7). Furthermore, it derives **recommendations about visualizations expectations** (C8) for different data categories through interviews with users.

Dashboard Design Principles for OGD: This thesis presents **16 design principles for dashboards in the OGD context** (C9), tailored to fulfill the conditions for OGD user engagement.

Requirements for End-to-End Data Storytelling Tool: The research offers a comprehensive list of **15 essential features required for designing a data storytelling**

tool (C10) in the context of open government data. Additionally, it identifies **six requirements for integrating the tool and converting conventional portals into storytelling portals** (C11).

In terms of practical contributions, this thesis presents six key contributions in the form of six tools. These tools were developed by implementing the requirements identified during various phases of the thesis, resulting in usable prototypes that served as proxies for validating the identified requirements. The **CitizenApps** prototype (T1) incorporates the requirements identified in theoretical contribution (C1) and, in turn, serves as a proxy for validating these requirements. Similarly, the **ODEON** prototype (T2) integrates the requirements identified in theoretical contribution (C4) and is utilized as a proxy for their validation. Additionally, the **QualityOGD** prototype (T3) encompasses the requirements identified in theoretical contribution (6) and acts as a proxy for validating these requirements. On the other hand, the **NBDash** (T4) prototype embodies the dashboard principles identified in theoretical contribution (C9) and acts as a proxy for validating these requirements. Furthermore, the **ODE** prototype (T5) integrates the requirements identified in theoretical contribution (10) and serves as a proxy for their validation. Lastly, the **StoryOGD** prototype (T6) combines the requirements identified in theoretical contribution (C11) with the ODE prototype, thereby extending its functionality to allow the conversion of conventional portals into storytelling portals.

8.2 Implications for Research

Concerning the potential implications for research, our diverse research activities enabled us to address the defined research questions effectively. A comprehensive mapping between the contributions and the research questions can be found in Table 8.1, outlining their interrelation.

Research Question 1 ("How can communication and collaboration among OGD stakeholders involved in data reuse be facilitated?") was addressed through several steps. Initially, a literature review was conducted to identify methods used to raise citizen awareness of OGD. Following that, a survey was conducted online involving 30 users, and the obtained results were compared with users' perceptions. This analysis allowed for the identification of the most favored communication methods for OGD awareness. Consequently, recommendations for effective communication methods were derived, addressing RQ1a. To delve deeper into the use of applications which is one of users' preferred communication methods for OGD awareness, interviews with 10 users were conducted allowing to gather requirements for such an application. These requirements were then implemented in the CitizenApps application, and its effectiveness in raising OGD awareness was evaluated with 25 users, answering RQ1b. The thesis also focused on the technical aspects of the lack of communication and collaboration between OGD stakeholders. A literature review identified impediments, and interviews with 9 stakeholders validated the requirements for a usable tool to address these issues. The ODEON prototype, incorporating these requirements, was evaluated with 22 stakeholders from different profiles, assessing its usability and usefulness in facilitating communication and collaboration, thus addressing RQ1c.

Table 8.1: Mapping between Research Questions and Contributions.

RQ	Theoretical contributions											Practical contributions					
	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
RQ1a	X	X															
RQ1b			X									X					
RQ1c				X									X				
RQ2					X	X								X			
RQ3a							X										
RQ3b								X									
RQ3c									X								
RQ3d									X						X		
RQ3e										X	X					X	X

Regarding Research Question 2 ("How to support automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user needs?"), we conducted a thorough literature review. From this review, we identified a comprehensive list of 14 dimensions and associated metrics to assess metadata and data content quality independently. Additionally, we pinpointed 7 features that could be incorporated into a quality assessment tool. These findings were then integrated into the QualityOGD prototype and validated with users, successfully addressing RQ2.

Research Question 3 ("How to facilitate the transformation of OGD into stories?") was approached in several stages. Firstly, we analyzed the OGD portal of Namur's usage statistics and conducted a complementary online survey with 43 participants to understand users' dataset needs. Additionally, we interviewed 10 users to gather opinions on well-designed visualizations of datasets, successfully identifying users' content and visualization expectations, addressing RQ3a and RQ3b respectively. Next, we conducted a comparison between individual visualizations and dashboards to investigate their influence on user engagement. This comparison prompted us to conduct a systematic literature review on dashboard principles, culminating in the formulation of 16 design principles tailored to OGD dashboards. Consequently, we proceeded to develop the Namur Budget Dashboard (NBDash) and evaluated its impact in comparison to individual visualizations. Our findings indicated that users showed a preference for using dashboards to engage with OGD, effectively addressing RQ3d. Lastly, our focus is on creating a comprehensive end-to-end tool that helps users transform data into compelling stories. We conducted a literature review and engaged with potential users to gather 15 essential features for a data storytelling tool in the open government data context. Implementing the ODE prototype and evaluating it with 11 participants addressed a portion of RQ3e. To fully tackle RQ3e, we identified 6 additional requirements through an exploratory study and integrated them into StoryOGD prototype. The StoryOGD prototype allows multiple dashboards to be grouped in a single interface, converting conventional portals' data into storytelling portals, thus fully addressing RQ3e for all OGD stakeholders considered in this thesis.

In addition to the specific contributions presented in this thesis, all the data collection instruments used throughout the research, including questionnaires, literature review protocol, and user testing procedure, are provided in full in the Part V for the sake of transparency and to aid replication. This valuable resource can serve as a foundation for other researchers embarking on similar investigations or seeking to validate and extend the study's findings. Moreover, the various requirements proposed and validated in this thesis for different purposes can serve as a robust framework for evaluating the suitability of specific tools. Researchers and practitioners in the field of open government data governance can use these requirements to assess whether a particular tool aligns with their intended purpose, ensuring its effectiveness in meeting specific needs. Additionally, these requirements offer a valuable means of comparison between different tools that cater to the same objectives. By benchmarking tools against the established requirements, stakeholders can make informed decisions on which solutions best align with their unique requirements and goals.

8.3 Implications for Practice

The practical contributions made in this thesis have several implications for the practice of OGD governance and usage:

Communication Methods for OGD Awareness: Public servants can utilize the literature review and recommended strategies to design more effective communication approaches, ensuring that users are well-informed about the existence and benefits of OGD. This can lead to increased awareness and engagement with OGD initiatives.

CitizenApps: Developers and OGD managers can follow the guidelines provided by the CitizenApps mobile application to implement various requirements effectively. The federated nature of CitizenApps allows infomediairies to showcase their OGD reuses, fostering greater visibility and understanding of the significance of OGD through real-world examples. CitizenApps also plays a pivotal role in advancing OGD goals. By raising awareness, it directly bolsters transparency, ensuring citizens are informed and can hold governments accountable. This tool also acts as a catalyst for economic growth, as informed individuals and businesses can leverage OGD for innovation. Moreover, it enhances citizen engagement by highlighting the relevance of OGD, and by doing so, helps refine public services through informed decision-making based on open data insights.

ODEON: Infomediairies, particularly developers, can leverage the use case diagram and source code of ODEON to facilitate collaboration among OGD stakeholders. By using ODEON as a starting point, developers can create their own online tools that promote cooperation and community involvement in OGD projects. This collaboration proposed in ODEON ensures government accountability, as consistent feedback mechanisms hold data providers responsible for the quality and relevance of their releases. By facilitating seamless interactions, the tool can stimulate economic growth, enabling stakeholders to co-create innovative solutions leveraging OGD. Citizen engagement is deepened as they actively participate in discussions and contribute to the data ecosystem. Lastly, this synergy leads to the design and delivery of public services that better cater to actual needs, making them more efficient and user-centric.

QualityOGD: Practitioners, such as developers and publishers, can benefit from the automation, integration of metadata and data quality dimensions, and customizable user needs provided by QualityOGD to assess OGD quality effectively and efficiently. This tool streamlines the assessment of OGD quality, leading to improved data reliability and usability. It underscores the government's commitment to accountability, giving stakeholders confidence in the integrity of the data provided. Economic growth benefits as businesses can reduce risks on relying only on high-quality data to drive innovation. The tool fosters citizen engagement by ensuring that the public accesses only the highest quality of data, enhancing their trust and participation.

Dashboard Design Principles: Dashboard designers and OGD managers can draw valuable insights from the NBDash case study, which showcases the application of design principles for OGD dashboards. This guidance enables the creation of more effective and user-friendly dashboards, presenting OGD in a meaningful and engaging manner to users.

ODE and StoryOGD: Infomediairies and publishers gain the ability to create dashboards with OGD and transform conventional portals into storytelling portals, enhancing user engagement with OGD. The availability of source code access encourages continuous improvement and customization, empowering developers and publishers to tailor these tools to specific needs. These tools significantly contribute to advancing OGD objectives. Firstly, they enhance transparency by presenting complex data in an accessible format, making government actions and data usage more comprehensible to the public. Secondly, they reinforce accountability by simplifying data-driven narratives, aiding in tracking government performance and expenditures, thus enabling better oversight. Thirdly, their capacity to convert data into engaging stories fosters economic growth by providing businesses with actionable insights, thereby stimulating innovation and productivity. Fourthly, they aid in informed decision-making by offering intuitive data visualization, enabling policymakers and citizens to make evidence-based choices. Fifthly, the tools boost citizen engagement by making OGD more engaging and relatable, thus encouraging public participation and collaboration. Lastly, by offering narratives instead of raw data, the tools improve public services by making information more accessible and relevant to citizens, resulting in better service delivery and more responsive governance.

Figure 8.1 showcases the integration and interoperability of the various tools developed in this thesis, specifically focusing on ODEON, QualityOGD, ODE, and CitizenApps. The process commences when a citizen introduces a service idea in the ODEON platform (1). This idea is then refined and elaborated upon with the assistance of other stakeholders such as fellow citizens and infomediaries, ensuring the requirements for the service are precisely defined (2). Once the requirements are well-articulated, an infomediary, such as a developer, may propose to build the service utilizing the ODE prototype. As an initial step, the developer verifies the quality of the datasets that would be used in the service implementation using the QualityOGD tool (3). The outcome of this evaluation could reveal potential issues with the data, which are then compiled and reported back in the ODEON platform (4). The publisher who is responsible for these datasets can then address these reported issues and make the necessary corrections (5). Once the quality of the datasets is assured, the developer proceeds to construct the service as a dashboard using the ODE tool (6). Upon its completion, the newly created service is automatically posted on the CitizenApps platform, making it readily available to a large audience of citizens (7). Regarding StoryOGD, it embodies a synthesis of various dashboards created using the ODE tool, according to the outlined process in Figure 8.1. Additionally, each of these applications can function independently. For example, an OGD publisher might employ ODEON solely to gather information about users' data needs or to identify issues in their datasets. Subsequently, they can utilize QualityOGD separately to assess their datasets' quality before online uploading. Similarly, ODE can be used in isolation to craft visualizations or dashboards, enhancing dataset comprehension for citizens. Table 8.2 highlights some of the key use cases of the developed tools, each aligned with a particular OGD stakeholder.

In summary, the practical contributions made throughout this thesis provide tangible tools, implementations, and frameworks that can be leveraged by infomediairies, OGD managers, public servants, and researchers. These resources not only improve OGD awareness and facilitate collaboration among stakeholders but also enhance the quality assessment process and the presentation of OGD through user-friendly dashboards and storytelling portals. The accessibility of source codes further encourages customization and improvement, ensuring that the work presented in this thesis becomes a valuable resource for the broader open government data community.

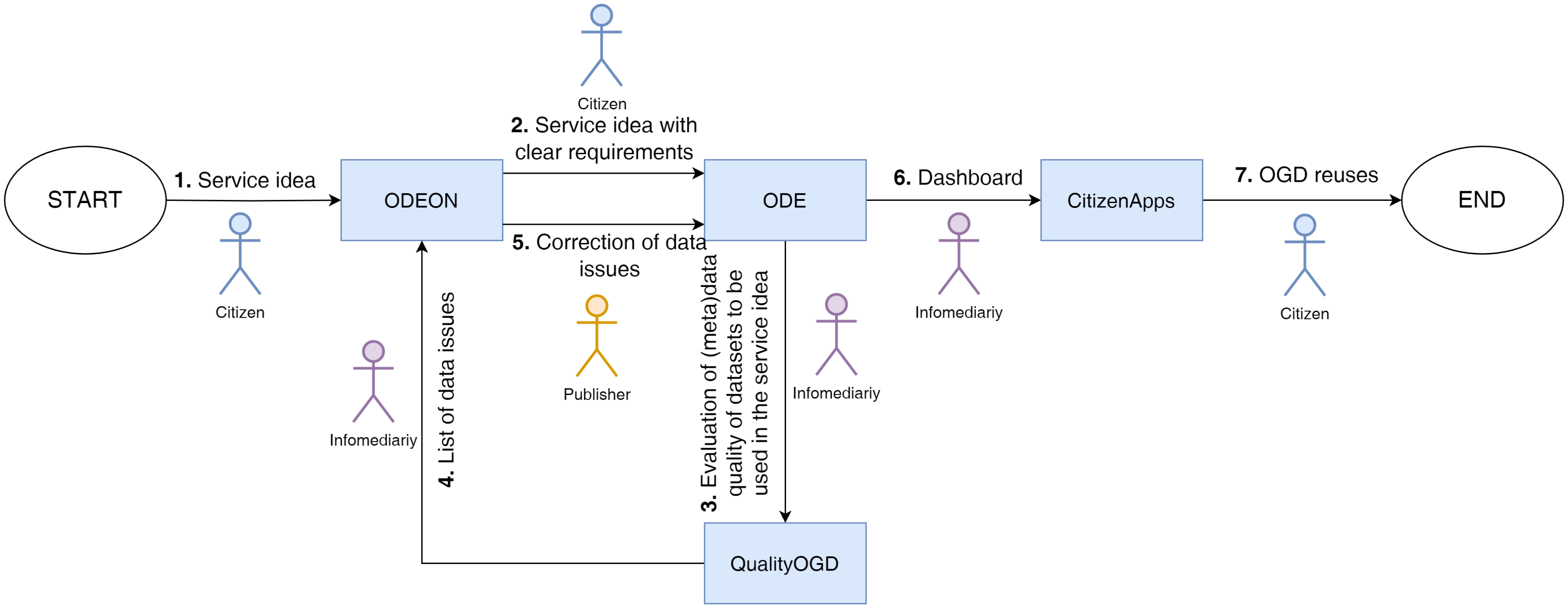


Figure 8.1: Interoperability between the different tools.

Table 8.2: Examples of how each OGD stakeholder can utilize the developed tools.

	ODEON/CitizenApps	QualityOGD	ODE/StoryOGD
Citizens	- Propose service idea - Stay updated on service development progress	Check the (meta)data quality	Create visualizations and dashboards using data
Infomediairies	- Collect user service needs - Validate or collect service idea	Check the (meta)data quality	Create visualizations and dashboards using data
Publishers	- Collect user data needs - Collect data issues	Check the (meta)data quality	Create visualizations and dashboards using data

8.4 Limitations and Future Research

The limitations of this thesis are primarily related to participant representativeness. In several studies conducted within the thesis, the number of participants was relatively small, which might impact the generalizability of the findings. While the number of participants may be considered small, previous studies (Faulkner, 2003; Nielsen, 2000) suggest that a minimum of five participants for usability tests is a good starting point, and Lallemand and Gronier, 2015 indicate that around 15 users is enough to study general tendency. To mitigate this limitation, future work should focus on increasing the number of participants by employing alternative communication channels or conducting on-site data collection at universities or public spaces. Unfortunately, the COVID-19 pandemic constrained the feasibility of such approaches during the research period.

Another limitation of this thesis is its limited focus on specific stakeholders, which may inadvertently overlook the perspectives of other potential users. The evaluation processes, including surveys and usability tests, were largely centered around a predefined user set, which might not holistically encapsulate the diverse range of stakeholders participating in OGD initiatives. Furthermore, in this thesis, the demographic data of the participants are often underutilized or completely disregarded. This could potentially omit valuable insights as demographic factors like age, profession, educational background, or location can significantly influence how different users perceive and utilize open government data. For future investigations, it would be beneficial to encompass a broader spectrum of user profiles to offer a more comprehensive understanding of user behavior and preferences. For instance, private sector enterprises, non-profit organizations, researchers, students, and citizens from varying socio-economic backgrounds could be included to ensure a more representative study population. In addition, future research should look into the implications of demographic factors on OGD usage. Understanding how these elements play a role in shaping the user interaction with open government data could yield valuable insights for tailoring data presentation and improving user engagement. By considering these factors, future work can offer more personalized and effective data delivery solutions, thus promoting wider and more efficient use of OGD.

In some studies, the research focused on specific technical aspects, overlooking other crucial dimensions, such as social, motivational, and emotional aspects related to the reuse of open government data. Addressing these non-technical aspects could provide a more comprehensive understanding of the challenges and opportunities in promoting open government data reuse. Future researchers should consider integrating these dimensions to provide a holistic view of user engagement with open government data.

Furthermore, many studies and evaluations in this thesis were specifically tailored to Namur, Belgium. This regional focus might limit the broader applicability of our results. For example, our study on identifying high-value datasets is rooted in the Wallonia context, making it potentially less relevant in different cultural, social, or governance settings. While the majority of our requirements were derived from a literature review, suggesting a certain level of generalizability, they would need further validation in diverse contexts for a comprehensive generalization. Consequently, future research should explore these solutions in varied regional or international settings to confirm and refine their applicability, ensuring they remain pertinent globally. Such endeavors would offer a more encompassing perspective on the efficacy of our proposed methods across different scenarios.

In essence, this thesis offers crucial knowledge and practical tools to boost open government data reuse, although its effectiveness could be further improved by addressing the highlighted limitations. Future work can amplify the reach and practicality of this research by broadening participant representation, accommodating varied user profiles, integrating non-technical aspects, and carrying out evaluations in different settings. In addition to improving upon these limitations, further research directions are proposed as follows.

Integration of tools with existing systems: The proposed prototypes within the thesis have been evaluated in isolation without considering their integration with existing open government data systems or platforms. The integration of implemented and other OGD tools with existing systems, such as OGD portals, is an essential area of future research that can greatly enhance the accessibility and usability of these tools. OGD portals serve as central repositories for government datasets (Erickson et al., 2013; Lněnička et al., 2021), making them an ideal platform to host and showcase collaboration, data quality, data analysis, visualization and other tools. One key aspect of this research topic revolves around incorporating the tools developed in this thesis directly into open government data portals. The specific mechanics of this integration will depend on the portal's existing infrastructure and the tool in question. The insights gathered during this process would provide valuable information on how such integrations could be made more efficient and effective, potentially leading to best practices for future tool integration in OGD portals. Moreover, research in this area should investigate user adoption and acceptance of integrated OGD tools. Understanding user needs, needs, and feedback is crucial for designing effective integrations. User studies, surveys, and usability testing can provide valuable insights into how well the integrated tools meet user expectations and what improvements can be made. Additionally, future research could explore the interoperability between the integrated OGD tools. As the volume of OGD continues to grow, the tools must be capable of handling large datasets and processing complex analyses efficiently. Research should focus on optimizing the performance of integrated tools to ensure they can handle increasing data volumes and user demands. In summary, future research could explore the technical and usability challenges associated with integrating OGD tools into existing systems, as well as the impacts of such integration on user engagement and data use. Studies could also examine the specific features and design elements that make integrated systems more user-friendly and effective. Moreover, researchers could investigate how integrated systems are adopted and used by different stakeholder groups, and how they can be designed to meet the diverse needs of these groups.

Integration of Generative AI for tools: Generative AI refers to artificial intelligence that can generate novel content, rather than simply analyzing or acting on existing data like expert systems (Gozalo-Brizuela and Garrido-Merchan, 2023). A prominent example is GPT-3, a language model developed by OpenAI, which generates human-like text based on a provided prompt. In the context of this thesis, the integration of Generative AI to the implemented tools can be detailed as below. Future research can explore the application of Generative AI to enhance user experience and interaction with the developed tools specially ODE and StoryOGD in several innovative ways. One of the ways is by developing sophisticated Natural Language Processing (NLP) interfaces that can understand and respond to user queries in natural language, thereby making data exploration and interaction more intuitive and accessible to non-technical users, which could potentially increase the range of data reuse. Another way Generative AI can be harnessed is by enabling the auto-generation of visualizations and insights based on user queries in ODE and StoryOGD prototypes. This would remove the need for users to manually create visualizations, allowing them to quickly gain insights from the data. Generative AI algorithms can also be used to learn from each user's interaction history and create personalized user experiences. This might involve recommending datasets, visualizations, or potential collaborations that align with a user's interests and needs. Furthermore, Generative AI could be employed to develop interactive tutorials and guidance for users of OGD tools. This could assist new users in learning how to use the tools effectively and provide ongoing support for all users. Lastly, the automatic generation of compelling data stories based on selected datasets could be enabled by Generative AI. This feature could significantly lower the barrier for users wanting to share their insights with a broader audience. Despite the potential benefits of these developments, future research would need to address several challenges. These include ensuring the accuracy and relevance of AI-generated content, managing the complexity of natural language interactions, and addressing privacy and security concerns associated with personalized experiences. Nevertheless, these research endeavors could contribute significantly to the creation of more user-friendly, powerful, and versatile OGD tools.

Gamification of Tools: Gamification refers to "a design approach of enhancing services and systems with affordances for experiences similar to those created by games" (Koivisto and Hamari, 2019). It mainly consists of the application of game-design elements and game principles in non-game contexts. In the context of OGD tools, the idea of gamification could introduce a whole new level of interaction and engagement. Implementing gamification in OGD tools can make the experience of using these tools more engaging and rewarding for users. By adding game-like elements such as points, levels, challenges, leaderboards, and rewards, users may feel more motivated to engage with the tool (Simonofski et al., 2022; Blazhko et al., 2017). This could increase usage and make users more likely to explore the full functionality of the tool, potentially leading to higher levels of data reuse. For instance, in the context of this thesis, users could be awarded points for interacting with different datasets, proposing service ideas, solving issues raised by other users, creating visualizations and dashboards, offering feedback on suggested visualizations or using advanced features of the tool (Simonofski et al., 2022). They could earn badges for reaching certain milestones, such as using the tool a certain number of times or interacting with a specific number of dataset (Simonofski et al., 2022). A leaderboard could display the users with the most points or badges, encouraging friendly competition. However, it's important to note that gamification isn't just about adding points and badges – it should be designed with the user's needs and motivations in mind. Understanding what motivates the tool's users, what challenges they face, and what they hope to achieve can inform the design of effective gamification strategies.

Future research in this area could involve conducting user research to understand these motivations and needs, developing and implementing gamification strategies, and evaluating the impact of these strategies on user engagement and data reuse. Moreover, the process of gamifying an OGD tool could offer insights for broader applications of gamification in the field of data analysis and visualization. It could help identify which types of game elements are most effective in this context, and how gamification impacts the way users interact with data.

CHAPTER 9

CONCLUSION

This thesis investigates three key barriers to the reuse of Open Government Data (OGD) - the lack of communication and collaboration among stakeholders, the insufficient (meta)data quality on OGD portals, and the difficulty for users to turn data into stories. These three barriers were handpicked based on a literature review and their pertinence to the situation in Wallonia (Belgium). They were tackled in a manner that ensured the research results held practical and theoretical value.

The first barrier tackled was the lack of communication and collaboration among stakeholders involved in OGD reuse. Firstly, by combining a literature review and an online survey of citizens, effective methods of raising OGD awareness were identified, underlining the potential of applications to convey its benefits. Secondly, through in-depth interviews with citizens, essential requirements for an effective awareness-raising application, termed CitizenApps, were outlined and evaluated, providing crucial insights for future development and assessments of similar applications. Lastly, we addressed the technical dimensions of stakeholder communication and collaboration. Employing a literature review, stakeholder interviews, and prototype development and evaluation, we offered a comprehensive list of requirements and a detailed use case diagram for a tool aimed at bridging the communication gap among stakeholders, thereby informing continuous improvement strategies. By providing recommendations for communication methods and a user-friendly application for OGD awareness and collaboration, this research has paved the way for more efficient and effective interactions between citizens, infomediaries, and OGD publishers.

The second barrier addressed was the insufficient (meta)data quality on OGD portals. The importance of high-quality data cannot be overstated as it is a critical element for unlocking the full potential of OGD. This study embarked on a comprehensive literature review to pinpoint the essential quality dimensions and metrics, as well as initial user requirements that would help gauge and improve data quality. These insights were then used to develop QualityOGD, a conversational agent designed to evaluate both the metadata and the data content quality while also considering user needs during the quality assessment procedure. By embedding these critical quality dimensions and metrics into the assessment tool, this research provides a dependable means for OGD stakeholders to accurately evaluate the quality of their data.

The third barrier tackled was the difficulty for users to transform data into meaningful stories. Data storytelling is crucial for making informed decisions. We addressed this objective from three angles. Initially, user needs for datasets and visualizations on OGD portals were identified, offering infomediaries and OGD publishers valuable insights for application development and data publication. Next, we distilled a comprehensive set of dashboard design principles to enhance citizen engagement with OGD, demonstrating their effectiveness through a practical tool, NBDash. Finally, we addressed the challenge of data storytelling by identifying the essential features for a data storytelling tool and integrated them into the ODE prototype and its extension, StoryOGD. These tools empower users to craft compelling narratives, regardless of their level of technical expertise, and also aid publishers in transforming their portals into storytelling platforms.

The contributions of this research encompass theoretical and practical aspects. The theoretical contributions include the identification of communication methods for OGD awareness, the requirements for OGD awareness applications, requirements for collaboration tools, metadata and data quality assessment dimensions and metrics, data and visualization expectations for OGD, dashboard design principles, and requirements for end-to-end data storytelling tools. On the other hand, the practical contributions consist of the development of CitizenApps, ODEON, QualityOGD, NBDash, ODE, and StoryOGD prototypes, serving as proxies for validating the identified requirements.

The implications for research extend beyond the specific contributions, as the findings have been disseminated through various channels, enabling other researchers to build upon and extend this work. Moreover, the provided data collection instruments and requirements offer valuable resources for future investigations in the realm of open government data governance. By utilizing these requirements, stakeholders can evaluate the suitability of specific tools and benchmark them against established standards to make informed decisions.

This thesis serves as a stepping stone towards a more transparent and accountable government by addressing key barriers to OGD reuse. By promoting effective communication and collaboration, ensuring data quality, and empowering users to transform data into compelling stories, this research contributes to the advancement of open government initiatives. However, the journey is far from over; there are still abundant opportunities for further research and enhancement. This study, although detailed, has limitations, both in terms of its overall approach and the specifics of individual contributions. Possible ways to alleviate these limitations have been discussed, along with the new research opportunities they present. In addition, three other directions for future work have been identified, which were not directly derived from the limitations: integration of tools with existing systems, incorporation of Generative AI, and gamification of tools.

Part V

Appendix

APPENDIX A

QUESTIONNAIRE FOR USER AWARENESS TO OGD

The purpose of these questions was to investigate the channels used and preferred by users to be aware of the existence and usefulness of open (government) data.

Q1. Are you aware of the existence or usefulness of open (government) data?

- O Yes
- O No

If "Yes", how (or through what channels) did you hear about the existence or usefulness of open (government) data?

Q2. How (or through what channels) would you have preferred to be informed

about the existence or usefulness of open (government) data?

Q3. How confident are you in computer science?

Not confident at all ☐ ☐ ☐ ☐ ☐ Very confident

Q4. How old are you?

- O 18 - 29
- O 30 - 49
- O 50 +

Q5. What is your level of education?

- O None
- O Primary
- O High School
- O High Education
- O PhD

APPENDIX B

QUESTIONNAIRE FOR EVALUATION OF CITIZENAPPS

The purpose of these questions was to evaluate the application CitizenApps.

Q1. What are your general impressions of CitizenApps? Does it meet your expectations?

Q2. To what extent do you agree with the following statements?

-In my opinion, it is desirable to use CitizenApps
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I think it is good for me to use CitizenApps
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-Overall, my attitude towards CitizenApps is favorable
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-CitizenApps enhances effectiveness in understanding the existence of OGD and its usefulness
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-CitizenApps makes it easier to understand the existence of OGD and its usefulness
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-CitizenApps enables me to understand more quickly the existence of OGD and its usefulness
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-Interacting with CitizenApps does not require a lot of mental effort
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I find CitizenApps to be easy to use
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I find it easy to get CitizenApps to do what I want to do
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I predict that I will use CitizenApps on a regular basis in the future
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I plan to use CitizenApps often
Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

Q3. Would you like to give some further explanation of your answers to the above questions?

Q4. Do you have any other comments and/or suggestions regarding CitizenApps? If so, can you write them down?

Q5. How confident are you in computer science?

Not confident at all ☐ ☐ ☐ ☐ ☐ Very confident

Q6. How old are you?

O 18 - 29
O 30 - 49
O 50 +

Q7. What is your level of education?

O None
O Primary
O High School
O High Education
O PhD

APPENDIX C

QUESTIONNAIRE AND INTERVIEWS FOR COMMUNICATION AMONG OGD STAKEHOLDERS

C.1 Interviews Questions during Requirements Validation

These questions served the purpose of validating the predefined requirements aimed at overcoming the impediments that arise due to the technical aspects of the lack of communication and collaboration among OGD stakeholders.

Collect user expectations
What are your expectations in terms of the functionalities to have in the OGD repository?
Collect user feedback on the predefined requirements
On a scale of 1 (Totally irrelevant) to 5 (Totally relevant), how relevant are the following requirements? And how to implement them?
Inform citizens of existing services based on OGD
Allow citizens to be involved in the service development process
Allow citizens to propose services
Allow developers to register existing services
Inform developers about existing services based on OGD
Allow developers to report issues related to the use of published datasets
Allow developers to request datasets that do not exist on the portals
Provide an open forum for discussion among OGD stakeholders
Inform developers about the datasets used in projects
Inform data providers about existing services based on OGD
Inform data providers of priority data to be published
Inform data providers of projects using their datasets

C.2 Survey Questions for Prototype Evaluation

These questions were utilized to assess the ease of use and effectiveness of the CitizenApps prototype in addressing the impediments that arise due to the technical aspects of the lack of communication and collaboration among OGD stakeholders, as well as to collect suggestions for additional features to be included in future versions.

What are your general impressions of the prototype presented?
Does it meet your expectations?

Evaluation of the usability of the prototype features
To what extent do you agree with the following statements?
Extremely Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Agree
As a citizen, this prototype is easy to
Discover existing and requested services using OGD
Suggest a project idea to develop
Follow the evolution of a service implementation
Give my feedback on project
Request access to a dataset
Give my general opinion or report a problem on a dataset
Register an existing project using OGD
As a developer, this prototype is easy to
Register an existing project using OGD
Suggest a service idea to get feedback from end users
Discover existing and requested services using OGD
Inform users on the evolution of the implementation of a service
Give my feedback on project
Know the datasets used in project
Request access to a dataset
Give my general opinion or report a problem on a dataset
As a publisher, this prototype is easy to
Discover existing and requested services using OGD
Know the projects using my datasets
Know the datasets requested by the users
Know the problems reported on a dataset
Respond to a reported problem

Would you like to give some further explanation of your answers to the above questions?

Evaluation of the usability of the prototype
To what extent do you agree with the following statements?
Extremely Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Agree
Learning how to use this prototype was easy for me
I find it easy to get this prototype to do what I want it to do
My interaction with this prototype is clear and understandable
I find this prototype clear and understandable
It would be easy for me to become skillful a of this prototype
Overall, I think this prototype is easy to use

Would you like to give some further explanation of your answers to the above questions?

Evaluation of the usefulness of the prototype features
To what extent do you agree with the following statements?
Extremely Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Agree
As a citizen, this prototype is useful for
Understanding the usefulness of OGD
Know the existing services using OGD
Participate in the development of new services
Communicate with other OGD stakeholders

As a developer, this prototype is useful for
Have ideas on projects to develop
Know the services using OGD and the datasets used in these services
Get feedback from users on an implemented or proposed service
Communicate with other OGD stakeholders
As a data provider, this prototype is useful for
Know which datasets to publish in order of relevance
Know the problems reported on a dataset
Know the projects using a dataset
Communicate with other OGD stakeholders

Evaluation of the usefulness of the prototype
To what extent do you agree with the following statements?
Extremely Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Agree
Overall, I think this prototype is useful in facilitating communication between OGD stakeholders

Would you like to give some further explanation of your answer to the above questions?
Do you have any other comments and/or suggestions regarding the presented prototype? If so, can you write them down?

Demographic Information	
How old are you?	[18 - 29 / 30 - 49 / 50 +]
What is your gender?	[Female / Male / Other]
What is your level of education?	[None / Primary / High School / High Education / PhD]
Indicate your occupation	[Student / Employed / Self-employed / Retired / Unemployed]
Indicate your organization type	[Public / Private / University]

APPENDIX D

Questionnaire for Evaluation of a Conversational Agent for OGD Quality Assessment

The aim of these questions was to evaluate the conversational agent developed to assess the (meta)data quality of OGD.

Q1. What are your general impressions of the prototype presented?

Q2. Evaluation of the usability of the prototype: To what extent do you agree with the following statements?

-I think that I would like to use this prototype frequently when working with OGD
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I found the prototype unnecessarily complex
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I thought the prototype was easy to use
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I think that I would need the support of a technical person to be able to use this prototype
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I found the various functions in this prototype were well integrated
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I thought there was too much inconsistency in this prototype
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I imagine that most people would learn to use this prototype very quickly
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I found the prototype very cumbersome to use
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I felt very confident using the prototype
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree
-I needed to learn a lot of things before I could get going with this prototype
Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

Q3. Evaluation of the usefulness of the prototype: To what extent do you agree with the following statements?

-*This prototype is useful to evaluate the overall quality of a dataset (content + metadata)*

Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

-*This prototype is useful for evaluating the quality of the metadata associated with a dataset*

Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

-*This prototype is useful to evaluate the quality of the content of a dataset*

Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

-*This prototype is useful to better highlight errors related to a dataset*

Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

-*This prototype is useful to take into account the user's preferences when evaluating the quality of a dataset*

Strongly Disagree ☐ ☐ ☐ ☐ ☐ Strongly Agree

Q4. Would you like to give some further explanation of your answers to the above questions?

Q5. Do you have any other comments and/or suggestions regarding the presented prototype? If so, can you write them down?

Q6. How confident are you in your data handling expertise?

Not confident at all ☐ ☐ ☐ ☐ ☐ Very confident

Q7. How old are you?

O 18 - 29
O 30 - 49
O 50 +

Q8. What is your level of education?

O None
O Primary
O High School
O High Education
O PhD

APPENDIX E

Questionnaire for Data Expectations on OGD Portals

The purpose of these questions was to understand the open (government) datasets that users are interested in.

Q1. Have you ever heard of Open Government Data?

- O Yes
- O No

Q2. Have you ever used an Open Government Data website? An OGD portal is a website where the municipality makes useful data available to all citizens.

- O Yes
- O No

Q3. What municipal data would you like to consult on an Open Government Data website?

Q4. Please indicate the level of importance of the data you would like to see open to the public on a website.

-*Communal Council reports*
Not at all important ☐ ☐ ☐ ☐ Very important
-Current and past legislation
Not at all important ☐ ☐ ☐ ☐ Very important
-*Municipal address guide*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Overview of organized events*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Restaurants and shopping places*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Sports facilities*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Location of public art*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Historical archives*
Not at all important ☐ ☐ ☐ ☐ Very important
-*List of historical monuments*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Grants and budget*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Waste containers*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Waste calendar*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Tree information*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Teaching quality*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Safety information*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Public toilets*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Dog-walking areas*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Spreading itineraries*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Parking spaces*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Street lighting*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Mobility information (traffic, roadworks)*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Information about COVID*
Not at all important ☐ ☐ ☐ ☐ Very important
-*Health care facilities*
Not at all important ☐ ☐ ☐ ☐ Very important

Q5. How confident are you in computer science?

Not confident at all ☐ ☐ ☐ ☐ ☐ Very confident

Q6. Would you like to take part in the next phase of this project via a 30-minute interview? The purpose of this interview will be to better understand how you would like your data to be presented on the Open Data site.

O Yes
O No

If yes, please enter your e-mail address

APPENDIX F

Questionnaire for Evaluation of ODE

The purpose of these questions was to evaluate the application ODE (Open Data Explorer).

Q1. What are your general impressions of the prototype presented? Does it meet your expectations?

Q2. Evaluation of the usability of the prototype: To what extent do you agree with the following statements?

-Learning how to use this prototype was easy for me

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-I find it easy to make this prototype do what I want it to do

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-My interaction with this prototype is clear and understandable

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-I find this prototype clear and understandable

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-It would be easy for me to become proficient in the use of this prototype

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Overall, I think this prototype is easy to use

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

Q3. Evaluation of the usefulness of the prototype: To what extent do you agree with the following statements? This prototype is useful for

-Have direct access to open data without access to portals

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Integrate multiple open data directly into a visualization or dashboard

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Have a general view on the content of an open data

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-To have an estimation of the quality of an open data

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Pre-processing of open data (e.g. deleting columns, grouping several open data)

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Create a visualization with open data

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Export a visualization

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Have visualization recommendations on an open data

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Share or export a dashboard

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Collect user feedback or give feedback on a dashboard

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

-Overall, I think this prototype is useful for manipulating open data and also turning it into visualizations and dashboards

Strongly Disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly Agree

Q4. Would you like to give some further explanation of your answers to the above questions?

Q5. Do you have any other comments and/or suggestions regarding CitizenApps? If so, can you write them down?

Q6. How confident are you in manipulating data?

Not confident at all ☐ ☐ ☐ ☐ ☐ Very confident

Q7. How old are you?

O 18 - 29
O 30 - 49
O 50 +

Q8. What is your level of education?

O None
O Primary
O High School
O High Education
O PhD

APPENDIX G

REPLICATION PACKAGE

This section presents a replication package, designed to guide researchers in reproducing the array of studies featured in this thesis. Alongside methodologies detailed in earlier sections (5.2.1, 5.4.2, 6.3, 7.3.1.1, and 7.4.2), and the source codes for the developed prototypes (accessible at https://github.com/chokkipaterne), as well as previously mentioned questionnaires in Part V, additional resources are provided. These resources encompass a protocol for literature reviews and a standard procedure for user testing during prototype evaluations. Subsequent sections further elaborate on these added materials.

G.1 Literature Review Protocol

The protocol for the literature reviews conducted in this thesis encompasses the subsequent steps.
Search Strategy: For each literature review in this thesis, we start by selecting appropriate keywords and databases for the search. Keywords are chosen based on their relevance to the research question. We primarily rely on databases such as "Scopus", "Science Direct", "Association for Computing Machinery", and "Google Scholar". After defining our primary search string using the selected keywords, we modify it to fit the specific criteria of each database, ensuring a comprehensive automated search.
Article Identification Process: Articles identified through our search will be assessed for relevancy across a three-tiered evaluation system:
Preliminary Screening: At this stage, articles are evaluated based on their publication type, domain specificity, and title. Exclusion criteria encompass: duplicate articles, titles bearing no direct correlation to our key search terms, studies anchored in domains not germane to our research focus (e.g., healthcare or advanced mathematics), and articles not penned in English.
Abstract Evaluation: Articles clearing the initial screening are then assessed by examining their abstracts. We filter out those whose abstracts diverge significantly from our primary keywords or research objectives.
In-depth Content Examination: Subsequent to the abstract evaluation, we conduct a thorough assessment of the full content of the remaining articles. This deeper dive helps in weeding out publications that, despite having relevant abstracts, do not align holistically with our study's objectives.

Iterative Search Approach: To bolster the comprehensiveness of our literature review, we employ the forward and backward search method as elucidated by Webster and Watson (2002). This entails inspecting references and citations within our primary set of articles to uncover additional germane publications.
Incorporation of Grey Literature: Recognizing the value of non-traditional sources, we supplement our core set of articles by delving into the grey literature. We utilize the Google search engine to identify pertinent reports, theses, and white papers not found in conventional academic repositories.

This rigorous, multi-pronged approach ensures a holistic and comprehensive review of the available literature pertinent to our study.

G.2 User Testing Protocol

Before conducting user testing, we engaged participants primarily via Facebook, emails, and by liaising with students across different faculties at the University of Namur. Our selection was influenced by the specific participant profiles we needed for evaluating the prototype. Additionally, we designed questionnaires for post-test feedback, ensuring they were clear and easy to understand. We also outlined clear tasks and guidelines for participants to follow during the assessment.

With these preliminary steps in place, our main user testing unfolded in these key phases:

Execution Phase:

- Each session commenced with a succinct orientation, highlighting the test's goals and ensuring participants that the evaluation was product-centric.
- Participants were directed to execute the tasks, simultaneously verbalizing their thought process.
- We meticulously observed and documented their interactions, pinpointing any hurdles or issues encountered.

Feedback Collection:

- Post task completion, we facilitated open discussions and provided our predesigned questionnaire to garner their feedback.
- We emphasized the value of transparent, both affirmative and critical, feedback.

Debrief:

- Post-test, participants were encouraged to offer any additional thoughts or feedback.

Data Analysis:

- We thoroughly evaluated the amassed notes, recordings, and other data. For metrics like Likert scale responses, we computed the median, mean, and standard deviation (SD). For subjective data, like spoken reflections and open-ended query responses, we summarized them into succinct descriptions, maintaining their original context and meaning.
- This analysis aided in identifying prevalent patterns, pain points, and areas for optimization.

BIBLIOGRAPHY

Tricia Aanderud, Boston Scientific, and Jaime D Agord. My Sharky Secrets for Telling Fabulous Data Stories. Technical report, SAS Global Forum, 2020.

Omer Hassan Abdelrahman. Open Government Data: Development, Practice, and Challenges. In **Open Data**. IntechOpen, nov 2021. ISBN 978-1-83968-316-9. doi: 10.5772/INTECHOPEN.100465. URL https://www.intechopen.com/online-first/78857.

Roobaea Alroobaea and Pam J. Mayhew. How many participants are really enough for usability studies? In **Science and Information Conference 2014**, pages 48–56, 2014. ISBN 9780989319317. doi: 10.1109/SAI.2014.6918171.

Lameck M Amugongo, Hippolyte N Muyingi, and Jürgen Sieck. Increasing open data awareness and consumption in Namibia : A hackathon approach. In **13th Culture and Computer Science Conference**, pages 187–198, 2015.

Bahareh Ansari, Mehdi Barati, and Erika G. Martin. Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research. **Government Information Quarterly**, 39 (1), 2022. ISSN 0740624X. doi: 10.1016/j.giq.2021.101657. URL https://doi.org/10.1016/j.giq.2021.101657.

Ana Carolina Araújo, Lucas Reis, and Rafael Cardoso Sampaio. Do transparency and open data walk together? An analysis of initiatives in five Brazilian Capitals. **Media Studies**, 7(14):65–83, 2016. ISSN 18485030. doi: 10.20901/ms.7.14.6.

Judie Attard, Fabrizio Orlandi, Simon Scerri, and Sören Auer. A systematic review of open government data initiatives. **Government Information Quarterly**, 32(4): 399–418, oct 2015.

Rosemeire Barbosa Tavares, Mark Hepworth, and Sely Maria De Souza Costa. Investigating citizens' information needs through participative research: a pilot study in Candangolândia, Brazil:. **American Journal of Health Promotion**, 27(2):125–138, may 2011. doi: 10.1177/0266666911401167. URL https://journals.sagepub.com/doi/abs/10.1177/0266666911401167.

Richard Baskerville. What design science is not. **European Journal of Information Systems**, 17:441–443, 2008.

Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. Methodologies for data quality assessment and improvement. **ACM Computing Surveys**, 41(3), 2009. ISSN 03600300. doi: 10.1145/1541880.1541883.

Martin Beno, Kathrin Figl, Jurgen Umbrich, and Axel Polleres. Open data hopes and fears: Determining the barriers of open data. **Proc. 7th International Conference for E-Democracy and Open Government CeDEM 2017**, pages 69–81, 2017. doi: 10.1109/CeDEM.2017.22.

Jorn Berends, Wendy Carrara, and Heleen Vollers. Analytical Report 6: Open Data in Cities 2. Technical report, European Commission, 2020. URL https://data.europa.eu/sites/default/files/edp{_}analytical{_}report{_}n6{_}-{_}open{_}data{_}in{_}cities{_}2{_}-{_}final-clean.pdf.

Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. **Political Analysis**, 20(3):351–368, 2012. ISSN 10471987. doi: 10.1093/pan/mpr057.

John C. Bertot, Paul T. Jaeger, and Justin M. Grimes. Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. **Government Information Quarterly**, 27(3):264–271, 2010. ISSN 0740624X. URL http://dx.doi.org/10.1016/j.giq.2010.03.001.

Sovit Bhandari, Navin Ranjan, Yeong Chan Kim, Jong Do Park, Kwang Il Hwang, Woo Hyuk Kim, Youn Sik Hong, and Hoon Kim. An automatic data completeness check framework for open government data. **Applied Sciences**, 11(19):1–17, 2021. ISSN 20763417. doi: 10.3390/app11199270.

Oleksandr Blazhko, Tetiana Luhova, Sergey Melnik, and Viktoriia Ruvinska. Communication model of open government data gamification based on ukrainian websites. In **2017 4th Experiment@ International Conference (exp. at'17)**, pages 181–186. IEEE, 2017.

Christoph Böhm, Markus Freitag, Arvid Heise, Claudia Lehmann, Andrina Mascher, Felix Naumann, Vuk Ercegovac, Mauricio Hernandez, Peter Haase, and Michael Schmidt. GovWILD: Integrating Open Government Data for transparency. **Proceedings of the 21st Annual Conference on World Wide Web Companion**, pages 321–324, 2012. doi: 10.1145/2187980.2188039.

Harry N. Boone and Deborah A. Boone. Analyzing Likert Data. **Journal of Extension**, 50(2), apr 2012. ISSN EISSN-1077-5315.

Richart Brath and Michael Peters. Dashboard design: why design is important. **DM Review Online**, oct 2004. URL http://cs.furman.edu/{~}pbatchelor/csc105/articles/TUN{_}DM{_}ONLINE.pdf.

Niall Brolcháin, Lukasz Porwol, Adegboyega Ojo, Tilman Wagner, Eva Tamara Lopez, and Eric Karstens. Extending open data platforms with storytelling features. In **18th Annual International Conference on Digital Government Research**, pages 48–53, 2017. ISBN 9781450353175. doi: 10.1145/3085228.3085283.

John Brooke. SUS - A quick and dirty usability scale. **Usability Evaluation In Industry**, 1986. doi: 10.1201/9781498710411-35.

John Brooke. SUS: A Retrospective. **Journal of Usability Studies**, 8(2):29–40, 2013. ISSN 00319376.

Jérôme Brugger, Marianne Fraefel, Reinhard Riedl, Hansjakob Fehr, Daniel Schöneck, and Christoph Stäehli Weissbrod. Current barriers to open government data use and visualization by political intermediaries. **6th International Conference for E-Democracy and Open Government CeDEM 2016**, pages 219–229, 2016. doi: 10.1109/CeDEM.2016.18.

Reinhard Budde, Karlheinz Kautz, Karin Kuhlenkamp, and Heinz Züllighoven. What is prototyping? **Information Technology and People**, 6(2/3):89–95, 1992.

Donald T. Campbell and Julian C. Stanley. **Experimental and Quasi-Experimental Designs for Research**. Chicago: Rand McNaUy, 1969.

Iván Cantador, Jesús Viejo-Tardío, María E. Cortés-Cediel, and Manuel Pedro Rodríguez-Bolívar. A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government. In **22nd Annual International Conference on Digital Government Research 2021**, pages 168–179, 2021. ISBN 9781450384926. doi: 10.1145/3463677.3463681.

Giulia Carsaniga, Eline N. Lincklaen Arriëns, Jochem Dogger, Mariska van Assen, and Gianfranco Cecconi. OPEN DATA MATURITY REPORT 2022. Technical report, EUROPEAN COMMISSION, 2022.

Zoey Chen and May Yuan. Psychology of word of mouth marketing. **Current Opinion in Psychology**, 31:7–10, 2020. ISSN 2352250X. doi: 10.1016/j.copsyc.2019.06.026. URL https://doi.org/10.1016/j.copsyc.2019.06.026.

Abiola Paterne Chokki and Benoît Vanderose. From Conventional Open Government Data Portals to Storytelling Portals : The StoryOGD Prototype. In **24th Annual International Conference on Digital Government Research**, pages 1–2, 2023.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open Government Data for Non-expert Citizens: Understanding Content and Visualizations' Expectations. **Research Challenges in Information Science**, 415 LNBIP:602–608, 2021. ISSN 18651356. doi: 10.1007/978-3-030-75018-3_42.

Abiola Paterne Chokki, Charalampos Alexopoulos, Stuti Saxena, Benoît Frénay, Benoît Vanderose, and Mohsan Ali. Metadata quality matters in Open Government Data (OGD) evaluation! An empirical investigation of OGD portals of the GCC constituents. **Transforming Government: People, Process and Policy**, pages 1–21, 2022a. doi: 10.1108/TG-09-2022-0118.

Abiola Paterne Chokki, Benoit Frenay, and Benoît Vanderose. Open Data Explorer: An End-to-end Tool for Data Storytelling using Open Data. In **Americas Conference on Information Systems 2022 Proceedings. 6**, 2022b.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, Benoît Frénay, and Benoît Vanderose. Fostering Interaction between Open Government Data Stakeholders: An Exchange Platform for Citizens, Developers and Publishers. In **Electronic Government**, 2022c.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Engaging Citizens with Open Government Data: The Value of Dashboards compared to Individual Visualizations. **Digital Government: Research and Practice**, 2022d. doi: https://doi.org/10.1145/3558099.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Open government data awareness: eliciting citizens' requirements for application design. **Transforming Government: People, Process and Policy**, 2022e. doi: https://doi.org/10.1108/TG-04-2022-0057.

Abiola Paterne Chokki, Anthony Simonofski, Benoît Frénay, and Benoît Vanderose. Increasing Awareness and Usefulness of Open Government Data: An Empirical Analysis of Communication Methods. In **16th International Conference on Research Challenges in Information Science**, 2022f.

Abiola Paterne Chokki, Anthony Simonofski, Antoine Clarinval, and Benoît Vanderose. Evaluating a Conversational Agent for Open Government Data Quality Assessment. In **29th Americas Conference on Information Systems**, pages 1–10, 2023.

Pin Yu Chu and Hsien Lee Tseng. A theoretical framework for evaluating government open data platform. In **EGOSE**, pages 135–142, 2016. ISBN 9781450348591. doi: 10.1145/3014087.3014099.

Antoine Clarinval, Anthony Simonofski, Benoît Vanderose, and Bruno Dumas. Public displays and citizen participation: a systematic literature review and research agenda. **Transforming Government: People, Process and Policy**, 15(1):1–35, 2021. ISSN 17506166. doi: 10.1108/TG-12-2019-0127.

Cihan Cobanoglu, Muhittin Cavusoglu, and Gozde Turktarhan. A beginner's guide and best practices for using crowdsourcing platforms for survey research: The Case of Amazon Mechanical Turk (MTurk). **Journal of Global Business Insights**, 6 (1):92–97, 2021. ISSN 2640-6470. doi: 10.5038/2640-6489.6.1.1177.

Jorgos Coenen, Eslam Nofal, and Andrew Vande Moere. How the arrangement of content and location impact the use of multiple distributed public displays. **Designing Interactive Systems Conference**, pages 1415–1426, 2019. doi: 10.1145/3322276.3322294.

Meghan Cook and Alexander Jurkat. An Open Government Research and Development Agenda Setting Workshop. Technical report, CTG, TWC and IILP, Albany, NY, 2011. URL https://ctg.albany.edu/media/pubs/pdfs/OGWorkshop{_}ActivityReport.pdf.

Gennaro Cordasco, Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Andrea Petta, Donato Pirozzi, Gianluca Santangelo, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo, and Luca Vicidomini. Engaging Citizens with a Social Platform for Open Data. **22nd Annual International Conference on Digital Government Research '17**, pages 242–249, 2017. doi: 10.1145/3085228.3085302.

Andreiwid Sheffer Corrêa, Pedro Luiz Pizzigatti Corrêa, and Flávio Soares Corrêa Da Silva. Transparency portals versus open government data. An assessment of openness in Brazilian municipalities. **15th Annual International Conference on Digital Government Research**, pages 178–185, 2014. doi: 10.1145/2612733.2612760.

Kevin Crowston. Amazon mechanical turk: A research tool for organizations and information systems scholars. **IFIP Advances in Information and Communication Technology**, 389 AICT:210–221, 2012. ISSN 18684238. doi: 10.1007/978-3-642-35142-6_14.

Jonathan Crusoe and Ulf Melin. Investigating open government data barriers: A literature review and conceptualization. **Electronic Government**, pages 169–183, sep 2018. ISSN 16113349. doi: 10.1007/978-3-319-98690-6_15. URL https://link.springer.com/chapter/10.1007/978-3-319-98690-6{_}15.

Jonathan Crusoe, Anthony Simonofski, Antoine Clarinval, and Elisabeth Gebka. The impact of impediments on open government data use: Insights from users. **13th International Conference on Research Challenges in Information Science**, pages 1–12, 2019. ISSN 21511357. doi: 10.1109/RCIS.2019.8877055.

Jonathan Crusoe, Elisabeth Gebka, and Karin Ahlin. Open Government Data from the Perspective of Information Needs - A Tentative Conceptual Model. **Electronic Government**, 12219 LNCS:250–261, 2020. ISSN 16113349. doi: 10.1007/978-3-030-57599-1_19.

Aba-Sah Dadzie and Matthew Rowe. Approaches to Visualising Linked Data: A Survey. **Semantic Web**, 2(2):89–124, 2010. URL http://vis.stanford.edu/protovis.

Data.europa.eu. Metadata Quality Assessment Methodology, 2020. URL https://data.europa.eu/mqa/methodology.

Nahla Davies. The 7 principles of UX design—and how to use them - 99designs, 2020. URL https://en.99designs.be/blog/web-digital/ux-design-principles/.

Tim Davies. Open data, democracy and public sector reform: A look at open government data use from data. gov. uk. Technical report, University of Oxford, 2010.

Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **Management Information Systems**, 13(3):319–339, 1989. doi: 10.2307/249008.

Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo, Luca Vicidomini, and Gennaro Cordasco. DatalEt-ecosystem provider (DEEP): Scalable architecture for reusable, portable and user-friendly visualizations of open data. **7th International Conference for E-Democracy and Open Government CeDEM 2017**, pages 92–101, 2017. doi: 10.1109/CeDEM.2017.14.

Makx Dekkers, Nikolaos Loutas, Michiel De Keyzer, and Stijn Goedertier. Open Data & Metadata Quality, 2014. URL https://joinup.ec.europa.eu/sites/default/files/document/2015-05/d2.1.2{_}training{_}module{_}2.2{_}open{_}data{_}quality{_}v1.00{_}en.pdf.

Aline Dresch, Daniel Pacheco Lacerda, and José Antônio Valle Antunes. Design Science Research: A Method for Science and Technology Advancement. **Springer International Publishing**, 2015.

Marc Dupuis, Barbara Endicott-Popovsky, and Robert Crossler. An Analysis of the Use of Amazon's Mechanical Turk for Survey Research in the Cloud. In **ICCSM**, 2013.

Sandra Durcevic. Top 20 Dashboard Design Principles, Best Practices & How To's, 2020. URL https://www.datapine.com/blog/dashboard-design-principles-and-best-practices/.

Erik Duval. Metadata Standards, What, Who & Why. **Journal of Universal Computer Science**, 7(7):591–601, 2001.

André Eberhardt and Milene Selbach Silveira. Show me the data! A systematic mapping on open government data visualization. **19th Annual International Conference on Digital Government Research Gov. Data Age**, pages 1–10, may 2018. doi: 10.1145/3209281.3209337.

Kami Ehrich. Mechanical Turk: Potential Concerns and Their Solutions, 2020. URL https://www.summitllc.us/blog/mechanical-turk-concerns-and-solutions.

Robert Shawn Wall Emerson. Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research? **Journal of Visual Impairment and Blindness**, 109(2):164–168, 2015.

John S. Erickson, Amar Viswanathan, Joshua Shinavier, Yongmei Shi, and James A. Hendler. Open government data: A data analytics approach. **IEEE Intelligent Systems**, 28(5):19–23, sep 2013. ISSN 15411672. doi: 10.1109/MIS.2013.134.

EUROPEAN COMMISSION. Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive. Technical report, EUROPEAN COMMISSION, 2020. URL https://www.access-info.org/wp-content/uploads/Deloitte-Study-2020.pdf.

European Environment Agency. Open data and e-government good practices for fostering environmental information sharing and dissemination. Technical report, European Environment Agency, 2019. URL https://unece.org/fileadmin/DAM/env/pp/a{_}to{_}i/Joint{_}UNECE-EEA{_}workshop/Draft{_}OD{_}EGOV{_}GP{_}.pdf.

Adir Even and G. Shankaranarayanan. Utility-Driven Assessment of Data Quality. **ACM SIGMIS Database: the DATABASE for Advances in Information Systems**, 38 (2):75–93, 2007. ISSN 00950033. doi: 10.1145/1240616.1240623.

Laura Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. **Behavior Research Methods, Instruments, and Computers 35**, pages 379–383, 2003.

Stephen Few. Common Pitfalls in Dashboard Design. **ProClarity Corporation**, page 31, 2006a. URL https://www.perceptualedge.com/articles/Whitepapers/Common{_}Pitfalls.pdf.

Stephen Few. Information Dashboard Design : The Effective Visual Communication of Data, 2006b.

Martin Fowler and Jim Highsmith. The Agile Manifesto. **Software development**, 9 (8):28–35, 2001.

FuturoCité. Baromètre de la Culture de la donnée – Edition 2022, 2022. URL https://www.futurocite.be/activites/enquete-culture-de-donnee/barometre-de-la-culture-de-la-donnee-edition-2022/.

Sukumar Ganapati. Use of Dashboards in Government. In **Fostering Transparency and Democracy Series**. IBM Center for the Business of Government, 2011. URL www.businessofgovernment.org.

Mila Gascó-Hernández, Erika G. Martin, Luigi Reggi, Sunyoung Pyo, and Luis F. Luna-Reyes. Promoting the use of open government data: Cases of training and engagement. **Government Information Quarterly**, 35(2):233–242, 2018. ISSN 0740624X. doi: 10.1016/j.giq.2018.01.003. URL https://doi.org/10.1016/j.giq.2018.01.003.

Elisabeth Gebka, Antoine Clarinval, Jonathan Crusoe, and Anthony Simonofski. Generating value with open government data: Beyond the programmer. **13th International Conference on Research Challenges in Information Science**, pages 1–2, 2019. ISSN 21511357. doi: 10.1109/RCIS.2019.8877054.

Elisabeth Gebka, Jonathan Crusoe, and Karin Ahlin. Open data reuse and information needs satisfaction: A method to bridge the gap. **CEUR Workshop Proceedings**, 2797:41–49, 2020. ISSN 16130073.

Olga Gil, María E. Cortés-Cediel, and Iván Cantador. Citizen Participation and the Rise of Digital Media Platforms in Smart Governance and Smart Cities. In **Research Anthology on Citizen Engagement and Activism for Social Change**, pages 1186–1202. IGI Global, 2022.

Felipe Gonzalez-Zapata and Richard Heeks. The multiple meanings of open government data: Understanding different stakeholders and their perspectives. **Government Information Quarterly**, 32(4):441–452, may 2015. ISSN 0740-624X. doi: 10.1016/j.giq.2015.09.001.

Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchan. Chatgpt is not all you need. a state of the art review of large generative ai models. **arXiv preprint arXiv:2301.04655**, 2023.

Alvaro Graves and Javier Bustos-Jiménez. Co-creating visual overviews for open government data. **16th Annual International Conference on Digital Government Research**, pages 37–42, may 2015. doi: 10.1145/2757401.2757407.

Alvaro Graves and James Hendler. Visualization tools for Open Government Data. **14th Annual International Conference on Digital Government Research**, pages 136–145, 2013. doi: 10.1145/2479724.2479746.

Alvaro Graves and James Hendler. A study on the use of visualizations for Open Government Data. **Information Polity**, 19(1-2):73–91, 2014. ISSN 18758754. doi: 10.3233/IP-140333.

Jonathan Gray, Lucy Chambers, and Liliana Bounegru. **The Data Journalism Handbook: How Journalists Can Use Data to Improve the News**. O'Reilly Media, Inc., 2012.

Pamela Grimm. Pretesting a questionnaire, 2010.

Greg Guest, Arwen Bunce, and Laura Johnson. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. **Field methods**, 18(1):59–82, 2006.

Panom Gunawong. Open Government and Social Media: A Focus on Transparency. **Social Science Computer Review**, 33(5):587–598, 2015. ISSN 15528286. doi: 10.1177/0894439314560685.

ADAM HAYES. Word-of-Mouth Marketing, 2021. URL https://www.investopedia.com/terms/w/word-of-mouth-marketing.asp.

Jeffrey Heer and Ben Shneiderman. A taxonomy of tools that support the fluent and flexible use of visualizations. **Interactive Dynamics for Visual Analysis**, 10:1–26, 2012. ISSN 1542-7730. URL http://queue.acm.org/detail.cfm?id=2146416.

Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. **MIS Quarterly: Management Information Systems**, 28(1):75–105, 2004. ISSN 02767783. doi: 10.2307/25148625.

Julien Hivon and Ryad Titah. Conceptualizing citizen participation in open data use at the city level. **Transforming Government: People, Process and Policy**, 11(1): 99–118, 2017. ISSN 17506166. doi: 10.1108/TG-12-2015-0053.

Yan Holtz and Healy Conor. From Data to Viz, 2018. URL https://www.data-to-viz.com/.

Andrea Janes, Alberto Sillitti, and Giancarlo Succi. Effective dashboard design. **Cutter IT Journal**, 26(1):17–24, 2013. ISSN 15227383.

Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, Adoption Barriers and Myths of Open Data and Open Government. **Information Systems Management**, 29(4):258–268, 2012.

Marijn Janssen, Ricardo Matheus, Justin Longo, and Vishanth Weerakkody. Transparency-by-design as a foundation for open government. **Transforming Government: People, Process and Policy**, 11(1):2–8, 2017. ISSN 17506166. doi: 10.1108/TG-02-2017-0015.

Peter Johnson and Pamela Robinson. Civic Hackathons: Innovation, Procurement, or Civic Engagement? **Review of Policy Research**, 31(4):349–357, 2014.

Peter A. Johnson and Sarah Greene. Who are government opendata infomediaries? A preliminary scan and classification of open data users and products. **URISA Journal**, 28(1):9–18, 2017. ISSN 10458077.

Marlen C. Jurisch, Markus Kautz, Petra Wolf, and Helmut Krcmar. An international survey of the factors influencing the intention to use open government. In **Proceedings of the Annual Hawaii International Conference on System Sciences**, volume 2015-March, pages 2188–2198. IEEE Computer Society, mar 2015. ISBN 9781479973675. doi: 10.1109/HICSS.2015.262.

Olga Kalatzi, Charalampos Bratsas, and Andreas Veglis. The Principles, Features and Techniques of Data Journalism. **Studies in Media and Communication**, 6(2): 36–44, 2018. ISSN 2325-8071. doi: 10.11114/smc.v6i2.3208.

Sophia Keyner, Vadim Savenkov, and Svitlana Vakulenko. Open Data Chatbot. In **ESWC 2019**, pages 111–115, 2019. ISBN 9783030323264. doi: 10.1007/978-3-030-32327-1_22.

Muzammil Khan and Sarwar Shah Khan. Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. **International Journal of Computer Applications**, 34(1), 2011.

Pearl Kitchenham, Barbara Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering-A systematic literature review. **Information and Software Technology**, 51(1):7–15, 2009.

Rob Kitchin and Gavin Mcardle. Urban data and city dashboards: Six key issues. In **Data and the City**. Routledge, 2016. URL http://progcity.maynoothuniversity.ie/.

Rob Kitchin, Tracey P. Lauriault, and Gavin McArdle. Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. **Regional Studies, Regional Science**, 2(1):6–28, 2015. ISSN 21681376. doi: 10.1080/21681376.2014.983149.

Jonna Koivisto and Juho Hamari. The rise of motivational information systems: A review of gamification research. **International Journal of Information Management**, 45(July 2018):191–210, 2019. ISSN 02684012. doi: 10.1016/j.ijinfomgt.2018.10.013. URL https://doi.org/10.1016/j.ijinfomgt.2018.10.013.

Sylvain Kubler, Jérémy Robert, Yves Le Traon, Jörgen Umbrich, and Sebastian Neumaier. Open data portal quality comparison using AHP. In **22nd Annual International Conference on Digital Government Research '16**, pages 397–407, 2016. ISBN 9781450343398. doi: 10.1145/2912160.2912167.

Sylvain Kubler, Jérémy Robert, Sebastian Neumaier, Jürgen Umbrich, and Yves Le Traon. Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. **Government Information Quarterly**, 35(1):13–29, 2018. ISSN 0740624X. doi: 10.1016/j.giq.2017.11.003. URL https://doi.org/10.1016/j.giq.2017.11.003.

Carine Lallemand and Guillaume Gronier. **Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs**. Eyrolles, 2015.

Josefin Lassinantti, Anna Ståhlbröst, and Mari Runardotter. Relevant social groups for open data use and engagement. **Government Information Quarterly**, 36 (1):98–111, 2019. ISSN 0740624X. doi: 10.1016/j.giq.2018.11.001. URL https://doi.org/10.1016/j.giq.2018.11.001.

Xiao Tong Li, Jun Zhai, Gui Fu Zheng, and Chang Feng Yuan. Quality assessment for open government data in China. In **ICIME**, pages 110–114, 2018. ISBN 9781450364898. doi: 10.1145/3285957.3285962.

Martin Lněnička, Renata Machova, Jolana Volejníková, Veronika Linhartová, Radka Knezackova, and Miloslav Hub. Enhancing transparency through open government data: the case of data portals and their features and capabilities. **Online Information Review**, 2021. ISSN 14684527. doi: 10.1108/OIR-05-2020-0204.

Devender Maheshwari and Marijn Janssen. Dashboards for supporting organizational development: Principles for the design and development of public sector performance dashboards. **Proceedings of the 8th international conference on theory and practice of electronic governance**, pages 178–185, 2014. doi: 10.1145/2691195.2691224.

Sébastien Martin, Muriel Foulonneau, and Slim Turki. Risk analysis to overcome barriers to open data. **Electronic Journal of e-Government**, 11(1):348–359, 2013.

Ricardo Matheus, Marijn Janssen, and Devender Maheshwari. Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. **Government Information Quarterly**, 37(3): 101284, 2020. ISSN 0740624X. doi: 10.1016/j.giq.2018.01.006. URL https://doi.org/10.1016/j.giq.2018.01.006.

Sehl Mellouli, Luis F. Luna-Reyes, and Jing Zhang. Smart government, citizen participation and open data. **Information Polity**, 19(1-2):1–4, 2014. ISSN 18758754. doi: 10.3233/IP-140334.

Ines Mergel. "A mandate for change": Diffusion of social media applications among federal departments and agencies. **Public Management Research Association Conference**, pages 1–27, 2011. URL https://www.maxwell.syr.edu/uploadedFiles/conferences/pmrc/Files/Mergel{_}final.pdf.

Chui Michael, Farrell Diana, and Jackson Kate. How government can promote open data and help unleash over $3 trillion in economic value. Technical report, McKinsey, 2014. URL https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-government-can-promote-open-data.

Petar Milic, Natasa Veljkovic, and Leonid Stoimenov. Comparative Analysis of Metadata Models on e-Government Open Data Platforms. **IEEE Transactions on Emerging Topics in Computing**, 9(1):119–130, 2021. ISSN 21686750. doi: 10.1109/TETC.2018.2815591.

Alaa M. Momani. The unified theory of acceptance and use of technology: A new approach in technology acceptance. **International Journal of Sociotechnology and Knowledge Development**, 12(3):79–98, jul 2020. doi: 10.4018/IJSKD.2020070105.

Benjamin Moreau, Nicolas Terpolilli, and Patricia Serrano-alvarado. A Semi-Automatic Tool for Linked Data Integration. **18th International Semantic Web Conference (ISWC2019)**, pages 4–8, 2019.

José Luis Moreno Cegarra, Juan Gabriel Cegarra Navarro, and José Rodrigo Córdoba Pachón. Applying the technology acceptance model to a Spanish City Hall. **International Journal of Information Management**, 34(4):437–445, 2014. ISSN 02684012. doi: 10.1016/j.ijinfomgt.2014.02.006. URL http://dx.doi.org/10.1016/j.ijinfomgt.2014.02.006.

Tamara Munzner. **Visualization Analysis and Design**. CRC Press Taylor and Francis Group, 2014. ISBN 9781466508934. doi: 10.1201/b17511.

Sebastian Neumaier, Jürgen Umbrich, and Axel Polleres. Automated quality assessment of metadata across open data portals. **Journal of Data and Information Quality**, 8(1):29, 2016. ISSN 19361963. doi: 10.1145/2964909.

Sebastian Neumaier, Lörinc Thurnay, Thomas J. Lampoltshammer, and Tomá Knap. Search, Filter, Fork, and Link Open Data: The ADEQUATe platform: Data- and community-driven quality improvements. **Web Conference 2018 - Companion World Wide Web Conference WWW 2018**, pages 1523–1526, 2018. doi: 10.1145/3184558.3191602.

Jakob Nielsen. Why You Only Need to Test with 5 Users, 2000. URL https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

Anastasija Nikiforova. Timeliness of Open Data in Open Government Data Portals through Pandemic-related Data: A long data way from the publisher to the user. **4th International Conference on Multimedia Computing, Networking and Applications**, pages 131–138, 2020. doi: 10.1109/MCNA50957.2020.9264298.

Anastasija Nikiforova. Open data hackathon as a tool for increased engagement of generation z: To hack or not to hack? In Fernando Ortiz-Rodríguez, Sanju Tiwari, Miguel-Angel Sicilia, and Anastasija Nikiforova, editors, **Electronic Governance with Emerging Technologies**, pages 161–175, Cham, 2022a. Springer Nature Switzerland. ISBN 978-3-031-22950-3.

Anastasija Nikiforova. Towards enrichment of the open government data: A stakeholder-centered determination of high-value data sets for latvia. In **ICEGOV '21**, ICEGOV '21, page 367–372, New York, NY, USA, 2022b. Association for Computing Machinery. ISBN 9781450390118. doi: 10.1145/3494193.3494243. URL https://doi.org/10.1145/3494193.3494243.

Anastasija Nikiforova and Martin Lnenicka. A multi-perspective knowledge-driven approach for analysis of the demand side of the Open Government Data portal. **Government Information Quarterly**, 38(4):101622, 2021. ISSN 0740624X. doi: 10.1016/j.giq.2021.101622. URL https://doi.org/10.1016/j.giq.2021.101622.

Anastasija Nikiforova, Nina Rizun, Magdalena Ciesielska, Charalampos Alexopoulos, and Andrea Miletič. Towards high-value datasets determination for data-driven development: a systematic literature review, 2023.

Javier Nogueras-Iso, Javier Lacasta, Manuel Antonio Urena-Camara, and Francisco Javier Ariza-Lopez. Quality of Metadata in Open Data Portals. **IEEE Access**, 9:60364–60382, 2021. ISSN 21693536. doi: 10.1109/ACCESS.2021.3073455.

Rajai Nuseibeh. What is a Chatbot?, 2018. URL https://chatbotsmagazine.com/what-is-a-chatbot-6dfff005bb34.

OECD. Engaging users: Promoting awareness and re-use of open government data. In **Government at a Glance 2019**. OECD Publishing, Paris, 2019.

Adegboyega Ojo and Marijn Janssen. Workshop about the understanding and improving the uptake and utilization open data. **15th Annual International Conference on Digital Government Research**, pages 350–351, 2014. doi: 10.1145/2612733.2612734.

Elena Ornig, Jolon Faichney, and Bela Stantic. Empirical Evaluation of Data Visualizations by Non-Expert Users. **International Journal on Advances in Software**, 10 (3):355–371, 2017.

Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. **Journal of Management Information Systems**, 24(3):45–77, 2007. ISSN 07421222. doi: 10.2753/MIS0742-1222240302.

Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. **12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008**, 2008. doi: 10.14236/ewic/ease2008.8.

Tomas Petricek. Tools for open, transparent and engaging storytelling. In **Companion to the first International Conference on the Art, Science and Engineering of Programming**, pages 1–2, 2017. ISBN 9781450348362. doi: 10.1145/3079368.3079382.

Donato Pirozzi and Vittorio Scarano. Support citizens in visualising open data. In **20th International Conference Information Visualisation**, pages 271–276. Institute of Electrical and Electronics Engineers Inc., aug 2016. ISBN 9781467389426. doi: 10.1109/IV.2016.45.

Axel Polleres, Jürgen Umbrich, Kathrin Figl, and Martin Beno. Perception of Key Barriers in Using and Publishing Open Data. **JeDEM - eJournal of eDemocracy and Open Government**, 9:134–165, 2017.

Simone Porreca, Francesco Leotta, Massimo Mecella, Stavros Vassos, and Tiziana Catarci. Accessing Government Open Data Through Chatbots. In **ICWE 2017**, volume LNCS 10544, pages 156–165. Springer International Publishing, 2018. ISBN 9783319744322. doi: 10.1007/978-3-319-74433-9_14. URL http://dx.doi.org/10.1007/978-3-319-74433-9{_}14.

Arie Purwanto, Anneke Zuiderwijk, and Marijn Janssen. Citizens' motivations for engaging in open data hackathons. In Panos Panagiotopoulos, Noella Edelmann, Olivier Glassey, Gianluca Misuraca, Peter Parycek, Thomas Lampoltshammer, and Barbara Re, editors, **Electronic Participation**, pages 130–141, Cham, 2019. Springer International Publishing. ISBN 978-3-030-27397-2.

Arie Purwanto, Anneke Zuiderwijk, and Marijn Janssen. Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. **Transforming Government: People, Process and Policy**, 14(1):1–30, apr 2020. ISSN 17506166. doi: 10.1108/TG-06-2019-0051.

Vigan Raca, Goran Velinov, Stefan Dzalev, and Margita Kon-Popovska. A Framework for Evaluation and Improvement of Open Government Data Quality: Application to the Western Balkans National Open Data Portals. **SAGE Open**, 12(2), 2022. ISSN 21582440. doi: 10.1177/21582440221104813.

Konrad Johannes Reiche and Edzard Hofig. Implementation of metadata quality metrics and application on public government data. In **IEEE 37th Annual Computer Software and Applications Conference Workshops**, pages 236–241, 2013. ISBN 9780769549873. doi: 10.1109/COMPSACW.2013.32.

Konrad Johannes Reiche, Edzard Höfig, and Schieferdecker Ina. Assessment and visualization of metadata quality for open government data. In **Proceedings of the 4th Conference for E-Democracy and Open Government (CeDEM 2014)**, pages 225–346, 2014.

Severino Ribecca. The Data Visualisation Catalogue, 2015. URL http://www.datavizcatalogue.com/.

Oliver C. Robinson. Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. **Qualitative Research in Psychology**, 11(1):25–41, 2014.

Jeff Rubin and Dana Chisnell. **Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests**. Wiley Publishing, Inc., 2008.

Igbal Safarov, Albert Meijer, and Stephan Grimmelikhuijsen. Utilization of open government data: A systematic literature review of types, conditions, effects and users. **Information Polity**, 22(1):1–24, 2017. ISSN 18758754. doi: 10.3233/IP-160012.

Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. What do we talk about when we talk about dashboards? **IEEE Transactions on Visualization and Computer Graphics**, 25(1):682–692, 2019. ISSN 19410506. doi: 10.1109/TVCG.2018.2864903.

Mark Saunders, Philip Lewis, and Adrian Thornhill. **Research Methods for Business Students (5th Ed.)**. Harlow: FT/Prentice Hall, harlow: ft edition, 2009.

Ashraf Shaharudin, Bastiaan van Loenen, and Marijn Janssen. Towards a Common Definition of Open Data Intermediaries. **Digital Government: Research and Practice**, 4(2):1–21, 2023. ISSN 2691-199X. doi: 10.1145/3585537.

Ben Shneiderman. Eyes have it: a task by data type taxonomy for information visualizations. **IEEE Symposium on Visual Languages, Proceedings**, pages 336–343, 1996. ISSN 10492615. doi: 10.1016/b978-155860915-0/50046-9.

Anthony Simonofski, Victor Amaral de Sousa, Antoine Clarinval, and Benoît Vanderose. Participation in Hackathons: A Multi-methods View on Motivators, Demotivators and Citizen Participation. **14th International Conference on Research Challenges in Information Science**, 385 LNBIP:229–246, sep 2020. doi: 10.1007/978-3-030-50316-1_14. URL https://link.springer.com/chapter/10.1007/978-3-030-50316-1{_}14.

Anthony Simonofski, Anneke Zuiderwijk, Antoine Clarinval, and Wafa Hammedi. Tailoring open government data portals for lay citizens: A gamification theory approach. **International Journal of Information Management**, 65(July 2021), 2022. ISSN 02684012. doi: 10.1016/j.ijinfomgt.2022.102511. URL https://doi.org/10.1016/j.ijinfomgt.2022.102511.

Elena Simperl and Johanna Walker. Analytical Report 8: The Future of Open Data Portals. Technical report, European Commission, 2020. URL https://data.europa.eu/sites/default/files/edp{_}analyticalreport{_}n8.pdf.

Janet M. Six and Ritch Macefield. How to Determine the Right Number of Participants for Usability Studies, 2016. URL https://www.uxmatters.com/mt/archives/2016/01/how-to-determine-the-right-number-of-participants-for-usability-studies.php.

Veronica S. Smith. Data dashboard as evaluation and research communication tool. In T. Azzam and S. Evergreen, editors, **Data visualization, part 2. New Directions for Evaluation**, pages 21–45. Wiley Online Library, 2013. doi: 10.1002/ev.

Tableau. 10 Best Practices for Building Effective Dashboards, 2021. URL https://www.tableau.com/learn/whitepapers/10-best-practices-building-effective-dashboards.

Maarja Toots, Keegan McBride, Tarmo Kalvet, and Robert Krimmer. Open data as enabler of public service co-creation: Exploring the drivers and barriers. **International Conference for E-Democracy and Open Government**, pages 102–112, 2017. doi: 10.1109/CeDEM.2017.12.

Usability.gov. User Interface Design Basics, 2021. URL https://www.usability.gov/what-and-why/user-interface-design.html.

UXPin. The Basic Principles of User Interface Design, 2021. URL https://www.uxpin.com/studio/blog/ui-design-principles/.

Vijay Vaishnavi and William Kuechler. Design science research methods and patterns: innovating information and communication technology. **Auerbach Publications**, 2007.

Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. User acceptance of information technology: Toward a unified view. **MIS Quarterly: Management Information Systems**, 27(3):425–478, 2003. doi: 10.2307/30036540.

Antonio Vetrò, Lorenzo Canova, Marco Torchiano, Camilo Orozco Minotas, Raimondo Iemma, and Federico Morando. Open data quality measurement framework: Definition and application to Open Government Data. **Government Information Quarterly**, 33(2):325–337, 2016. ISSN 0740624X. doi: 10.1016/j.giq.2016.02.001.

Di Wang, Deborah Richards, Ayse Aysin Bilgin, and Chuanfu Chen. Implementation of a conversational virtual assistant for open government data portal: Effects on citizens. **Journal of Information Science**, pages 1–21, 2023. ISSN 17416485. doi: 10.1177/01655515221151140.

Richard Y. Wang and Diane M. Strong. Beyond accuracy: What data quality means to data consumers. **Journal of Management Information Systems**, 12(4):5–33, 1996. ISSN 07421222. doi: 10.1080/07421222.1996.11518099.

Jane Webster and Richard T Watson. Analyzing the Past to Prepare for the Future: Writing a Literature Review. **MIS Quarterly**, 26(2):xiii – xxiii, 2002. ISSN 02767783. doi: 10.1.1.104.6570.

Lisa Wenige, Claus Stadler, Michael Martin, Richard Figura, Robert Sauter, and Christopher W. Frank. Open Data and the Status Quo – A Fine-Grained Evaluation Framework for Open Data Quality and an Analysis of Open Data portals in Germany. **ACM Journal of Data and Information Quality**, 1(1), 2021. URL http://arxiv.org/abs/2106.09590.

Claus O Wilke. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. O'Reilly Media, 2019. ISBN 1492031089.

Bernd W. Wirtz, Jan C. Weyerer, and Michael Rösch. Citizen and Open Government: An Empirical Analysis of Antecedents of Open Government Data. **International Journal of Public Administration**, 2018. ISSN 15324265. doi: 10.1080/01900692.2016.1263659.

Gareth W. Young and Rob Kitchin. Creating design guidelines for building city dashboards from a user's perspectives. **International Journal of Human Computer Studies**, 140(March 2019):102429, 2020. ISSN 10959300. doi: 10.1016/j.ijhcs.2020.102429. URL https://doi.org/10.1016/j.ijhcs.2020.102429.

Gareth W. Young, Rob Kitchin, and Jeneen Naji. Building City Dashboards for Different Types of Users. **Journal of Urban Technology**, 28(1-2):289–309, 2021. ISSN 14661853. doi: 10.1080/10630732.2020.1759994. URL https://doi.org/10.1080/10630732.2020.1759994.

Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. Quality Assessment for Linked Data: A Survey. **Semantic Web**, 1: 1–31, 2012.

Hui Zhang and Jianying Xiao. Quality assessment framework for open government data: Meta-synthesis of qualitative research, 2009-2019. **Electronic Library**, 38(2): 209–222, 2020. ISSN 02640473. doi: 10.1108/EL-06-2019-0145.

Anneke Zuiderwijk and Marijn Janssen. Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. **Open Government**, pages 115–135, 2014. doi: 10.1007/978-1-4614-9563-5_8.

Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks. Socio–technical Impediments of Open Data. **Electronic Journal of Electronic Government**, 10(2):156–172, 2012.

Anneke Zuiderwijk, Marijn Janssen, and Yogesh K. Dwivedi. Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. **Government Information Quarterly**, 32(4):429–440, 2015. ISSN 0740624X.