The Window Theory: Factors and Design Principles for Digital Transparency

Matheus, R.

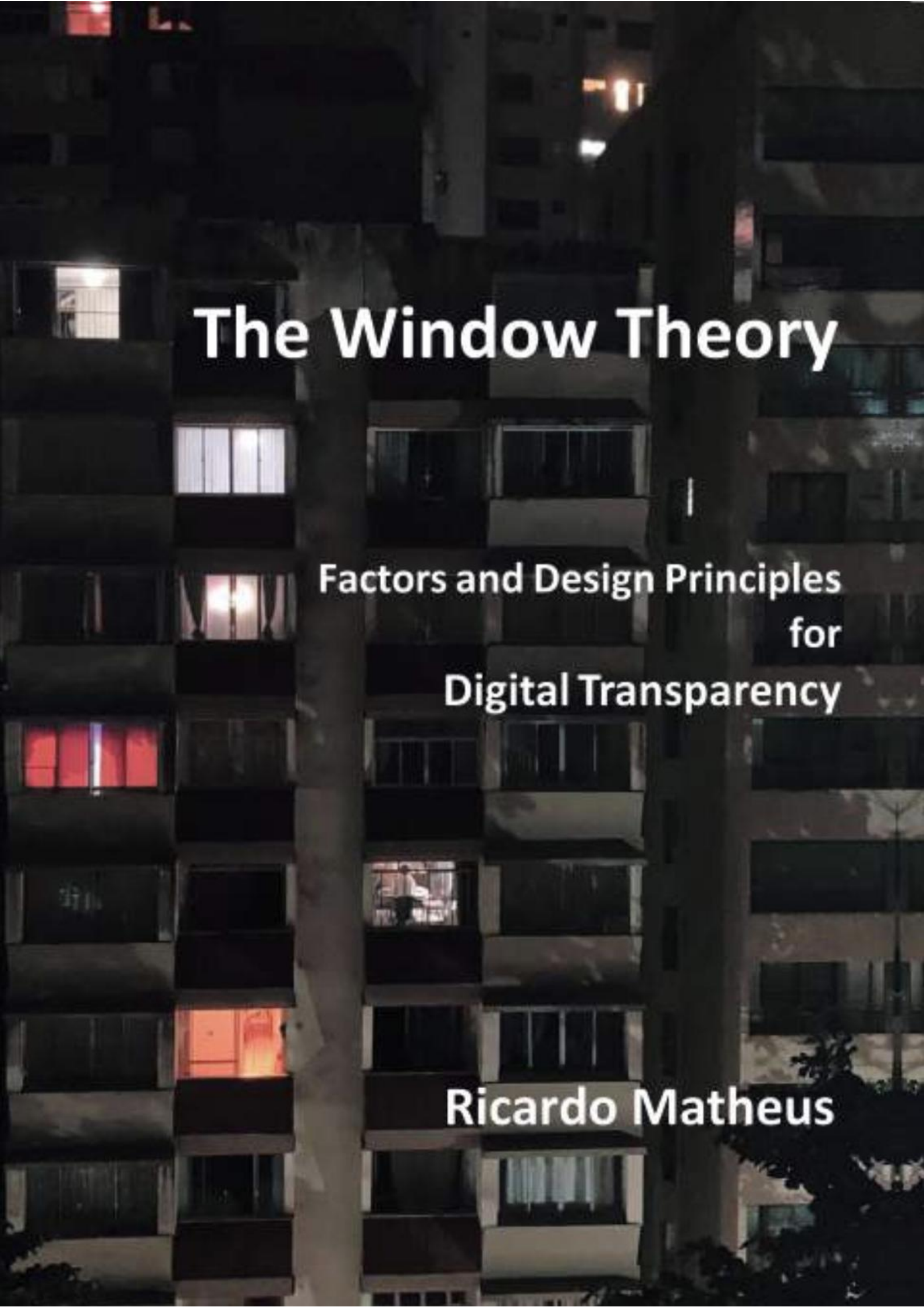
DOI
10.4233/uuid:bea85fde-fbaf-4302-9c6e-d0f469d1ac11

Publication date
2024

Document Version
Final published version

Citation (APA)
Matheus, R. (2024). *The Window Theory: Factors and Design Principles for Digital Transparency*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:bea85fde-fbaf-4302-9c6e-d0f469d1ac11

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.



The Window Theory

Factors and Design Principles for Digital Transparency

Ricardo Matheus

The Window Theory: Factors and Design Principles for Digital Transparency

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
10 September at 12:30 o'clock

by

Ricardo MATHEUS
Master of Science in Administration, University of São Paulo, Brazil
born in Salto, Brazil

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. Dr. Ir. M.F.W.H.A. Janssen	Delft University of Technology, promotor
Prof. Dr. T. Janowski	Gdańsk University of Technology, promotor

Independent members:

Prof. Dr. Ir. Nitesh Bharosa	Delft University of Technology
Prof. Dr. Hans de Bruijn	Delft University of Technology
Dr. Ir. Bastiaan van Loenen	Delft University of Technology
Prof. Dr. Yogesh Kumar Dwivedi	Swansea University
Prof. Dr. José Carlos Vaz	University of São Paulo

Keywords: Transparency; Open Government Data; Open data; Factors; Determinants; Design Principles; The Window Theory

ISBN: 978-94-93391-34-5

Printed by Proefschrift Specialist || https://www.proefschriftspecialist.nl/

ACKNOWLEDGEMENTS

I want to express my eternal gratitude to my supervisor Prof. Dr. Marijn Janssen. During the ten-year trip, Marijn was central to completing my Ph.D. Thanks to my co-supervisor, Tomasz Janowski, for all the valuable contributions when writing my papers and my personal life. A warm thank you to Prof. Dr. Yogesh for his kindness and generosity. Thanks to the Doctoral Committee Prof. Dr. Ir. Nitesh Bharosa, Prof. Dr. Hans de Bruijn, Dr. Ir. Bastiaan van Loenen, Prof. Dr. José Carlos Vaz and chairperson Prof. Dr. Ir. Bart Van Arem. Eternal appreciation to my family for all their support during my life. A huge thanks to my cousins, Prof. Dr. Oscar Della Pasqua, for bringing me to The Netherlands, and Prof. Msc. Carlos Della Pasqua.

This thesis has been written in the context of two European Union-funded projects. The first project within the Horizon 2020 framework is called OpenGovIntelligence (OGI – www.opengovintelligence.eu). The second project within the Erasmus+ framework is called CAP4CITY (www.cap4city.eu). The author and promotors thank the European Commission (EC) for funding and supporting both projects and this PhD thesis.

A special thanks to my friends at TU Delft, Dr. Dhata Praditya and Fernando Kleiman. I must also thank my Delft colleagues: Jolien, Mark, Aaron, Yao-Hua Tan, Jafar, and secretaries Laura Burns and Ellen Schwencke-Karlas. Still in Delft, my friends Wout Wijnands, Paula Schaling, Cees Kok, Channa Wijnands, and, Kirsten Wijnands. My best roommates Taco Dommerholt, Widodo, Stephanie, Louise, Charlotte and Angeline Rahardjo-Koelman (in memorial).

I also want to thank the people from my former lecturer position at Rotterdam School of Management: Prof. Dr. Ting Li, Dr. Rodrigo Bello, Dr. Dimitrios Tsekouras, and Yanick Kuper. Thanks to my OpenGovIntelligence (OGI) colleagues and my CAP4CITY colleagues. My eternal gratefulness to Carlinhos Cecconi, Juliano Cappi, Vinicius, Everton, Jean, Luiza, Paula Liebert, Vagner Diniz and Reinaldo Ferraz.

My appreciation to my former colleagues at Rio de Janeiro IBM Center of Operations: Pedro Arias, Pedro Bittencourt, Marcelo Granja, Dr. Sergio Bastos, Rafaela Bastos, Alexandre Caderman, George, Igor Krauss (in memorial due to COVID-19). My former colleagues at Salvador City Hall, Treasury Secretariat Giovanna Testa Victer, Sub-Secretariat Walter Cairo, Innovation Director MSc. Mauricio Correia, Treasury Director Ulysses Areas, CTO Jorge Ubiratan, Deise Marziane Johann Capelletti, Robson Oliveira and Javier Rodriguez. Kisses to

my Brazilian friends Dr. Manuella, Monica, Marina, Mariana, Dr. Luiz Pinheiro Junior, Dr Vanderlei Rovigatti, Murilo Lemos de Lemos, Allan, Renan, Pietra, Rogério, Dr. Milena Locci, Kátila Kormann Rafael Schroeder, Dr. Hérsio Massanori Iwamoto, Thiago Borges, Suellen, Marisa, Bruno and Miro Cazelatto (in memorial). My former colleagues at Blumenau City Hall: Governmental Management Secretary Paulo Costa, GDPR Data Controller Andre, Mayor Mário Hildebrandt, and Vice-mayor Maria Regina Soar.

A big thanks to my former supervisors Prof. Dr. José Carlos Vaz, Dr. Jaime Crozatti, Dr. Cesar Alexandre Souza, and Dr. Nicolau Reinhard, and my former professors Prof. Dr. Fernando Coelho, Dr. Marcelo Nerling, Dr. Hiro Sano, Dr. Alessandro Soares da Silva, Dr. Cristina Kerches, Dr. Eduardo Caldas, Dr. Flavia Mori, Dr. Martin Jayo, Dr. Ursula Peres, Dr. Wagner Iglecias, Dr. Isak Kruglianskas, Dr. Martinho Isnard Ribeiro de Almeida, Dr. Paulo Feldmann, and Dr. Rosa Maria Fischer.

To all of you, thanks!!!

Ricardo Matheus

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Summary

Transparency is important for scrutinizing government activities by enabling it to look inside the government. Opening-up data has been heralded for its ability to contribute to digital transparency by opening data about the functioning of government. Digital transparency is a way to create transparency using digital means, i.e., collecting, processing, and presenting data to provide insight into the functioning of government. In turn, transparency can improve accountability, participation, innovation, and corruption detection and deliver other benefits. The many barriers that open government data (OGD) initiatives encounter show that digital transparency is hard to implement. Data disclosure in Open Government Data Portals (OGDP) and the development of applications (apps) based upon such data will not create transparency *per se*. Digital transparency is more complex than often understood, and many factors influence its creation. The research aim pursued in this thesis *is to develop a set of design principles to deliver digital transparency using OGD-based applications*. The utilitarian aim is to help software designers create such applications for digital transparency.

This doctoral dissertation consists of four published papers aimed at helping software designers deliver digital transparency. In those papers, we look at various views on digital transparency as a concept distinct from transparency. Transparency is defined, interpreted, and measured in different ways. Most models of transparency do not consider the technical aspects and how data is released, or focus on the technical aspects only. In contrast, organizational and technical aspects are often viewed as significant barriers to creating transparency based on OGD.

The first paper conducts a systematic literature review (SLR) to develop a comprehensive model of determinants that enable or impede OGD-based transparency and their expected effects. Public Administration (PA) and Information Systems (IS) scientific literature were surveyed. The model, called *The Window Theory*, integrates a broad range of determinants and unifies existing models. In the model, the window provides a view inside the government, which uncovers its internal functioning. Different insights might be created depending on how, when, and who looks through the window. The window theory can be used to develop context-specific models that are comprehensive and parsimonious. The models can help designers determine relevant factors when delivering digital transparency for individual situations.

In the second paper, a model was developed to understand the factors that influence the usefulness of digital transparency. Using three case studies, the many factors identified in the

Window Theory were used to understand how digital transparency was created. The case studies were selected for their complexity and covering various elements, based on the expected effects of transparency: accountability, co-creation, and improved decision-making. The main factors that influenced transparency were identified for each case study. The case studies showed how the users struggled with complexity, e.g. different interpretations of the same data triggering discussions about the meaning of such data. Surprisingly, the case studies prioritized different factors, suggesting that the context and requirements matter for delivering transparency. Also, what constitutes digital transparency was found to be dependent on the context. The case studies were used to identify the most important factors for digital transparency and to arrive at a parsimonious model that was subsequently tested using a survey. The survey was distributed to the OGD application users in five countries. The results showed that the usefulness is influenced by perceived functionality, transparency, and efficiency. Surprisingly, the functionality and efficiency of the apps were found to be more important than transparency. This suggests that transparency is treated as the means, not the end. Usefulness can be created without high levels of transparency, as the public wants answers to their questions without consulting all kinds of data and applications.

The survey's results suggest the need for a balance between releasing raw data and creating apps. In the survey, using 218 responses, we found that even if the raw data is not easy to use, higher levels of transparency are possible by creating additional views similar to apps. Apps are easy to use but provide pre-defined views, whereas users might prefer different options. The results suggest that apps and OGDP raw data releases are needed to create an open government. The findings show that given a particular situation, different types of end-users and their objectives influence the most relevant determinants from the Window Theory, suggesting that digital transparency depends on the context and the desired stakeholder view. While apps can help reduce the effort of using OGD, they also provide a pre-determined view that can limit transparency. The contexts and stakeholders' needs determine the best method to deliver transparency.

The third paper focuses on the design of dashboards for creating digital transparency. Governments can use dashboards to interact with the public and support their decision-making and policy processes. Two smart city cases are investigated, showing that dashboards can improve transparency and accountability. However, realizing these benefits was cumbersome, and various risks and challenges were encountered in the process. These challenges can easily

produce misconceptions, wrong decision-making, and a blurred picture, thus resulting in less transparency, accountability, and, ultimately, less trust in government.

Dashboards should help citizens understand the situation and avoid tedious information searches and overload. Thus creating citizen dashboards requires a design focused on relevance and accounting for individual circumstances. Balancing issues, such as privacy protection and information overload, and designing an overview to create digital transparency at a glance is challenging. Although dashboards can have different shapes and forms, some users might want to gain access to the raw data. Due to the diversity of users and uses of data, it is paramount that dashboards have access to such data. Then dashboards could be used to communicate and stimulate interaction with the public. However, organizational changes must accompany the introduction and use of dashboards. The paper presents the first set of principles to guide the design of dashboards.

In the fourth and final paper, based on the literature review, case studies, and survey results, we derived design principles to help deliver digital transparency. In total, 16 principles were identified and validated through the case studies in three European countries. The most used design principles were related to privacy and metadata. The validation showed that many principles were important and easy to implement, including opening raw data with high levels of granularity, assigning stewardship, visualizing different views, and providing feedback mechanisms. A principle with high organizational impact and importance for digital transparency is transparency-by-design, i.e. transparency must be accounted for in every phase of the design process. Counter-intuitively, some efforts aimed at more transparency produce the opposite effect, e.g. creating an open data portal with many datasets, leading to information overload and difficulties in finding and selecting datasets, leading to less transparency.

This research shows that digital transparency is highly context- and stakeholder-dependent. Given the context and constellation of stakeholders, the Window Theory and design principles can help designers build applications with higher levels of digital transparency. The research identified many factors determining digital transparency and the principles for delivering digital transparency by government applications.

In future work, we recommend investigating situational and stakeholder views for digital transparency. We also recommend investigating the relationships between the factors in the Window Theory and the expected impact on digital transparency including accountability, anti-

corruption, innovation, decision-making, and public policy improvements. Last but not least, we recommend testing and refining the Window Theory in practice.

Samenvatting

Transparantie is belangrijk voor de controle op overheidsactiviteiten. Het openstellen van data wordt geprezen vanwege het vermogen om bij te dragen aan digitale transparantie. Digitale transparantie is een manier om met digitale middelen transparantie te creëren door het verzamelen, verwerken en presenteren van gegevens om inzicht te geven in het functioneren van de overheid. Op zijn beurt kan transparantie resulteren in voordelen zoals betere verantwoordingsplicht, participatie, het opsporen van corruptie en innovatie. De vele barrières die open overheidsdata (OGD) initiatieven tegenkomen, laten zien dat digitale transparantie moeilijk in de praktijk te brengen is. Het openbaar maken van gegevens in OGD Portals en de ontwikkeling van applicaties (apps) voor bepaalde doeleinden zullen op zichzelf niet tot transparantie leiden. Digitale transparantie is complexer dan vaak wordt gedacht, en veel factoren beïnvloeden het creëren van digitle transparantie. Het onderzoeksdoel van dit proefschrift is het ontwikkelen van een reeks ontwerpprincipes ter ondersteuning van het ontwerp van digitale transparantietoepassingen met behulp van OGD. Dit zou ontwerpers moeten helpen OGD-applicaties te creëren met een hogere mate van digitale transparantie.

Dit proefschrift bestaat uit 4 gepubliceerde artikelen die bedoeld zijn om softwareontwerpers te helpen de digitale transparantie te verbeteren. Transparantie is een dubbelzinnige term en kan op verschillende manieren worden geïnterpreteerd, wat tot uiting komt in de verschillende opvattingen over digitale transparantie in de vier artikelen. Er bestaat een verscheidenheid aan transparantiedefinities en -modellen met uiteenlopende standpunten en verschillende manieren om het niveau van transparantie te meten. De meeste modellen voor het verklaren van transparantie houden geen rekening met de technische aspecten en de manier waarop gegevens worden vrijgegeven, of richten zich alleen op de technische aspecten. Daarentegen worden organisatorische en technische aspecten vaak gezien als belangrijke belemmeringen voor het creëren van digitale transparantie vanuit OGD. Om van generieke transparantie te onderscheiden gebruiken we de term digital transparantie.

In het eerste artikel werd een systematisch literatuuronderzoek uitgevoerd om een alomvattend model te ontwikkelen van determinanten die transparantie voor OGD en de verwachte effecten ervan mogelijk maken of belemmeren. De wetenschappelijke literatuur van Bestuurskunde en Informatiesystemen werd onderzocht. Het alomvattende model kreeg de naam *The Window Theory*, dat bestaande modellen verenigt en een breed scala aan determinanten in één enkel

model integreert. Het raam biedt zicht op de overheid vanuit een bepaalde invalshoek. De naam is gekozen omdat er, afhankelijk van hoe, wanneer en wie er in een raam kijkt, een ander antwoord gevonden kan worden. De verwachting is dat de factoren per situatie verschillen. De Window Theory kan worden gebruikt als basis voor het ontwikkelen van contextafhankelijke modellen die tegelijkertijd alomvattend en spaarzaam zijn. Dit kan ontwerpers helpen bij het ontwikkelen van digitale transparantie naar de relevante factoren te kijken.

In het tweede artikel werd een model ontwikkeld om factoren te begrijpen die de bruikbaarheid van digitale transparantie beïnvloeden. De enorme hoeveelheid factoren die in de Window Theory werden geïdentificeerd, werd in drie verschillende casestudies gebruikt om te begrijpen hoe digitale transparantie werd gecreëerd. De casestudies zijn geselecteerd vanwege hun complexiteit en omdat ze een diversiteit aan elementen bestrijken, met name op basis van de verwachte effecten van transparantie: verantwoording, co-creatie en verbetering van de besluitvorming. Voor elke casestudie werden de belangrijkste factoren geïdentificeerd die van invloed waren op de transparantie. De casestudies laten de worsteling van gebruikers zien om met de complexiteit om te gaan om digital transparantie te creëren. Zo zijn er verschillende interpretaties van dezelfde gegevens mogelijk, waardoor er discussie ontstaat over de verschillende betekenissen. Verrassend genoeg bleken verschillende factoren belangrijk te zijn in de casestudies. Deze bevinding suggereert dat de context en de behoefte bepalen welke factoren de transparantie beïnvloeden. Ook bleek wat digitale transparantie inhoudt afhankelijk te zijn van de context. De cases zijn gebruikt om de belangrijkste factoren te identificeren die van invloed zijn op digitale transparantie en om te komen tot een spaarzaam model dat kan worden getest met behulp van een enquête.

Deze analyze van de factoren in de cases resulteerde in een model dat is getoetst door middel van een enquête. De enquête werd verspreid onder gebruikers van OGD-applicaties in 5 verschillende landen. De resultaten laten zien dat waargenomen functionaliteit, transparantie en efficiëntie het nut beïnvloeden, maar dat functionaliteit van apps en efficiëntie belangrijker zijn dan transparantie. Nut kan worden gecreëerd zonder een hoge mate van transparantie, omdat het publiek antwoorden op hun vragen wil en transparantie geen doel an sich is.

Uit onze onderzoeksresultaten blijkt dat er een balans nodig is tussen het vrijgeven van ruwe data en het maken van apps. Uit een onderzoek onder 218 respondenten hebben we ontdekt dat zelfs als ruwe gegevens niet eenvoudig te gebruiken zijn, deze een hoger niveau van

transparantie kunnen bieden doordat het openen van ruwe data het mogelijk maakt dat het publiek extra inzicht krijgt. Apps zijn eenvoudig te gebruiken, maar bieden een vooraf gedefinieerde weergave. Gebruikers geven de voorkeur aan verschillende opties, sommige geven de voorkeur aan ruwe data, terwijl anderen app als voorkeur hebben. De resultaten suggereren dat zowel apps als het vrijgeven van ruwe data in OGD nodig zijn om een open overheid te creëren. De bevindingen laten zien dat, gegeven een bepaalde situatie, verschillende soorten eindgebruikers en hun doelstellingen de meest relevante determinanten uit de Window Theory beïnvloeden, wat erop wijst dat dat digitale transparantie afhangt van de context en de gewenste stakeholderview. Apps kunnen de drempel voor het gebruik van OGD helpen verlagen, maar ze bieden ook een vooraf bepaald beeld dat de transparantie beperkt. De contexten en behoeften van belanghebbenden bepalen welke factoren de grootste invloed hebben op transparantie.

In het derde artikel ligt de nadruk op het ontwerp van dashboards voor het creëren van digitale transparantie. Dashboards kunnen door overheden worden gebruikt om hun besluitvormings- en beleidsprocessen te ondersteunen of om met het publiek te communiceren en interactie aan te gaan. Er worden twee smart city-cases onderzocht waaruit blijkt dat dashboards de transparantie en verantwoording kunnen verbeteren. Het realiseren van deze voordelen was echter omslachtig en bracht risico's en uitdagingen met zich mee. Deze uitdagingen kunnen gemakkelijk resulteren in misvattingen, verkeerde besluitvorming, het creëren van een wazig beeld, resulterend in minder transparantie en verantwoording, en uiteindelijk in nog minder vertrouwen in de overheid.

Dashboards moeten burgers helpen inzicht te krijgen in de situatie en langdurige zoekprocessen en een overdaad aan informatie te voorkomen. Om deze reden heeft het creëren van dashboards voor burgers een ontwerp nodig dat gericht is op relevantie en rekening houdt met de situatie van de burger. Het balanceren van zaken als privacy, informatie-overload en het ontwerpen van een overzicht in één oogopslag is een uitdaging. Hoewel dashboards op verschillende gebieden verschillende vormen kunnen hebben, kunnen er gebruikers zijn die toegang willen krijgen tot de onbewerkte gegevens. Vanwege de diversiteit aan mogelijke gebruikers van data is het van belang dat er verschillende dashboards kunnen worden samengesteld en dat toegang tot de ruwe data kan worden gegeven. De introductie van dashboards zou nutteloos kunnen zijn als de introductie ervan niet gepaard gaat met organisatorische veranderingen. Dashboards moeten niet alleen worden gebruikt om met het publiek te communiceren, maar ook om

feedback van hen te krijgen en interactie te stimuleren. Een eerste set van principes die als leidraad dienen voor het ontwerp van dashboards wordt gepresenteerd.

In het laatste artikel hebben we middels het literatuuronderzoek, casestudies en enquêteresultaten ontwerpprincipes afgeleid die ontwerpers kunnen helpen transparantie te creëren. In totaal zijn 16 principes geïdentificeerd en getest in drie reële scenario's binnen Europese landen. De meest gebruikte principes hadden betrekking op privacy en metadata. Uit de evaluatie blijkt dat veel principes belangrijk en gemakkelijk in de praktijk te brengen zijn, waaronder het openstellen van ruwe data met een hoge mate van granulariteit, het toekennen van stewardship, het visualiseren van verschillende standpunten en feedbackmechanismen. Een principe dat door de ondervraagden belangrijk werd gevonden voor het creëren van digitale transparantie en het hebben van een grote impact op de organisatie, was transparantie-by-design. Dit suggereert de noodzaak om transparantie in elke fase van het ontwerpproces op te nemen. Bovendien resulteerden sommige inspanningen gericht op het creëren van meer transparantie juist in minder transparantie. Het creëren van een open dataportaal met een groot aantal datasets kan bijvoorbeeld een informatie-overload voor gebruikers creëren, waardoor het moeilijk wordt om de juiste dataset te vinden en te selecteren, wat resulteert in minder digitale transparantie.

Uit dit onderzoek blijkt dat digitale transparantie in hoge mate afhankelijk is van de context en de behoefte van de betrokkenen. De Window Theorie and de ontwerpprincipes kunnen helpen softwareontwikkelaar helpen om tot een hogere mate van transparantie te komen. Het onderzoek helpt het grote aantal factoren en de complexiteit van digitale transparantie te begrijpen en biedt richtlijnen voor het ontwikkelen van digitale transparantie bij de overheid.

We bevelen verder onderzoek aan om de situatie en standpunten van belanghebbenden te onderzoeken die de digitale transparantie beïnvloeden. We raden ook aan om inzicht te krijgen in de relatie tussen de factoren in de Window Theorie en de verwachte effecten van digitale transparantie, zoals verantwoording, corruptiebestrijding, innovatie, besluitvorming en verbetering van het overheidsbeleid. Daarnaast raden wij aan om de Window Theorie in de praktijk te testen en te verfijnen.

1 Introduction

Open government data (OGD) has been heralded for creating transparency (J.C. Bertot, P.T. Jaeger, & J.M. Grimes, 2010). Transparency has been used as a tool to look into the functioning of public institutions. Many institutions launch transparency initiatives, including politicians and bureaucrats (executive branch), congressmen and parliamentary (legislative branch), members of the Judiciary, or the Civil Society (entrepreneurs, media, the public, etc.) (Curtin & Meijer, 2006). Transparency often represents a need for good governance and trustworthiness in modern times (Giddens, 2013).

Transparency is a widely used term by practitioners and academics (Zuiderwijk, Gascó, Parycek, & Janssen, 2014). However, different concepts of transparency are used, and misunderstandings of transparency have been identified (Bannister & Connolly, 2011a; Ricardo Matheus & Janssen, 2015a). What is transparent for one person might not be transparent for another person. In extreme cases, transparency has even been suggested as a magical concept for solving everything (Ward, 2014). Transparency can be understood as an output (product) of the public sector for monitoring and targeting performance and for enabling government innovation (De Bruijn, 2007, p. 9). A similar view on transparency as a type of output is given by Sunstein (2018, p. 1), where the public benefits from disclosing public sector information without having to make *Freedom of Information Act* (FOIA) requests.

Some studies consider transparency to be an effect of government digitalization, e.g., electronic government or e-gov for short (Bertot et al., 2010). Other studies consider transparency as a precondition for macroeconomic and fiscal sustainability (Craig & Kopits, 1998), or describe electronic processes for transparency to result in more efficient public service delivery and use (Scholl, 2005). Transparency is used in many domains for different purposes and can be created using digital technologies. In this thesis, we use the term *digital transparency* to refer to software-based applications aimed at looking inside the government by the public. Digital transparency is challenging, as can be expressed by many barriers found in the OGD literature (Barry & Bannister, 2014). Although much has been written about transparency, system designers and architects have no support for creating digital transparency using OGD.

This research aim*s to develop a set of design principles to support the design of digital transparency applications using OGD*. The basis for the design principles is the "*Window Theory*", which gives an overview of the factors influencing digital transparency and the

outcomes resulting from digital transparency. From this, principles are derived that guide designers in their work, and they are evaluated using pilots and a survey. This research has both scientific and practical relevance - it should help system designers and policy-makers create more transparent OGD portals and applications.

This chapter is structured as follows. In section 1.1, we define and discuss digital transparency. As our research is focused on design, section 1.2 defines and discusses the concept of transparency-by-design, which some view as an approach to achieving digital transparency. In section 1.3, we discuss principle-agent theory and information asymmetry, which are the key theoretical notions underlying transparency. The barriers and challenges for developing software applications that should result in digital transparency are presented in section 1.4. This provides the basis for the problem definition in the section 1.5. We end this chapter by providing an overview of the thesis in the section 1.6.

1.1 Digital Transparency and Open Government Data (OGD)

The driver of this research is the boost that digital transparency gained after the initial hype of the Open Data movement (Davies, 2010). Open Government Data (OGD) has been heralded to contribute to digital transparency. Yet, its actual contribution remains limited (Bannister & Connolly, 2011a; Ricardo Matheus & Janssen, 2015a). Several international initiatives have recently prompted milestones in the Open Data (OD), Open Government (OG), and OGD movements. The Obama Memorandum of Open Government and Transparency (Coglianese, 2009; Obama, 2009, 2013), the Public Sector Information (PSI) European Directive (Cretu & Manolea, 2013; K. Janssen, 2011; Manolea & Cretu, 2013), the Open Government Partnership (OGP), and the Tim Berners-Lee Open Data speech (Bizer, Heath, & Berners-Lee, 2009) were major initiatives driving the need to put open data into practice. The most visible results were the increasing number of *Open Government Data Portals* (OGDPs) in different countries: 75 countries established OGDPs between 2009 and 2024 (DataPortals.Org (2024) (See https://www.data.gov/open-gov/).

In Europe, the Directive on the re-use of public sector information (PSI Directive) was specifically drafted to address the needs of re-users (Van Loenen & Grothe, 2014) towards value-creating and sustainable open data ecosystems (Van Loenen et al., 2021). A benchmarking-driven analysis of open government data initiatives among European countries can be seen in the study of Lnenicka et al. (2024).

Besides that, exploring open government data ecosystems across data, information, and business is described by Fang et al. (2024). Moreover, an exploration of government officials' information behaviours in an open data policy implementation study was conducted by Yang and Wu (2021), and Ansari et al. (2022) aimed to identify how to enhance the usability and usefulness of open government data.

The creation of transparency by governments using digital means is commonly referred to as transparency (see section 3.1 for details). This thesis considers transparency using digital means to contribute to the public's insights into the government. In this thesis, we use digital transparency as a more appropriate term to describe and differentiate from in-person transparency (offline versus online transparency) (Shkabatur, 2013, p. 82).

This thesis focuses on creating digital transparency in governments using OGD (see sections 2.1 and 2.2 for details). The FOIA is out of the scope of this thesis. In the year 2024, there are more than 100 countries with FOIA Legislation, enabling citizens to query their governments (Ackerman & Sandoval-Ballesteros, 2006; Foerstel, 1999). Whereas OGD is about making available data for use in various forms, FOIA is about users requesting data from governments. FOIA can result in transparency for the requester, but the data might not be open to all. FOIA requests can open data for the public, but this might not always be the case.

The scientific literature and practitioners in charge of OGDPs show that the digital transparency concept is considered more as a background or an effect than the main focus of many studies (Tolbert & Mossberger, 2006). Rader, Cotter, and Cho (2018, p. 1) suggest that digital transparency is usually conceptualized as an outcome, and the mechanisms to achieve digital transparency are not explained. Some papers refer to enhancements of democracy and governance (Heeks, 2001), others to improvements in the quality of the government's social control (accountability) (Pina, Torres, & Acerete, 2007). This brings us to our first definition of digital government transparency, or digital transparency for short.

Definition 1: Digital government transparency refers to any initiative using digital means to contribute to the public's insights into the government.

Our definition takes a broad view of digital transparency. This closely matches the ideas behind open data, which is also a broad concept. We use the open data definition, operationalized by Ubaldi (2013, p. 5), as the data opened by the public sector with "*the lowest level of abstraction from which information and then knowledge are derived*".

In the United States of America (USA), transparency is considered as a "*sacred value*" to its democracy (Schudson, 2020, p. 5). Open Government (OG) and OGD policies are sometimes efforts from governments aimed at regaining the trust of the public (O'Hara, 2012). As such, governments have promoted digital transparency through the development of OGDPs, where a vast number of datasets are made available (Thorsby, Stowers, Wolslegel, & Tumbuan, 2017, p. 53). Some research investigates the effects of digital transparency. OGD is sometimes viewed as a way to create transparency, which should increase trust in government (Abelson, Gauvin, MacKinnon, & Watling, 2004, p. 2). However, this is debatable, as digital transparency can also result in less trust (Bannister & Connolly, 2011b).

The opening of data might not result in digital transparency per sé. Some authors criticize OGD and suggest that governments have been creating efforts of transparency in non-relevant areas and keeping secrecy (or less transparency) of critical areas such as "*full and accurate information on the nature, composition and workings of Expert Groups*" in the European Commission (Kierkegaard, 2009, p. 17). Hence, open data might not result in more transparency and less secrecy but in less transparency and more secrecy. Hence, the impact of OGD is debatable.

Data is often opened using centralized portals to provide a focus point for the public to find what they need. The simple disclosure of datasets in a centralized web portal, such as OGDPs, might not address the variety of citizens' and other stakeholders' information-seeking needs (Bertot, McDermott, & Smith, 2012). Liebwald (2015, p. 310) also concluded that transparency cannot be understood as simply "*providing online access to the vast volume of legal documents*".

Processes before the disclosure of data influence the level of digital transparency, as data might not be opened, pre-processed, or only partly opened (Ricardo Matheus & Janssen, 2015a). It is also necessary to understand the most common users of OGD and OGDPs and how they use the data and portals. This means that the design of OGD policies and OGDPs influences digital transparency. This thesis coined the term "*transparency-by-design*" to describe such design. Transparency-by-design is initially discussed in the section 1.2 and presented in the editorial co-authored by the author (Janssen, Matheus, Longo, & Weerakkody, 2017). Besides that, transparency-by-design was identified as a design principle to increase the level of digital transparency, presented in section 3.4 "Paper 4 – Design Principles for Creating Digital Transparency in Government".

Digital transparency is usually created in the form of dashboards to simplify and reduce the scope to specific audiences, with the same or similar databases behind each dashboard.

We also operationalize some concepts related to digital transparency, as follows.

The first concept is open government data. We use the definition provided by Attard, Orlandi, Scerri, & Auer (2015, p. 8): "*Open government data is a subset of Open Data, and is simply government-related data that is made open to the public*".

The second concept is open government. According to Hansson, Belkacem, & Ekenberg (2015, p. 4), "*the open government concept encompasses participatory aspects of government [..] more informed but also [..] data production [..] distributed to a diversity of actors both in the public and private sectors*", i.e., open government provides public access to data to create transparency and accountability. It encourages engagement between the government and the public. Our open government focus is on the creation of transparency.

The third concept is perceived usefulness. We use the definition by Davis (1989, p. 320): '*the degree to which a person believes that using a particular system would enhance his or her job performance'".*

The fourth concept is perceived transparency. We use the definition provided by Parris, Dapko, Arnold, & Arnold (2016, p. 24) that perceived transparency is "[organizational] *low effort opportunities for stakeholders to increase perceptions of organization transparency*".

The fifth concept is perceived efficiency. We used the definition provided by Phan & Ngu (2021, p. 4): perceived efficiency is "*related to personal analysis and judgment of a favorable adaptive outcome*".

The sixth concept is digital transparency. We use the definition provided by Matheus, Faber, Ismagilova, & Janssen (2023, p. 124): "*digital transparency as a stakeholder's ability to understand what is happening in the government using portals or apps".*

1.2 Transparency-by-Design

Even though transparency is a "*solution to the problems of lack of information*" (Fung, 2013, p. 184), the disclosure of data might not be sufficient. Fung, Graham, and Weil (2007) described examples using cases from the financial sector. An example is the Enron case from 2001. In this case, a North American company, Enron, misrepresented its earnings and modified its balance sheet to indicate favourable performance. The accounting firm Arthur Andersen failed to detect this fraud. Hence, the partial opening of data might not result in transparency and even in creating a biased view. However, the full disclosure of data might not be possible, costly, or may not be of interest to the organization in charge of disclosure. The latter is the case for banks aiming to provide minimal transparency to society to maintain their competitive position (Faust & Svensson, 2001, p. 371).

The Enron case exemplifies the need to simplify and balance the transparency of organizations. Therefore, concepts like *digital transparency-by-design* are emerging (Janssen, Matheus, et al., 2017), postulating that digital transparency should be embedded in the systems architecture design and improve open government data public policies (Saxena, 2017). Transparency-by-design suggests that instead of adding a layer of open data afterward, transparency should be included in the design efforts from the start (e.g., Janssen, Matheus, et al. (2017, p. 4)). Ideally, this should result in the automatic opening of data, simplifying the opening process without any additional cumbersome steps (e.g., Janssen, Matheus, et al. (2017, p. 4). This suggests that transparency should be considered a design principle throughout the project life cycle.

A review was conducted on Google Scholar and Scopus using the keywords "digital transparency" and "transparency-by-design" to identify scientific papers discussing the concept of transparency-by-design. This SLR showed that authors have been publishing about transparency-by-design since 2010 in a wide range of fields, such as Medicine and Biochemistry (Petyuk, Gatto, & Payne, 2019), Law (Liebwald, 2015, p. 312; Veale, Binns, & Ausloos, 2018), Business and Management (Castiglia & Turi, 2011), Education (McCormick, 2010), Urban Studies (Robbins & Henschke, 2017), and Computer Science (de Mingo & Cerrillo-i-Martínez, 2018; Hildebrandt, 2012, 2013; Janssen, Matheus, et al., 2017; Mascharka, Tran, Soklaski, & Majumdar, 2018; Saxena, 2017; Stoyanovich & Howe, 2018).

The overview shows a diversity of definitions for both digital transparency and transparency-by-design concepts. Some papers only define digital transparency or discuss the positive effects or challenges of digital transparency. Petyuk et al. (2019, p. 202) suggests that "*Public trust in scientific research is affected by the clarity of published conclusions and also the perceived transparency of the method*". Veale et al. (2018, p. 121) discussed the "*risk transparency fallacy*", pointing out that "*Oversight is unlikely to be useful if only provided at an individual level. Just as individuals suffer from consent fatigue, many of the solutions for the increasingly complex processing ecosystem today risk of a 'transparency fallacy', where the responsibility for obtaining and digesting complex information about computational systems falls, unhelpfully, on the data subject*".

Castiglia and Turi (2011, p. 124) provided a simple, utility-driven definition of digital transparency (together with accountability), pointing out that "*the goal of transparency and*

accountability is to enable stakeholders to obtain clear and relevant information about college and university performance". Robbins and Henschke (2017, p. 585) described the instrumental value of digital transparency, considering digital transparency as a governmental goal to "*enable ethical values which are lacking due to the informational deficits*". These authors also suggested that "*transparency can be used to ensure that privacy is not being overridden without justification and authorization and to assure citizens that there are appropriate policies in place*".

Some papers have definitions for both transparency and transparency-by-design concepts. Transparency-by-design can be defined as "*breaking a complex chain of reasoning into a series of smaller sub-problems, each of which can be solved independently and composed, is a powerful and intuitive means for reasoning*" (Hildebrandt, 2012, p. 4944). de Mingo and Cerrillo-i-Martínez (2018, p. 256), in their paper about records management, pointed out that transparency-by-design refers to "*the incorporation of transparency obligations into a record's lifecycle, from the moment it is created, to guarantee effective public access to public information, and to also provide a guarantee of records' integrity and their traceability to the original source*".

Liebwald (2015, p. 312) discussed the role of technology standards in integrating legal information systems. Liebwald (2015) concluded that "*to force the precision and rationality of law by introducing a 'transparency-by-design' approach, thus demanding transparency of law by design technology*". Stoyanovich and Howe (2018) concluded that transparency-by-design "*focuses solely on the final step of the data science lifecycle (called "analysis and validation"), and is limited by the assumption that input data sets are clean and reliable*".

Transparency-by-design should result in digital transparency. In the literature, some support for creating transparency-by-design can be found. The main characteristics of transparency-by-design are given by and compiled by Saxena (2017, p. 422), describing the Sri Lankan initiative:

1. *"Openness should be the guiding vision of any OGD initiative";*
2. *"Guidelines should be laid down for identifying the sensitivity of the data and the concomitant need for privacy";*

3. *"Insights of relevant actors involved in data should be taken into account during the publishing process";*
4. *"Data publishing process should be counted as a standardized and routine procedure", and;*
5. *"The process of data re-use should be monitored regularly".*

These characteristics allow organizations to achieve transparency-by-design and avoid the challenges described by Janssen, Matheus, et al. (2017):

- *"Late involvement of the government body in the publishing process";*
- *"Lack of guidelines for publishing data, especially with reference to the publication of sensitive data";*
- *"Lack of insight into the activities of other actors involved in the publishing process";*
- *"Different approaches to publishing data and a lack of focus on the outcomes of publishing the data sets".*

Digital applications (software) and business processes should be designed so that transparency becomes not an afterthought but is integrated into the heart of the public sector, allowing transparency by design. We used these characteristics to derive our definition:

Definition 2: Digital transparency-by-design influences and shapes all the steps of a data cycle, from data collection to data disclosure.

This SLR showed there is a gap connecting digital transparency and transparency-by-design. While some authors see the potential benefits of both, there is no literature discussing how transparency-by-design would help governments increase the level of digital transparency or how transparency-by-design could be a principle helping policy-makers and designers of OGDPs. These aspects are discussed in Paper 4, section 3.4.

The next section discusses some theoretical concepts related to digital transparency, e.g., agency theory and information asymmetry.

1.3 The Quest for Creating Digital Transparency

Transparency is aimed at overcoming the discrepancy in information positions between the public and the government, thus enabling the public to view what is happening within the government. Agency theory comprises the agency problem and its solution (Michael C. Jensen & Meckling, 1976, p. 309). Eisenhardt (1989, p. 58) explains that agency theory has one party (the principal) delegating work to another party (the agent) who performs that work. In our situation, the principal is the public, and the agent is the government. Agency theory is also known as the principal-agent model, which is schematically shown in Figure 1.

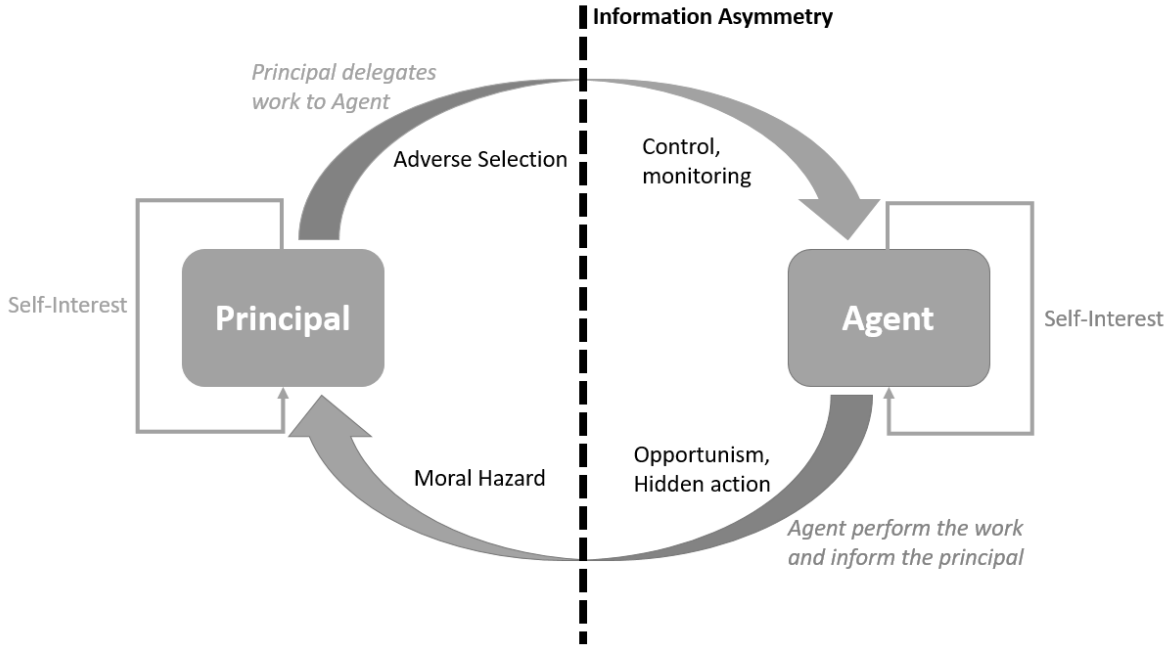


Figure 1 – Principal-agent model adapted from Snippert, Witteveen, Boes, and Voordijk (2015, p. 573)

The agency problem is as old as human civilization, when humans started to practice trade and business, aiming to maximize their interests (Panda & Leepsa, 2017). This self-interest action can lead to issues such as adverse selection and moral hazard, in particular, due to the information asymmetry between principal and agent. *Adverse selection* refers to the "misrepresentation of ability by the agent" (Eisenhardt, 1989, p. 61). An example, given by

Eisenhardt (1989, p. 61), is that adverse selection can happen when the principal cannot wholly verify specific characteristics of someone or a product. A typical example is buying a used car. If you are not a car specialist, you might ignore issues that might impact the costs and quality of the car in the future.

Moral hazard refers to the "*lack of effort on the part of the agent*" (Eisenhardt, 1989, p. 61). An example given by Eisenhardt (1989, p. 61), a moral hazard happens when a scientist works on a research project, but the research is so complex that whoever is in charge of monitoring and controlling the scientist cannot detect what the scientist is doing.

Many examples of adverse selection and moral hazards influence transparency (Klein, Lambertz, & Stahl, 2016). Hence, it is difficult to make decisions when there is an imbalance of information between the principal and the agent. This imbalance is called "Information asymmetry" in agency theory. Information Asymmetry refers to a situation in which one party has more information than another party (Michael C. Jensen & Meckling, 1976).

Agency theory is used in many fields, such as accountability and internal audit (Adams, 1994), healthcare and life insurance industry (Pottier & Sommer, 1997), and finance (Michael C Jensen & Smith, 2000). In the government, agency theory was employed to solve issues such as information asymmetry between vendors and government contracting (procurement) (Amagoh, 2009, p. 2), food safety and food quality (Hobbs, 2004, p. 397), healthcare regulation (Bloom, Standing, & Lloyd, 2008, p. 2083), and improving the relationship between public sector universities and industries (Abramo, D'Angelo, Di Costa, & Solazzi, 2011, p. 85).

Reducing information asymmetry between the government (agent) and the public (principal) can improve trustworthiness (Abelson, Gauvin, MacKinnon, & Watling, 2004) and governance (B. G. Peters & Pierre, 1998). However, when poorly designed or implemented, such reduction can cause excessive bureaucratic paperwork that might "hinder *monitoring activity*" of the public sector (Attila, 2012, p. 711). For OGD, this suggests that governments (agents) might reduce the level of transparency to the public when dumping many datasets to the parties monitoring them (principal), as explained by (Attila, 2012, p. 711).

The parties involved in digital transparency can vary, depending on the situation. As an example of the Chilean OGD ecosystem described by Gonzalez-Zapata and Heeks (2015, p.

447), digital transparency stakeholders are primary (those who have a formal relationship with OGD) and secondary (those who have an informal relationship with OGD).

The primary stakeholders are politicians, public officials, public sector practitioners, and international organizations. The secondary stakeholders are civil society activists, funding donors, ICT providers, and academics. Gonzalez-Zapata and Heeks (2015, p. 447) analyzed the power and interests of all these stakeholders. The power position can be based on the information available. The more information, the more powerful a stakeholder can be.

This is similar to information asymmetry, which found that the gap between the information position of the principal and agent needs to be bridged. In a similar vein, OGD is aimed at bridging the gap between the government and the public. The next section 1.4 shows the barriers and challenges of developing software applications (apps) to create digital transparency and why creating apps for digital transparency is difficult.

1.4 The Limitations of Digital Transparency: Barriers and Challenges for Digital Transparency

Apps and OGD portals have been created so the public can view the government's functioning. Many different barriers and challenges are found, which result in digital transparency not delivering on the initial expectation (B. Worthy, 2010). Some authors call this “opacity” or “secrecy” (Birchall, 2011; Fenster, 2005). Although transparency looks appealing and straightforward, in practice, it is more challenging to achieve (Bertot et al., 2010). An underlying question is what the public wants to see and what type of data is provided. The creation of digital transparency encounters all kinds of challenges and risks.

Although more and more OGD is being released, there is a gap between the expected effects of OGD and the current reality (Barry & Bannister, 2014). The literature identified many barriers to open data use. For example, Barry and Bannister (2014) identified technical, cultural, legal, administrative, and risk-related barriers. These barriers are diverse and range from the technical aspects of data, such as metadata (Strathern, 2000) and system characteristics such as usability (Bertot, Gorham, Jaeger, Sarin, & Choi, 2014) to internal resistance to transparency (Navarro-Galera, Alcaraz-Quiles, & Ortiz-Rodríguez, 2016). Cultural and legal barriers are out of the scope of this thesis.

Digital transparency is dependent on the data quality (Sivarajah, Kamal, Irani, & Weerakkody, 2017), the way the data is processed and visualized (Alharthi, Krotov, & Bowman, 2017), and the public using the data (Mol, 2010). To lower the usage barrier, simple apps are often developed to create digital transparency. These apps often give a predefined view to the public. This provides insights from a single perspective, whereas others might have provided different or complementary insights. For example, if only budget information is shared, then no insight is gained into resource utilization. The insight needed is often difficult to determine (Cahlikova & Mabillard, 2019). As such, the quality of data and what is shared in an easy-to-access manner influences digital transparency, among many other factors.

Users play a crucial role in creating digital transparency, and their skills, experience, and understanding might be barriers to it. Sometimes, the mere opening of data is viewed as a form of transparency, whereas transparency requires the public to understand and “interpret the data” (Zyl, 2014, p. 347). Mol (2010, p. 136) pointed out that transparency only happens when certain conditions are met. How transparency is designed is essential, but the stakeholders' use

of information should be given attention. Users might have different digital skills and understanding of the situation at hand. Using the same data, they might arrive at different interpretations. Which views are appropriate for creating transparency depends on the needs and might change over time.

Apart from the challenges, digital transparency might encounter risks. Abu-Shanab (2015) identified an increased chance of violating privacy when publishing public sector data. Such publishing is difficult since there is no consensus about what constitutes privacy or the trade-offs between transparency and data protection (Conradie and Choenni (2014, p. 516). R. Meijer, Conradie, and Choenni (2014) and Janssen and van den Hoven (2015) found "significant barriers to release open data, such as privacy, opaque ownership or judicial issues" when searching for barriers in real-life case studies. In line with the "radical transparency" concept, Mergel (2012, pp. 285-286) pointed out the possibility of overexposing the government if all the internal procedures were transparent. The design of open government portals and apps is hindered by a lack of knowledge of which factors influence transparency.

The various challenges and risks have been given much attention in the literature (Barry & Bannister, 2014). However, factors influencing transparency have been given less attention. Robbins and Henschke (2017, p. 585) argue that we must understand how these factors influence digital transparency positively or negatively. Hence, such factors should be identified when designing digital transparency.

1.5 Problem Definition and Research Objective

The multi-faceted barriers presented in the section 1.4 make the achievement of digital transparency challenging. It is hard for app and portal designers to determine how digital transparency can be created. However, a myriad of examples of digital transparency exist. They range from regular citizens monitoring the pollution in their neighbourhood to computer scientists creating fraud detection software using artificial intelligence (see project Serenata de Amor https://github.com/okfn-brasil/serenata-de-amor). More than 600 open data portals were identified by DataPortals.Org (2024). They are based on a range of platforms, use a variety of datasets, and are aimed at different user bases. This diversity calls for open data apps and portals to be built upon design principles that ensure digital transparency by design. Such principles and the influencing factors can drive the creation of OGDPs for digital transparency.

With this in mind, **the research objective of this thesis is to develop design principles to support the design of OGD-based portals and applications for digital transparency**.

As a utilitarian objective, this thesis aims to support website designers and policymakers in creating such applications.

1.6 Ph.D. Dissertation Outline

This dissertation is structured into four chapters.

The first chapter introduces the dissertation; it presents the research topics and theories around them and derives the research problem and objectives.

The second chapter describes the research approach, including the research objective, research questions, research philosophy, research phases, and the published papers that comprise this dissertation, including an overview of the research methods and findings.

The third chapter contains all papers published and included in this dissertation. The papers are slightly modified to match the style and format needs of the dissertation.

The fourth chapter discusses the answers to the research questions. It also contrasts the principle-based vs. rule-based approaches, discusses why full transparency cannot be realized in practice, and anticipates the future of transparency. Furthermore, the chapter presents research limitations and recommendations for future research.

Figure 2 contains an overview of the chapters, research questions, methods, and findings. The next chapter explains the papers and their research methods and findings.

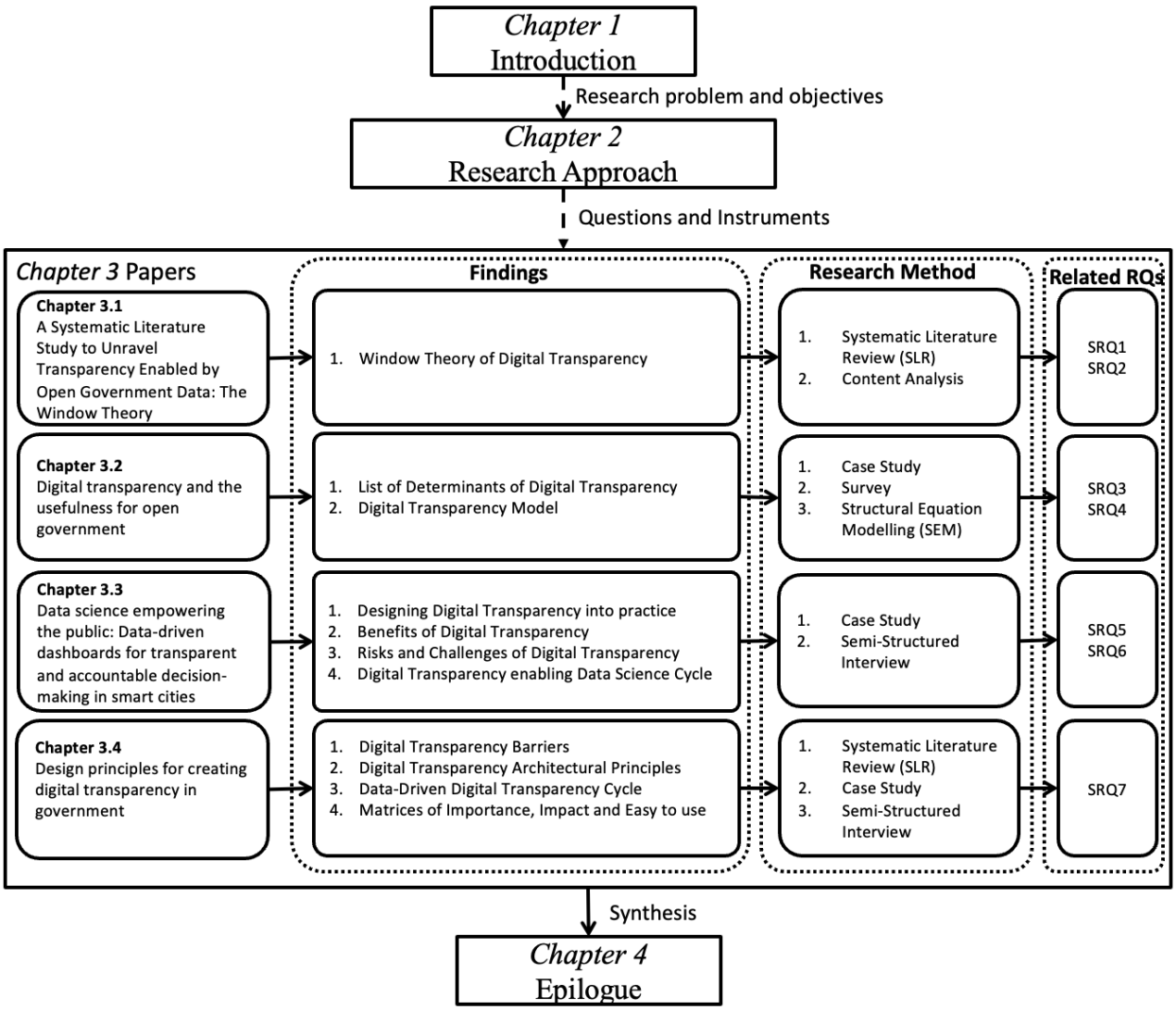


Figure 2 – PhD thesis outline

2 Research Approach

This chapter describes the research objective, questions, philosophy, and strategy adopted by this dissertation and an overview of the research methods adopted by the four included papers. To derive design principles, the dissertation follows a design science research approach. The rest of the chapter is structured as follows. The research objectives are presented in section 2.1, followed by the research questions in the section 2.2, and the research philosophy and strategy in the section 2.3. After that, the research methods and findings of the papers that make up this dissertation are presented in the sections 2.4 and 2.5 respectively.

2.1 Research Objective

More and more OGD is opened all over the world. Data from local, regional, and national governments can potentially be used to deliver digital transparency. Despite such efforts, no design support for delivering digital transparency exists. This thesis aims to develop design principles to support the design of OGD-based portals and applications for digital transparency. The designers of such portals and applications can use the principles to deliver digital transparency, and policy-makers can use them to improve government transparency in general.

2.2 Research Questions

The dominant approach for delivering digital transparency in government is OGD-based portals and apps. However, there is an ongoing discussion about whether and how such portals and apps deliver transparency. Addressing the research objective of this thesis – to develop design principles to support the design of OGD-based portals and applications for digital transparency – should help designers and policymakers deliver more transparency through OGD-based portals and apps. To address the research objective, we formulate five research questions that will be answered in the remainder of this dissertation.

RQ1 – What is digital transparency?

Digital transparency is an intuitively appealing concept, but many different definitions exist. To answer this RQ, the literature was reviewed, existing definitions and conceptualizations were identified, and the definition to serve the needs of this research was identified.

RQ2 – Which factors influence the creation of digital transparency?

Digital transparency has no clear framework to be used as a theoretical and practical model for implementing OGD-based portals and apps. This prompted the identification of the factors that influence digital transparency by conducting a systematic literature review. The review created the Window Theory, which contains all the factors found in the literature.

RQ3 – What types of mechanisms result in digital transparency?

Three case studies were analysed to understand how digital transparency is created using OGD. The factors identified in RQ2 were used to investigate the cases and to understand the mechanism for creating transparency. Indeed, the level of transparency created varies per case study. However, the cases provided deep insight into what users consider as transparency and which mechanisms worked for them to create transparency. The case studies show the importance of considering the context and the users, as the context determines what is considered transparent and which factors are of influence.

RQ4 – What are the factors influencing the usefulness of digital transparency?

The case studies provided insights into the mechanisms and factors influencing transparency. They also showed that the outcomes differ per case. To generalize the factors found in the case studies, a parsimonious model was created that could be tested through the case studies, and a survey was administered among the end-users of OGD-based portals and applications.

RQ5 – What are the risks and benefits when creating transparency dashboards?

Dashboards can improve transparency and accountability. However, various risks and challenges are encountered, and realizing these benefits is cumbersome. Challenges include insufficient data quality, lack of data understanding, poor analysis, wrong interpretation, confusion about the outcomes, and imposing a pre-defined view. These challenges can easily result in misconceptions, wrong decision-making, and a blurred picture, resulting in less transparency, accountability, and, ultimately, less trust in government.

RQ6 – What are the design principles for transparency dashboards?

Design principles guiding the design of transparency dashboards help solve ill-structured or ‘complex’ problems, including digital transparency. Design principles can also guide the design of more effective public sector dashboards, complemented by citizen engagement, data interpretation, governance, institutional arrangements, and other mechanisms.

RQ7 – What are the design principles for creating digital transparency?

Digital transparency is a multi-dimensional concept that is hard to put into practice. Based on the factors that influence the level of transparency, design principles are developed to guide developers in creating portals and apps that deliver transparency depending on the scenario characteristics and the expected effects.

The research questions are answered by four research papers that comprise this dissertation. The mapping of the papers and research questions is shown in Table 1.

Table 1 – Papers and research questions addressed

#	Paper Title	Paper URL	Research questions (RQ) addressed
1	A systematic literature study to unravel transparency enabled by open government data: The Window Theory	Paper 1	RQ1 – What is digital transparency? RQ2 – Which factors influence the creation of digital transparency?
2	Digital transparency and the usefulness for open government	Paper 2	RQ3 – What types of mechanisms result in digital transparency? RQ4 – What are the factors influencing the usefulness of digital transparency?
3	Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities	Paper 3	RQ5 – What are the risks and benefits of creating digital transparency? RQ6 – What are the design principles for Dashboards?
4	Design principles for creating digital transparency in government	Paper 4	RQ7 – What are the design principles for creating digital transparency?

2.3 Research Philosophy and Strategy

After deciding on the topic, a researcher faces several important decisions that shape the research, considering the selected topic and contextual conditions, such as data and infrastructure available, department colleagues, and funding. Such decisions can be labelled as a paradigm (Chilisa & Kawulich, 2012; Guba & Lincoln, 1994), i.e., "a summary of assumptions, beliefs, and values that will guide the scientific research" (Chilisa & Kawulich, 2012, p. 1). Since the paradigm describes the researcher's point of view, some important beliefs that make up a research paradigm are derived (Chilisa & Kawulich, 2012, p. 1):

1. *Ontology* refers to the philosophical assumption about the nature of the studied reality.
2. *Epistemology* is the way we justify knowledge in logical discourse.
3. *Axiology* is the ethics and values considered in our system.

The beliefs that constitute this research's chosen paradigm are summarized in Figure 3.

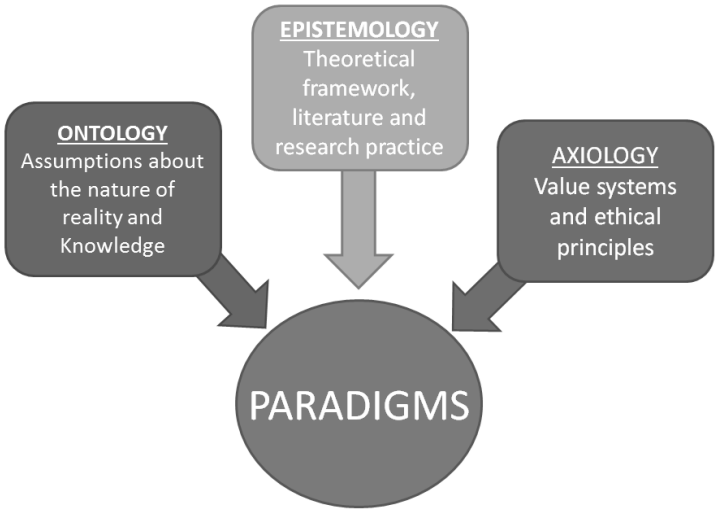


Figure 3 – Basic beliefs influencing the choice of a research paradigm adapted from Chilisa and Kawulich (2012, p. 1)

The *research methodology* is the approach used to systematically inquire the researcher's point of view, combining ontology, epistemology, and axiology (Chilisa & Kawulich, 2012, p. 1). The methodology is described in the section 2.4.

The selection of ontology, epistemology, and axiology determine the boundaries of knowledge and research limits. This research is focused on the factors influencing transparency and the ability to design applications to create transparency. Before selecting our research philosophy, we compare the post-positivist (positivist) and constructivist (interpretative) paradigms.

The *post-positivist paradigm* can be distinguished from positivism according to whether the focus is on theory verification (positivism) or theory falsification (post-positivism) (Lincoln, Lynham, & Guba, 2011, p. 168; see table 6.3). Positivism sees everything as perfect; it is influenced by the philosophy of August Comte, where "*true knowledge is based on the experience of senses and can be obtained by observation and experiment*" (Dash, 2005, p. 1). Post-positivism sees the imperfection of the reality (Lincoln et al., 2011, p. 168 see table 6.3).

The post-positivist paradigm emphasizes the absolute certainty of probability recognized by the positivist paradigm. Guba and Lincoln (1994) explain the positivist paradigm using the example that one million white swans cannot prove that all swans are white. Only one black swan can prove the contrary. It means that scientists construct knowledge instead of identifying and describing the laws of nature. Chilisa and Kawulich (2012) describe the post-positivist paradigm as a less strict form of positivism, in which, no matter how scientists adhere to the scientific method, the research outcomes are not totally objective or cannot be challenged. The authors suggest that post-positivists share common understandings with positivists. However, the current social science practices and research approach match the post-positivist paradigm.

The *constructivist* paradigm understands the world as other people experience it. According to Chilisa and Kawulich (2012), constructivists differ from positivists in their assumptions about the nature of reality, such as knowledge and its sources, and values and their role in the research process. The Constructivist Paradigm is connected with Edmund Husserl, Willhem Dilthey, Martin Heidegger and Max Weber, the scientists who studied human consciousness and self-awareness (Chilisa & Kawulich, 2012, p. 9).

The ontology in the post-positivist paradigm is characterized by only one single point of view (reality) that is constant across time and context (Chilisa & Kawulich, 2012, p. 8). It means the reality is objective and independent of the researcher's interest. The object can be measured and can be broken into variables. However, unlike positivists, post-positivists understand that reality exists. However, it is imperfect due to human limitations such as information asymmetry or lack of proper tools to measure certain phenomena (Chilisa & Kawulich, 2012, p. 9). The

ontology in the Constructivist Paradigm is based on the belief that reality is a social construction and there are as many realities as people constructing them. It means reality will be influenced by social aspects such as cultural background, personality, and mindset, e.g., scientific background, social sciences, and hard sciences. Chilisa and Kawulich (2012, p. 10) suggest that constructivist understanding refers to "*reality is limited to context, space, time and individuals or group in a given situation and cannot be generalized into one common reality*".

The epistemology in the post-positivist paradigm is defined as "not achieved but approached" (Chilisa & Kawulich, 2012, p. 9). It differs from positivists because positivists believe that statements can be tested empirically, verified, confirmed, or disconfirmed with a stable generalization, identifying laws and principles that govern the universe (Chilisa & Kawulich, 2012, p. 9). As an example, if any research has the same data and tools, they will arrive at the same results (Chilisa & Kawulich, 2012, p. 9). The epistemology of the constructivist paradigm considers that knowledge is subjective, socially constructed, and mind-dependent. Each human experience will bring different versions of the truth. Within a context, humans can state what is true or false without a high level of generalization (Chilisa & Kawulich, 2012, p. 10).

The axiology in the post-positivist paradigm believes that theories, hypotheses, and background knowledge can strongly influence the researcher and subject of study. Positivism demands all inquiries to be value-free. Researchers should use scientific methods to gather data and, neutrally, achieve objectivity. The axiology in the constructivist paradigm considers reality as a social construction. Constructivists believe that researchers are influenced by their values, influencing the paradigm (topic selection, methods to collect and analyse data).

The Methodology in the post-positivist paradigm aims to predict results, test a theory identifying cause-and-effect relationships, and find the strength of relationships between variables. Research approaches include quantitative, experimental, quasi-experimental, correlational, causal-comparative, and survey designs. The data collection techniques usually are questionnaires, observations, tests, and experiments. The methodology in the constructivist paradigm tries to understand people's experiences. The methodology is influenced by the assumption of the existence of multi-realities based on the human experience and context. Common methodologies are ethnography, phenomenology, biography, case study, and grounded theory. Data gathering techniques depend on the design, respondents, and problem,

but usually, they are interviews, observations, visual aids, personal and official documents, photographs, drawings, information conversations, and artifacts (Chilisa & Kawulich, 2012).

The basic beliefs and assumptions from the post-positivist and constructivist paradigms are summarized in Table 2.

Table 2 – Comparison of Post-Positivist and Constructivist Paradigms adapted from Chilisa and Kawulich (2012) and Guba and Lincoln (1994)

	Post-Positivist Paradigm	**Constructivist Paradigm**
Reason for doing the research	Discover laws that are generalizable and govern the universe	To understand and describe human nature
Ontological assumptions	One reality, based on probability	Multiple realities based on human experience (values, culture, etc.)
Goodness or quality criteria	Conventional benchmarking of scientific "rigor". Internal and external validity, reliability, and objectivity	Trustworthiness and authenticity, and misapprehensions
Ethics	Extrinsic, tilt toward deception	Intrinsic process tilt toward revelation, identifying particular problems
Nature of knowledge (Epistemology)	Objective	Subjective
What counts as truth (Axiology)	Based on precise observation and measurement of what is verifiable	Truth is context-dependent
Methodology	Mainly quantitative, correlational, quasi-experimental, comparative studies and surveys	Mainly qualitative. phenomenology, ethnographic, and symbolic interaction
Techniques of gathering data	Mainly questionnaires, observations, tests, and experiments	Mainly interviews, participant observation, and documents.

2.3.1 Research Philosophy in this Dissertation

To select the research philosophy and strategy, we used the method suggested by Dash (2005, p. 1 website), who provides a list of questions that researchers should ask themselves when looking for the paradigm and methodology:

1. *"What is the nature or essence of the social phenomena being investigated?"*
2. *"Are social phenomena objective in nature or created by the human mind?"*
3. *"What are the bases of knowledge corresponding to the social reality, and how can knowledge be acquired and disseminated?"*
4. *"What is the relationship of an individual with her environment? Is she conditioned by the environment, or is the environment created by her?"*

This research takes a post-positivism philosophy because of the assumption that our transparency model rules the world. The transparency model was created after a literature review. The model has 72 factors influencing transparency and eight expected transparency effects (see section 3.1). However, this research also assumes that no full transparency is possible and that transparency is context-dependent and in the eye of the stakeholder. The transparency model will always need to be updated in accordance with the context. For example, specific profiles or cultures may desire a particular type of technology to access data, e.g., developers *vs.* citizens and Western *vs.* Eastern civilizations. Stakeholders might have different needs for transparency, and what is transparent for one might be opaque for another.

Besides cultural and personal influences, the available technologies will evolve in the following years, and how data can be presented will be changed by the novelty of ICTs. In the past 30 years, humankind has witnessed the obsolescence of CD-ROM, DVD-ROM, and blue-ray discs. Even pen drives and hard drives are becoming obsolete since the majority of the data is stored in the cloud, and Internet access has become almost ubiquitous. Internet of Things (IoT) and Big Data, Artificial Intelligence (AI) and Machine Learning (ML) algorithms will influence the level of transparency. IoT and Big Data are about collecting and storing data, while AI and ML are about processing data.

The model of transparency presented by Ricardo Matheus and Janssen (2013, p. 167) presents dimensions that influence transparency apart from the factors. Further, transparency is also a sense-making concept. What may be transparent to someone may be opaque to somebody else, and vice-versa. Thus, the survey performed in Paper 2 (section 3.2) is the constructivist part of this research. Considering the transparency effect of transparency portals using OGD, let us reconstruct the model we created in Paper 1 (sections 3.1) and Paper 2 (section 3.2).

2.4 Design Science Research Approach

This research uses a *design science research* (DSR) approach to address a practical problem, create a real-world solution and contribute to theory development. There is a potential tangible impact on industries and society. However, there are many misconceptions about what constitutes design science. To clarify design science, we follow Baskerville (2008) by explaining what characteristics do not belong to design science.

First, design science is not only design but "*is more than design alone*, [where] *information systems is one problem arena of many possible arenas*" (Baskerville, 2008, p. 1).

Second, design science is not design theory because design theory might have "*precise components, (...) to be grounded (...), or have to simply intertwine design and development principles*" (Baskerville, 2008, p. 1).

Third, design science is not an IT artifact; it is "*problem-driven and leads to an artifact that solves the problem* [and] *the IT artifact alone is not design science*" (Baskerville, 2008, p. 1).

Fourth, design science is not a methodology because "*methods are usually thought to involve pre-defined processes, or methodical ways of 'doing' things* [although it is] *possible to 'do' design science in methodical ways*" (Baskerville, 2008, p. 1).

Fifth, design science is not action research because "*action research is a methodology* [and] *is clearly centered on discovery-through-action* [while design science is] "*focused on problem-solving by creating and positioning an artifact in a natural setting*" (Baskerville, 2008, p. 1).

Sixth, design science is not computer science because "*computer science is a discipline centered on computer-related artifacts* [while] *design science is much broader*" (Baskerville, 2008, p. 1).

Seventh, design science is not a separate academic discipline, ranging "*across many academic disciplines such as architecture, engineering and information systems*" and it is not new since "the Design Research Society was already founded in 1966" (Baskerville, 2008, p. 1).

Historically, design science originates in the studies presented by Simon (1996) called "The Science of the Artificial"; the first edition was released in 1969. The "artificial" means something built by humans (human artefact) and not naturally happening in nature. Simon (1996, p. 114) considers that "*design theory is aimed at broadening the capabilities of computers to aid design, drawing upon the tools of artificial intelligence and operations research*". Besides the conceptualization, Simon (1996, p. 114) sustains that "*engineers are not the only professional designers. Everyone designs who devises courses of action aimed at changing existing situations into preferred ones*" and exemplifies that there is no difference when constructing artificials for public policy, medicines or sales plans. Since then, many fields have been using design science, including education (Collins, 1992), pharmacy (Lapão, Da

Silva, & Gregório, 2017), administration sciences (Barzelay & Thompson, 2010), and information systems and software engineering (Hevner & Chatterjee, 2010; Wieringa, 2014).

2.4.1 Design Science Research (DSR) approach in this dissertation

Gregor (2006) argues that design theory is a theory for design and action. This thesis aims to develop a design theory that helps policy-makers and developers design systems to create digital transparency. Section 3.2 materializes this goal in the Paper 2 – Digital Transparency and the Usefulness for Open Government, published in Government Information Quarterly.

This thesis followed a DSR method provided by Chanson, Bogner, Bilgeri, Fleisch, and Wortmann (2019), providing a design cycle to create design principles for digital transparency applications. We follow Chanson et al. (2019) who used DSR in information systems to derive design requirements for an IoT Sensor Data Protection System (SDPS). These requirements would generate design principles to help design system features.

Our goal is not to derive principles for one case study but to arrive at a set of generalizable principles that together form a design theory for digital transparency. This is why we opted for demonstrating and evaluating the principles in three case studies instead of a single one (see section 3.2). This enables us to generalize our findings, as in the section 3.2.

Chanson et al. (2019, p. 1279) suggest that "a *design principle that is instantiated by an explicit design feature can be understood as an explanation (design principle) of why a specified piece (design feature) leads to a predefined goal (design requirement)*". This process, depicted in Figure 4, was adapted by Chanson et al. (2019) using studies from (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), (Beck, Weber, & Gregory, 2013) and (Meth, Mueller, & Maedche, 2015).

"Data and system interoperability" is an example of a design principle presented by Ricardo Matheus, Janssen, and Janowski (2021). This principle is connected to some barriers, such as "Privacy issues due to information sharing risks" and "Difficulties in processing vast volumes of data". More examples can be seen in the section 3.4.4.

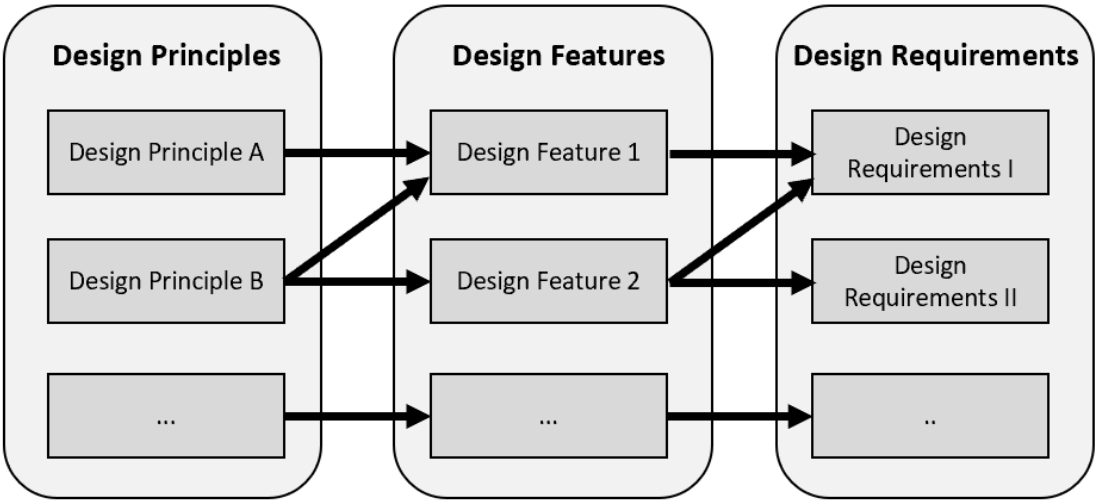


Figure 4 – Design cycles for principles, features, and requirements in information systems adapted from Chanson et al. (2019, p. 1279)

Bharosa, van Wijk, Janssen, de Winne, and Hulstijn (2011, p. 1) view principles as "*meant to guide stakeholders in proactively dealing with […] issues*". Principles are a kind of heuristic that deals with concrete issues. They are directed toward solutions for the issues at hand. Therefore, barriers to creating transparency are taken as a starting point. The research steps shown in Figure 5, were followed to derive principles.

Whereas most design approaches take an inductive approach and analyse the situation to derive barriers, we opted for a deductive approach. There is already much knowledge about barriers available in the literature (e.g., privacy issues identified by Mittelstadt and Floridi (2016) and Zakim and Schwab (2015) or data quality issues discussed by Sivarajah et al. (2017) and Alharthi et al. (2017) and this would result in a better generalization of our findings.

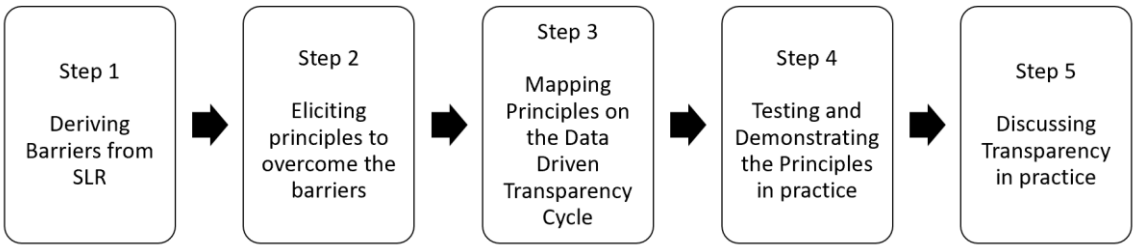


Figure 5 – Overview of the research approach

Figure 5 provides an overview of the design approach. In the first step, issues were identified, justifying the need for design principles. After that, literature was identified to address the issues found. After that, design principles were derived, becoming artefacts. Finally, the design principles were evaluated.

Step 1 was a systematic literature review (SLR) conducted to derive barriers to creating digital transparency using OGD. The SRL is described in the section 2.6.

In Step 2, principles were derived using the literature review to overcome the barriers and create digital transparency using OGD. In order to classify the principles, all were mapped on the Data-Driven Transparency cycle in Step 3 (Figure 26). Step 2 and Step 3 used the DSR approach described in the section 2.4.

In Step 4, after the principles were derived and classified, they were tested and demonstrated in practice in three real-life case studies in European countries. Step 5 discusses transparency in practice. Both Step 4 and Step 5 used case study research, as described in the section 2.8.

2.5 Research Methods Overview

This section presents an overview of all research methods used in the four papers comprising this dissertation, presented in the sections 3.1, 3.2, 3.3, and 3.4. The papers used a range of data collection methodologies. These are described in the remainder of this chapter:

- Systematic Literature Review (SLR) – Section 2.6;
- Content Analysis – Section 2.7;
- Case Study – Section 2.8;
- Semi-structured interviews – Section 2.9;
- Survey data collection method – Section 2.10; and
- Structural Equation Modelling (SEM) – Section 2.11.

Using a multi-method or mixed-method approach has a well-established tradition in the social sciences (Alexander, Thomas, Cronin, Fielding, & Moran-Ellis, 2008). According to McDonnell, Scott, and Dawson (2017, p. 1), a multi-method approach can bring various benefits, e.g., robust triangulation between methods, complementary studies of a phenomenon using the same data, and exploring "*complex social experiences and lived realities along various dimensions – so multi-dimensional realities can be captured*". For these reasons, we opted for a multi-method design science approach.

Table 3 – Overview of the research methods of the published papers

Paper Title	Paper number	Systematic Literature Review (SLR)	Content Analysis	Case Study	Survey	Semi-structured interview	Structured Equation Modelling (SEM)
A systematic literature study to unravel transparency enabled by OGD: The Window Theory	Paper 1	X	X				
Digital transparency and its usefulness for open government	Paper 2				X		X

Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities	Paper 3	X		X		X	
Design principles for creating digital transparency in government	Paper 4	X		X		X	

Error! Reference source not found. summarizes the research questions, methods, and results of this dissertation.

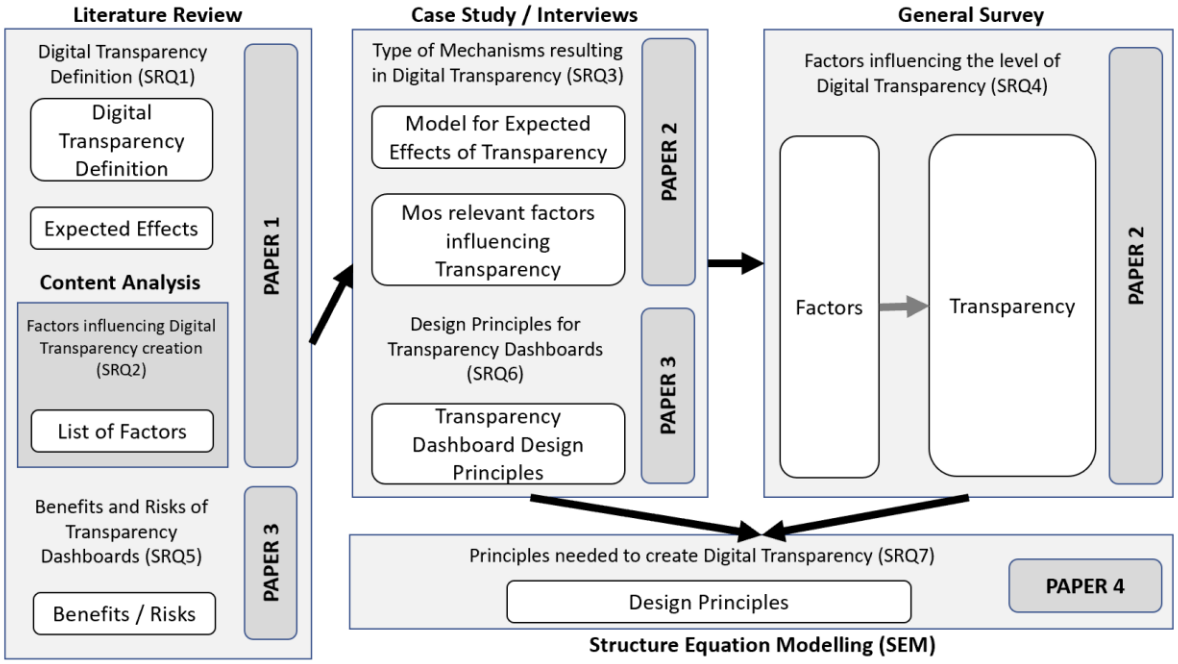


Figure 6 – Research questions, methods, and results

2.6 Systematic Literature Review (SLR) Method

Since the dissertation is based on the collection of papers, methodology sections are included in the published papers and specific methods are presented in the current chapter.

Four instances of the Systematic Literature Review (SLR) are documented in three published papers: one each in Paper 1 and 3 and two in Paper 4:

1. Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory; details are included in the section 2.6.1
2. Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities; details are in the section 2.6.2
3. Paper 4 – Design Principles for Creating Digital Transparency in Government; details are included in the sections 2.6.3 and 2.6.4

According to Fink (2019, p. 6), SLR is a "*systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars and practitioners*". Although Fink (2019) and Kitchenham (2004) provide suitable approaches to SLR, we adopted Petticrew and Roberts (2008) which has an easy step-by-step guide and provides the best fit for all four SLRs in this dissertation.

Petticrew and Roberts (2008, p. 1) mention three reasons for conducting a SLR. First, SLR provides a "*method of making sense of large bodies of information*" that might contribute to answering questions of what influences what (Petticrew & Roberts, 2008, p. 2). Second, SLR can bring us to reality when "*we think we know more than we [in reality] do*" (Petticrew & Roberts, 2008, p. 2). Third, SLR can bring more acceptance than single studies that might be "*far removed from real-life setting*" (Petticrew & Roberts, 2008, p. 3).

To conduct a SLR, Petticrew and Roberts (2008, p. 27) recommended seven stages, which are summarized in Figure 7.

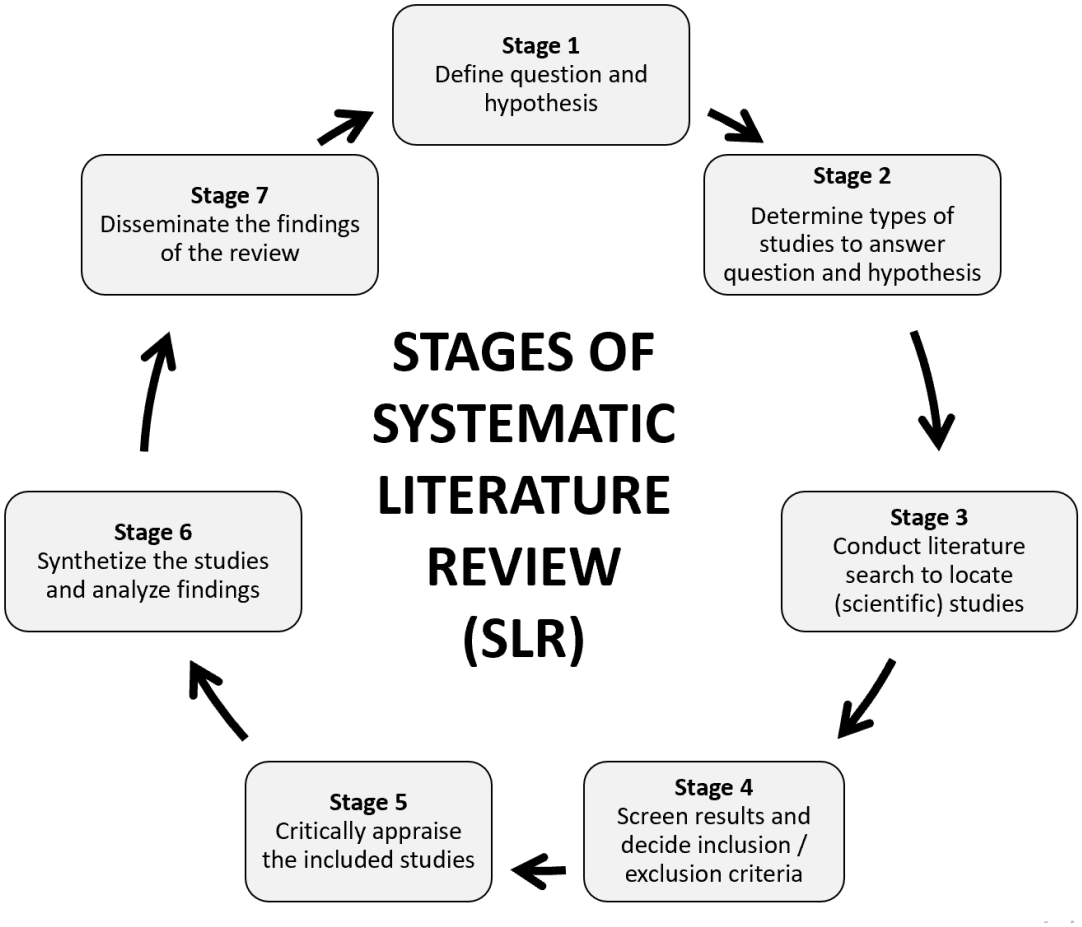


Figure 7 – Stages of systematic literature review adapted from Petticrew and Roberts (2008, p. 27)

2.6.1 SLR to Identify the Determinants Influencing Transparency and Expected Effects of Transparency – Paper 1

Following the approach by Petticrew and Roberts (2008), the SLR conducted in the Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory" comprises the stages presented in Table 4.

Table 4 – Phases and stages of SLR – Paper 1

Stage #	Stage Name	Description
1	Define questions and hypothesis	There is no question or hypothesis, but an objective of "identifying factors influencing digital transparency and expected effects of digital transparency".

2	Determine types of studies to answer questions and hypothesis	The top 25 journals with an average impact factor higher than 1,0 in the fields of Public Administration (PA) and Information Systems (IS), based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus).
3	Conduct a literature search to locate (scientific) studies	The search resulted in an initial selection of 173 papers. The number of papers was further reduced to 73, including the papers published between 2007 and 2017 only and following the inclusion/exclusion criteria mentioned in stage 4.
4	Screen results and decide the inclusion/exclusion criteria	Inclusion criteria: keywords used "Transparency" and "Government". Exclusion criteria: papers that are out of scope
5	Critically appraise the included studies	The 73 papers were analysed using content analysis, looking for sentences with "factors influencing digital transparency and expected effects of digital transparency". This content analysis is described in section 3.1.2.2, following the methods by Elo and Kyngäs (2008).
6	Synthesize the studies and analyse the findings	The content analysis resulted in a long list of factors influencing digital transparency and the expected effects of digital transparency. Synonyms were found to be used in different papers to depict the same determinant or effect. These synonyms were grouped into 42 factors and counted to create a framework, presented in Figure 21 and Table 11. Positive and negative associations were registered to help "how" factors and expected effects were influencing digital transparency.
7	Disseminate the findings of the review	The paper was published in 2019 and can be freely accessed here: https://www.tandfonline.com/doi/full/10.1080/15309576.2019.1691025

For this SLR, we adopted an instrumental view in which digital transparency is positioned between determinants (or factors) and expected effects. The initial model is shown schematically in Figure 8. Digital transparency is positioned in the centre. Determinants are variables that might enable or impede transparency, whereas expected effects refer to the variables showing digital transparency's intended and unintended consequences.

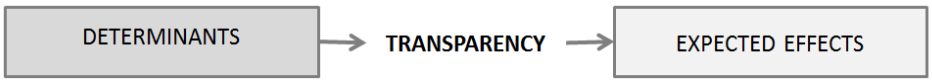


Figure 8 – Basic model of transparency

The literature surveyed included the top 25 journals with an average impact factor above 1,0 in the fields of Public Administration and Information Systems, based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus) – https://www.scimagojr.com/journalrank.php.

The search keywords were kept broad by including "transparency" and "government" to avoid missing relevant papers. The terms were only searched in the title, abstract, and keywords. The SLR was conducted in April 2017. Papers published in 2017 were excluded to avoid any confusion about which subset of papers from 2017 should be included. Searching on "transparency" and "government" returned many papers, as transparency and government are commonly used words outside the OGD field. For example, chemistry and environmental science use these words within a different context, e.g. transparency of glass or water.

Some journals did not have a search function, and in those cases, the search was performed manually by accessing each paper and looking for the keywords. The results show that much work is available in the financial domain. Budget transparency is based on well-defined measures, standards, financial reports, and regulations. Articles that focused on budget transparency were excluded, as these do not represent the typical challenges faced by not well-defined open data (Conradie & Choenni, 2014; Harrison et al., 2012; Janssen, Charalabidis, & Zuiderwijk, 2012). This resulted in an initial selection of 173 papers. The number of papers was further reduced to 73, only including papers published between 2007 and 2017.

An online Google Sheet was used to register the long list of factors and expected effects of digital transparency. Afterwards, we conducted content analysis, which resulted in a framework with factors influencing digital transparency and the expected effects of digital transparency. The framework is presented in Figure 21 and the content analysis in section 2.7.

2.6.2 SLR to Identify Design Principles for Dashboards – Paper 3

The SLR conducted in Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" comprises the stages from the Petticrew and Roberts (2008) approach that are presented in Table 5.

Table 5 – Phases and stages of SLR – Paper 3

Stage #	Stage Name	Brief Description

1	Define questions and hypothesis	There is no research question or hypotheses, but an objective to “identify transparency dashboards use in smart cities”.
2	Determine types of studies to answer questions and hypothesis	The literature review was performed using the top 20 journals of 2015 in the Scimago Journal Rank (SJR) in Information Systems, Information Systems and Management, and Library and Information Sciences. Since this option resulted in only a few papers, we conducted an open and not extensive search using Google Scholar to support the principles identified in section 3.3.3.
3	Conduct a literature search to locate (scientific) studies	The SLR was conducted from the 1st of October until the 1st of November 2016. Around 130 papers were found, considering titles, abstracts, keywords and citations. Most of the papers only mentioned dashboards in the text, whereas the research was not focused on dashboards.
4	Screen results and decide on the inclusion and exclusion criteria	Inclusion criteria: “dashboard” keyword Exclusion criteria: papers out of scope, e.g., not related to smart cities or transparency
5	Critically appraise the included studies	Out of 130 papers found, only 19 were related to smart cities and principles for dashboards. The scant literature shows that the topic is underexplored, whereas dashboards are essential in data science.
6	Synthesize the studies and analyse the findings	Since we could not find any literature supporting the creation of dashboards in smart cities, we identified from the literature review a list of 10 design principles helping to create transparency dashboards (see Table 24).
7	Disseminate the findings of the review	The paper was published in 2020 and can be freely accessed here: https://www.sciencedirect.com/science/article/pii/S0740624X18300303

In this SLR, we aimed to identify “transparency dashboards use in smart cities”. This SLR used the keyword “dashboard”. The literature review was performed using the top 20 journals in the 2015 Scimago Journal Rank (SJR) in Information Systems, Information Systems and Management, and Library and Information Science. The SLR was conducted from the 1st of October until the 1st of November 2016. It uncovered 130 papers, but only 9 were related to smart cities. The scant literature shows that the topic of dashboards is underexplored, whereas dashboards are essential in data science. Since we could not find any literature supporting the

creation of dashboards in smart cities, we identified from the literature review a list of 10 design principles that help designers create transparency dashboards (see Table 24).

2.6.3 SLR to Identify Barriers to Digital Transparency

The SLR conducted in Paper 4 – Design Principles for Creating Digital Transparency in Government" comprises the Petticrew and Roberts (2008) stages that are presented in Table 6.

Table 6 – Phases and stages of SLR – 1st SLR Paper 4

Stage #	Stage Name	Brief Description
1	Define questions and hypothesis	There is no research question or hypotheses, but the objective is to "identify barriers for creating digital transparency".
2	Determine types of studies to answer questions and hypothesis	The top 25 journals with an average impact factor higher than 1,0 in the Public Administration and Information Systems fields, based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus). We limited the publication years to the period between 2007 and 2018.
3	Conduct a literature search to locate (scientific) studies	The SLR was conducted from the 1st of April until the 31st of May 2019. This SLR resulted in a list of 50 relevant articles that helped uncover 364 barriers to digital transparency.
4	Screen results and decide on inclusion/exclusion criteria	Inclusion criteria: keywords "big data" or "open data", and "barriers" and "transparency". Exclusion criteria: papers out of scope, e.g., on finances.
5	Critically appraise the included studies	Table 25 shows the 42 different types of barriers found on SLR. Only barriers mentioned in at least two papers were included in the table, removing 40 barriers that were only mentioned once and considered irrelevant to the research.
6	Synthesize the studies and analyse the findings	The 42 identified barriers were grouped into seven categories: Data Quality (DQ), Economic (EC), Ethical (ET), Human Resources (HR), Organizational (OR), Political and Legal (PL), Technical (TE), and Usage category (US).
7	Disseminate the findings of the review	The paper was published in 2021 and can be freely accessed here: https://www.sciencedirect.com/science/article/pii/S0740624X20303294.

In this SLR, we aimed to identify barriers to creating digital transparency. We view digital transparency as a way to create transparency using digital means by collecting, processing, and presenting data. Barriers were considered as any factor impeding the creation of digital

transparency. This SLR was conducted using the keywords “big data” or “open data”, and “barriers” and “transparency”. These keywords were used to limit our search to topics related to open data, transparency, and barriers that impede the creation of digital transparency.

This SLR used the Scopus, JSTOR, SpringerLink, and Web of Science databases as sources of papers. From these databases, we surveyed the top 25 journals with an average impact factor higher than 1,0 in the Public Administration and Information Systems fields, based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus).

The SLR was conducted from the 1st of April until the 31st of May 2019. This SLR resulted in a list of 50 relevant articles, which included 364 barriers to transparency. These barriers were reduced by similarity or synonyms to 42 types of barriers, which, in turn, were grouped into eight categories – data quality, economic, ethical, human, organizational, political and legal, technical, and usage. Figure 9 summarizes the creation of such categories through paper searching, identification of barriers, and their grouping into 42 types and eight categories:

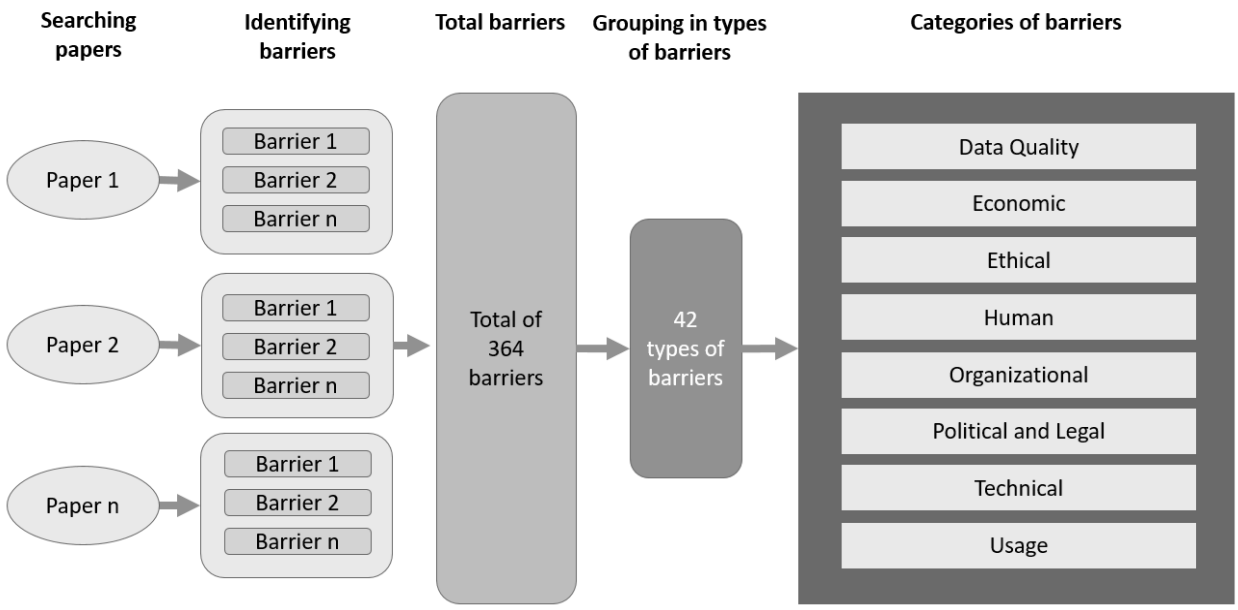


Figure 9 – Flowchart of categorizing barriers

2.6.4 SLR to Identify Principles for Creating Digital Transparency

This section presents the second SLR conducted in Paper 4 – Design Principles for Creating Digital Transparency in Government". The summary of the stages following the approach by Petticrew and Roberts (2008) is presented in Table 7.

Table 7 – Phases and stages of SLR – 2nd SLR Paper 4

Stage #	Stage Name	Brief Description
1	Define questions and hypothesis	There is no question or hypothesis, but an objective to "identify principles for creating transparency".
2	Determine types of studies to answer questions and hypothesis	This SLR used the Scopus, JSTOR, SpringerLink, and Web of Science databases as sources of papers. From these databases, we surveyed the top 25 journals having an average impact factor higher than 1,0 in the field of Public Administration and Information Systems based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus).
3	Conduct a literature search to locate (scientific) studies	The SLR resulted in 29 papers, 22 of which were found to be relevant.
4	Screen results and decide the inclusion and exclusion criteria	Inclusion criteria: keywords 'open data', 'portal', 'design' or 'architecture' and 'principle'. Exclusion criteria: out of scope papers, e.g., in biological or medical fields
5	Critically appraise the included studies	Each paper was read by two researchers, who identified the potential principles derived from the papers. Many of the principles were found in the case studies conducted in the United States of America (USA), Sweden, the Netherlands, and Denmark. In total, 186 forms of principles were found. However, only 64 of them were used.
6	Synthesize the studies and analyse the findings	Following the Open Group Architecture Framework (TOGAF) (TOGAF, 2009), the identified barriers to digital transparency were used to derive principles to overcome them. In the end, 16 principles were used in the paper (see Table 26).
7	Disseminate the findings of the review	The paper was published in 2021 and can be freely accessed here: https://www.sciencedirect.com/science/article/pii/S0740624X20303294

The SLR used the terms ‘open data’, ‘portal’, 'design' or 'architecture', and ‘principle’ to identify principles for creating transparency. This SLR used the Scopus, JSTOR, SpringerLink, and Web of Science databases as sources of papers. From these databases, we surveyed the top 25 journals having an average impact factor higher than 1,0 in the field of Public Administration and Information Systems based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus). The papers found were published in the period from 2007 to 2018.

This SLR resulted in 62 papers, of which 50 were found to be relevant to this research. Some papers had to be excluded because they were out of the scope, such as those related to the biological, e.g., (Ellison, 2010) or medical, e.g., (Asante-Korang & Jacobs, 2016) fields.

The two researchers then read each paper and identified the potential principles. Many of the principles were found in the case studies conducted in the United States of America (USA), Sweden, the Netherlands, and Denmark. In total, 186 principles were found and 64 of them were found to be relevant. Following the Open Group Architecture Framework (TOGAF) (TOGAF, 2009), barriers identified in the SLR, presented in the section 3.4.3, were used to derive principles to overcome these barriers, presented in the section 3.4.4. In the end, 16 principles were used (see Table 26). Figure 10 summarizes the process used in this SLR.

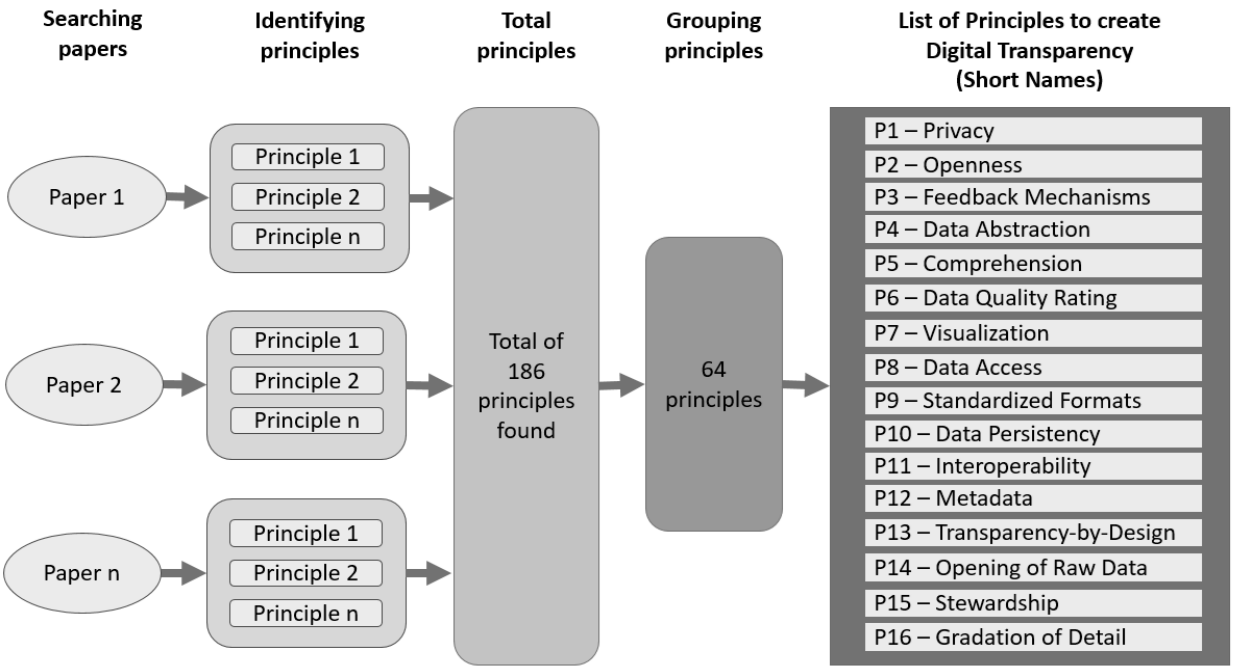


Figure 10 – Flowchart of principles to create digital transparency

2.7 Content Analysis Method

This section contains descriptions of the content analysis methods, their advantages, disadvantages, and differences with other methods such as thematic analysis. Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory" and particularly the paper's methodology section 3.1.2.2 was used as a source of the content analysis methods selected in this dissertation.

Cole (1988, p. 1) describes the content analysis method as "*a research method for analyzing written verbal or visual communication messages*". Elo and Kyngäs (2008, p. 108) also classify the content analysis method as "*a method of analysing documents*" and identify two approaches for content analysis – inductive and deductive. In terms of data, both quantitative and qualitative data can be used as a source for content analysis. An inductive approach is considered "*If there is not enough former knowledge about the phenomenon or if this knowledge is fragmented*" (Elo & Kyngäs, 2008, p. 109). A deductive approach is recommended if "*the structure of analysis is operationalized on the basis of previous knowledge and the purpose of the study is theory testing*" (Elo & Kyngäs, 2008, p. 109). We selected the inductive approach from Elo and Kyngäs (2008, p. 108) because of its easy step-by-step guide and to achieve the best fit for our research. Figure 11 shows the overview of the content analysis and model construction for Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory". The four steps followed in the paper are adapted from the steps in Elo and Kyngäs (2008, pp. 109-111). They are described in the sections 2.7.1 to 2.7.4.

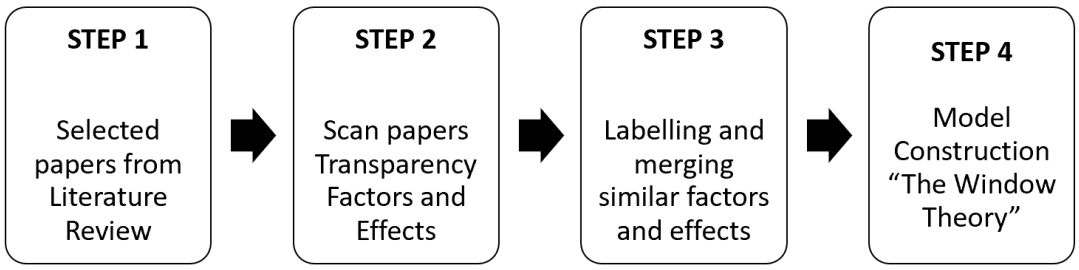


Figure 11 – Content analysis and model construction overview

2.7.1 Step 1 – Selecting Papers for SLR

In step 1, we use the papers selected for this research by SLR, as described in the section 2.6.1. Our SLR relies on previously published research, and the selection/exclusion criteria limit the number of papers that can be managed. Some potentially relevant papers might have been missed because they were not published in the outlets surveyed. We excluded the literature concerning budgetary data, which is well-structured and well-defined, unlike most OGDs.

2.7.2 Step 2 – Scanning Transparency Factors and Expected Transparency Effects

Elo and Kyngäs (2008, p. 109) pointed out that "*there are no systematic rules for analysing data*", but the main objective is to classify words from a text into much smaller content categories. The preparation phase has two phases. The first is selecting the unit of analysis, such as a theme, topic or specific word (Elo & Kyngäs, 2008, p. 109). The second is making sense of the data. In this phase, Elo and Kyngäs (2008, p. 109) advocate researchers "*to learn 'what is going'* [on the text] *and obtain a sense of whole*", aiming to "*become immersed in the data*". After these decisions, researchers must decide if they will follow an inductive or deductive approach. The second phase reports the model, conceptual system, conceptual map, or categories that emerged from both inductive and deductive approaches.

In step 2, both Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory" authors conducted a scan of transparency factors and expected transparency effects. All potential factors and effects were individually registered in an online spreadsheet to keep the list broad. This resulted in a long list of factors presented in the section 3.1.3.

The authors of Paper 1 had some difficulties when scanning these factors and expected effects. Many papers lack clear definitions and conceptualization, and this complicates the SLR. It was assumed that they would use the same definitions in such situations.

We also identified papers with any association between the factors and expected effects, especially papers with written (text) or graphical (figures, equations) associations. This identification helped construct the model described in Step 4 (see section 2.7.4). Also, it was registered if these associations were positive or negative.

2.7.3 Step 3 – Labelling and Merging similar factors and effects

This step consists of labelling and merging similar factors and effects. After registering all potential factors and effects, the authors read all of them individually, aiming to find synonymous or similar terms. Some of these factors and expected effects registered were synonyms. While some papers used one term, others used similar terms that could be aggregated into one common term or merge terms occurring with minor frequency into terms occurring with bigger frequency. On some occasions, exact wording was used for slightly different situations. For example, primary data can refer to raw data that has not yet been processed but also to data collected from the source and made ready for use. Instead of trying to define each element, we would label them both as primary data.

2.7.4 Step 4 – Model Construction

Step 4 creates a model of associations between transparency factors and expected transparency effects. The results are presented in section 3.1.4 of Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory".

A tree diagram was created using the Visual Understanding Environment software (http://vue.tufts.edu/). This diagram named 'The Window Theory' is presented in Figure 21. It reflects the determinants of transparency and their possible effects. The Window Theory is not a tested theory but an overview of the relationships uncovered by the surveyed papers.

All factors and expected effects were manually included in the VUE software. Associations were visually identified from scientific literature during Step 2 (see the list of scientific papers in Table 12). Some papers expressed clear associations written in text or graphically in figures. A few had statistical evidence of the relationship between factors. After studying all the models found in scientific literature, the content analysis resulted in four clusters of determinants of transparency and one cluster of expected effects on transparency.

Hsieh and Shannon (2005) recommend creating relationships between categories and subcategories based on their concurrence, influence, and consequences. Factors and effects were catalogued accordingly, including whether a determinant had a positive or negative relationship with transparency. Factors having similar semantics were clustered. After reading the papers, 42 determinants and 8 expected effects were found. Hsieh and Shannon (2005) suggest creating a diagram showing the relationships between the clusters and factors.

The Window Theory aggregates determinants and effects into a single mode but does not consider contextual factors. Some factors might be relevant to a certain situation only. In addition, the relevance and significance of the factors could not be aggregated due to the small number of quantitative studies that employ different constructs.

Besides the poor definitions and conceptualization found during Step 2, in many papers where the literature suggested a causal or correlated relationship, the explanatory mechanisms were poorly described or could be challenged. For example, it is often stated that access should be given to raw data. Although publishing raw data enables finding new insights and avoids predefining views, it might not per se result in transparency, as raw data might not be easy to understand and use by most people. As such, the mechanisms for creating transparency might not work in all circumstances. Due to this, we decided to include all potential statements, as transparency might work in different ways depending on the context, i.e., people, data, objective, etc. (Matheus and Janssen, 2015a).

2.8 Case Study Research Method

This section presents the case study research method used for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" and Paper 4 – Design Principles for Creating Digital Transparency in Government" introduced in the sections 3.3 and 3.4 respectively. After presenting the method in this section, two sub-sections explain which decisions were made for each paper.

2.8.1 The Method

We followed Yin (2013) as the primary source for the case study research approach. Yin (2013, pp. 45-46) views case study research as a comprehensive research strategy that follows a predefined research method and is often used based on opportunism rather than rational grounds. Case studies can be considered empirical due to the existence of a formal description of the research design, which is divided into six steps: plan, design, prepare, collect, analyse, and share (see Figure 12).

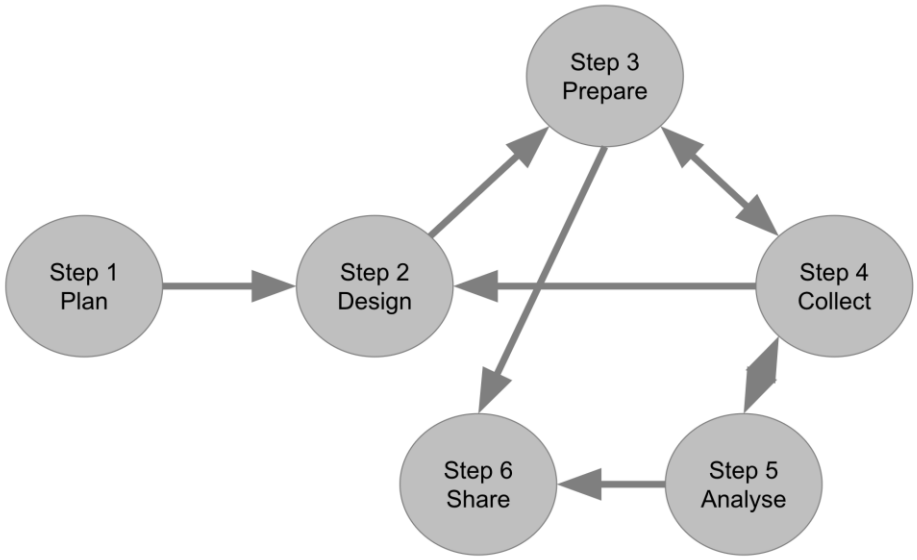


Figure 12 – Case study flow adapted from Yin (2013, p. 31)

In Step 1 – Case Study Plan, Yin (2013, p. 33) points out that the first condition, "What is the form of the Research Questions?" indicates that case studies should have a "how" or "why" in the research question, and/or this question needs an "*in-depth description of some social phenomenon*" Yin (2013, p. 33).

In Step 2 – Designing Case Study, Yin (2013, p. 57) describes the design step as a "*link between data collected* […] *to the initial questions*" and states that the case study work is analytical rather than statistical generalization.

In Step 3 – Case Study Preparation, Yin (2013, p. 118) recommends following determined actions related to expected skills and values to follow, training to do the case study, and developing the case study protocol. Following Yin (2013, p. 119), there are five recommendations for ensuring the data collection will occur evenly:

1. *Desired skills and values of case study investigator;*
2. *Training for a specific case study;*
3. *Developing a protocol for the study;*
4. *Screening candidate cases; and*
5. *Conducting a pilot case study.*

Regarding the *desired skills and values of the case study investigator*, a case study is not a simple research method, and the difference between research case study and non-research case studies is demonstrated by Yin (2013, pp. 22 and 50-52). Besides that, we also followed a basic list of desired attributes is given by Yin (2013, p. 120): 1) *Ask good questions and interpret them fairly;* 2) *Be a good listener leaving out any type of bias; 3) Stay adaptive to find and use opportunities as soon as they appear; and 4) Conduct research ethically.* This dissertation demonstrates this in all papers that have already been published.

In Step 4 – Case Study Collection, Yin (2013, p. 152) recommends six sources of evidence: 1) documents, 2) archived records, 3) interviews, 4) direct observations, 5) participant observations, and 6) physical artefacts. We used documents from the literature review, interviews in the format of digital questionnaires, and physical artefacts in the format of transparency applications. Besides that, Yin (2013, p. 170) recommends four principles for Data Collection, which were followed in this research, except principle 4 since we didn't use data from social media sources:

1. Principle 1: Use multiple sources of evidence
2. Principle 2: Create a case study database

3. Principle 3: Maintain a chain of evidence
4. Principle 4: Exercise care when using data from social media sources

In Step 5 – Case Study Analysis, Yin (2013, p. 211) recommends creating an analytical strategy. While quantitative research aims at statistical generalization, qualitative research aims to provide analytical generalization. A case study has a qualitative instead of quantitative logic through data analysis categorization, tabulating, testing, or recombining data to draw empirical conclusions (Yin, 2013, p. 215). In this dissertation, we tabulated with Google Sheets for all online Google Forms questionnaire questions.

Regarding analytic techniques, Yin (2013, p. 223) recommends using one or more of five techniques: 1) pattern matching, 2) explanation building, 3) time-series analysis, 4) logic models, and 5) cross-case synthesis. In Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities", we opted for pattern matching when selecting the research question "*How digital transparency initiatives overcome the various barriers hindering digital transparency and create a window for the public to view the internal functioning of government*". In Paper 4 – Design Principles for Creating Digital Transparency in Government", we opted for pattern matching when selecting the research question "What are the principles to overcome such barriers to digital transparency?" or, in other words, "What are the principles enabling digital transparency?".

In Step 6 – Case Study Sharing, Yin (2013, p. 272) recommends creating a plan for potential audiences and orienting the case study to this specific audience. Since we have an academic audience, most of our sharing strategy is based on scientific publications in journals.

2.8.2 Paper 3

In Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities", following Yin (2013, p. 33), we set forth the case study question: "*How can dashboards support the creation of digital transparency?*". The question is justified since dashboards are used by civil servants and public audiences (people) as tools to create digital transparency. We understand that this is an event of an in-depth description of a social phenomenon, as recommended by Yin (2013, p. 33).

We decided to use a multiple case studies approach in Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities". Two cases in the Rio de Janeiro city hall were selected. The first is related to the Car Traffic Jams Dashboard, and the second to the Public Bus Transport Traffic Dashboard. These cases were selected to highlight diverse audiences of digital transparency. While the first case focused on private urban mobility (cars), the second on public urban mobility (buses). We considered these two audiences for the same data: traffic jams and time spent in urban mobility. Besides that, the first author worked in the Rio de Janeiro city hall and had access to all the data to conduct this study.

Following Yin (2013, pp. 83-84), our case study for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" is holistic. This is because there is a single unit of analysis: digital transparency using dashboards. A holistic case was selected considering the research objective of studying a governmental dashboard that discloses data to diverse audiences: users of private cars or public transportation (buses) and civil servants planning and monitoring traffic jams at the IBM Center of Operations Rio de Janeiro.

The multiple-case study research was selected considering diverse audiences (users of private cars and public buses and civil servants) for the same data (traffic jams) and the objective of using similar digital dashboard tools for digital transparency. We aimed to understand if digital transparency works in diverse scenarios for the same problem to reduce traffic jams and the time spent in urban mobility.

The four criteria for maintaining the quality of a case study recommended by Yin (2013, p. 79) are construct validity, internal validity, external validity, and reliability.

For *Construct Validity*, we followed recommendations when using multiple sources of evidence. First, two cases in different Rio de Janeiro City Hall scenarios were selected. Second, we surveyed more than one person in each case, collecting data to gather more diverse opinions regarding digital transparency using public digital dashboards for traffic jams and the time spent in urban mobility.

For *Internal Validity*, we also followed recommendations when addressing rival explanations and theories from a literature review conducted in Paper 3 – Data Science

Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities. This literature review brought a list of digital dashboards' benefits, risks and challenges. A pattern matching between the barriers identified in the literature review and design principles was made in section 3.3.4. We divided the barriers and created design principles to overcome them when creating digital transparency dashboards.

For *External Validity*, we also followed Yin's recommendations when using replication logic in multiple case studies conducted in Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities. If desired, other authors could compare our approach and make comparisons, analytic generalization or statistical generalization similar to Paper 4 – Design Principles for Creating Digital Transparency in Government".

For Reliability, we followed recommendations and created a case study protocol with transparent methodology, survey questionnaire, and results, keeping a chain of evidence.

Figure 13 shows the overview of the case study research approach of Paper 3.

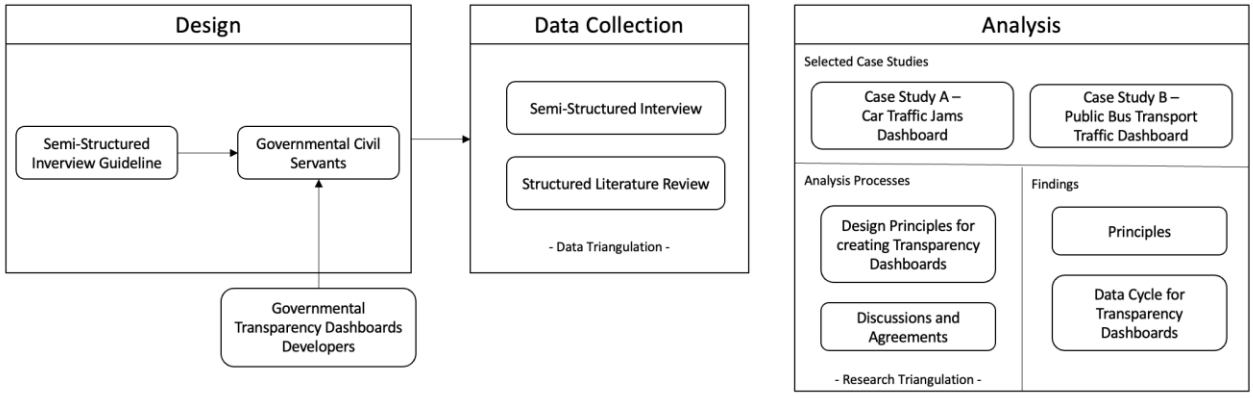


Figure 13 – Overview of the case study approach taken in Paper 3

2.8.3 Paper 4

In Paper 4 – Design Principles for Creating Digital Transparency in Government", following Yin (2013, p. 33), we set forth a case study question "*How digital transparency initiatives overcome the various barriers hindering digital transparency and create a window for the public to view the internal functioning of government*". The question is justified since digital transparency is understood in this dissertation as a behavioural event; it is a

contemporary event due to technological development and general access to ICTs, such as the Internet, 3G/4G/5G Internet, smartphones, and computers, to major parts of the population.

We decided to use a multiple case study approach in Paper 4 – Design Principles for Creating Digital Transparency in Government". Three countries were selected to conduct such case studies – Belgium, England and Ireland – because they provide diverse cases and are part of the OpenGovIntelligence (OGI) project. One of the authors worked as a work package leader in the OGI project. These cases are described in the section 3.4.5.

The three cases were selected based on the following criteria:

1. The cases must represent a governmental website that discloses data to the public, i.e. citizens, entrepreneurs, etc.;
2. Each case represents a different expected transparency effect: co-creation, accountability, and decision-making; and
3. Organizations must provide researchers with access to data.

The purpose of the case study interview is to identify the architectural principles driving the creation of transparency websites. Three cases were selected due to their contextual relevance and recent phenomenon. Transparency architectural factors and principles for creating transparency are a new area of study, and there is no clear description in scientific literature explaining the boundaries between the phenomenon and context.

According to Yin (2013, pp. 83-84), our case study for Paper 4 is holistic. We selected a holistic case study because there is only one unit of analysis – design principles for digital transparency applications using OGD, and considering the research objective of studying a governmental website that discloses data to the public. This case study provides deep insights into how transparency mechanisms work, including exploring the architectural factors and principles influencing transparency mechanisms and the reasons (why) for their influence and enabling the creation of transparency for their audiences.

The multiple-case study research was selected considering the second selection criteria about the three expected effects of transparency (co-creation, accountability, and decision-making). We also aimed to identify whether the different expected transparency effects would result in different types of transparency mechanisms or a different range of factors influencing digital transparency. The research uses the replication logic instead of the sampling logic.

The first case is that of the Flemish government, which aims to improve decision-making and accountability by monitoring air pollution with close to real-time data. The case website is available at https://www.milieuinfo.be/emissiepunten/.

The second case was selected to investigate the decision-making and expected effect of transparency. The Trafford Council, part of Greater Manchester, provides services to around 226.000 people and works with neighbouring councils to share ideas, innovations, and, in some cases, services. A priority for Trafford is economic growth: to support businesses, create jobs, and tackle unemployment and the accompanying social challenges. This application aims to help policy-makers and street-level bureaucrats improve decision-making to reduce the *worklessness* in Trafford Council and Great Manchester. The transparency website is available at http://www.trafforddatalab.io/opengovintelligence/.

The third case was selected to analyse co-creation as an expected effect of transparency. Through the governmental agency Marine Institute, the Irish government created an app called the Irish National Tide Gauge Network. This app aims to enhance the value of the marine data assets for scenario purposes by structuring and enriching the data with vocabulary and semantic value to aid the requirements of the scenarios. This app's creation was influenced by internal and external stakeholders such as civil servants in the Marine Institute, enterprises in the Leisure sector, and programmers in the maritime sector. The app can be accessed here: http://vis.marine.ie/dashboards/#/dashboards/wave_spectral?buoy=AMETS%20Berth%20A%20Wave%20Buoy&measurement=SignificantWaveHeight.

The four criteria for maintaining the quality of a case study recommended by Yin (2013, p. 79) are construct validity, internal validity, external validity, and reliability.

For *Construct Validity*, we followed recommendations when using multiple sources of evidence. First, we choose three cases in different European countries. Second, we surveyed more than one person in each case, collecting data for more diverse opinions.

For *Internal Validity*, we also followed recommendations when addressing rival explanations and theories from a literature review conducted in Paper 1. This paper brought a list of factors influencing digital transparency. A pattern matching between the barriers identified in the literature review and design principles was made in the section 3.4.4. We identified the barriers to digital transparency and created design principles to overcome them.

For *External Validity*, we also followed Yin's recommendations when using replication logic in multiple case studies conducted in Paper 4 – Design Principles for Creating Digital Transparency in Government. If desired, other authors could compare our approach and make comparisons and analytic or statistical generalizations.

For *Reliability*, we followed Yin's recommendations and created a case study protocol with transparent methodology, survey questionnaire, and documented results, keeping a database and chain of evidence to back them up.

Figure 14 shows the overview of the case study research approach. The cases were selected to cover different types of influence on transparency, considering SLR in Paper 1.

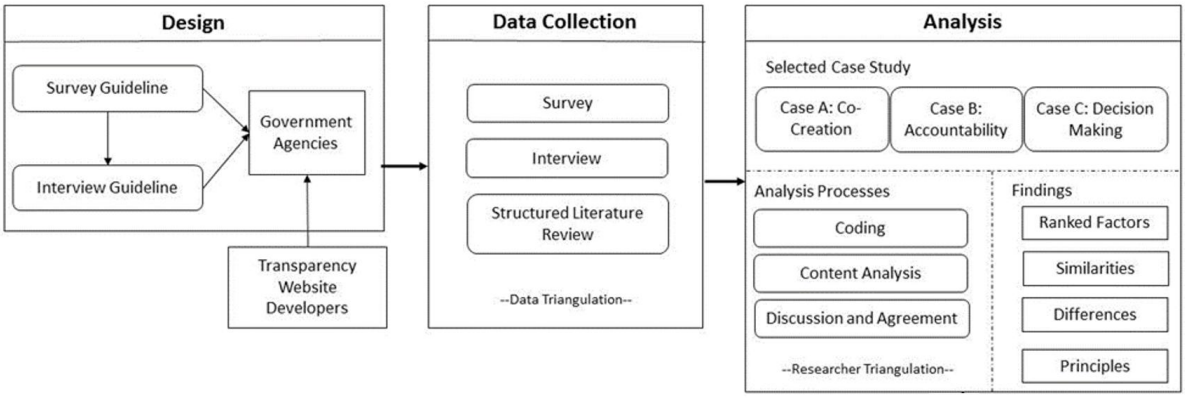


Figure 14 – Overview of the case study approach adopted in Paper 4

2.9 Semi-Structured Interviews Method

This section presents the semi-structured interview method used in Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" and Paper 4 – Design Principles for Creating Digital Transparency in Government" included in the sections 3.3 and 3.4 respectively. In the remainder of this section, we first describe the general methodology and then explain it for both papers.

2.9.1 The Method

Although authors such as Louise Barriball and While (1994) or Longhurst (2003) are providing semi-structured interview methodologies, we opted to use the methodology provided by Kallio, Pietilä, Johnson, and Kangasniemi (2016) because we did not have in-depth information on

how the two digital transparency applications worked in practice. This fits what Kallio et al. (2016) recommended as in-depth information and participants' knowledge of the topic.

Following Longhurst (2003, p. 143), semi-structured interviews are "*a verbal interchange where one person, the interviewer, attempts to elicit information from another person by asking questions. Although the interviewer prepares a list of predetermined questions, semi-structured interviews unfold in a conversational manner offering participants the chance to explore issues they feel are important*".

For Kallio et al. (2016, p. 2955), semi-structured interviews can be conducted individually or in groups, with a variety of structures, depending on the study purpose and research questions. The advantages listed by Kallio et al. (2016, p. 2955) are:

- It enables reciprocity between the interviewer and participant;
- It enables the interviewer to improvise follow-up questions based on the participants' responses; and
- It allows space for participants' verbal expressions.

Besides the positive aspects, Kallio et al. (2016, p. 2955) highlighted such issues as:

- The need for the interviewer's knowledge of previous studies in the research area;
- Proper preparation of guides and questions before the interview; and
- The high level of rigour when conducting and reporting qualitative studies.

Nonetheless, Kallio et al. (2016, p. 2962) proposed a five-step framework for the development of a qualitative semi-structured interview guide, summarized in Table 8. These steps were followed by Paper 3 and Paper 4, described in the sections 3.3 and 3.4.

Table 8 – Overview of the steps for the qualitative semi-structured interviews adapted from Kallio et al. (2016, p. 2962)

Step Name	**Summary description**
Identifying the prerequisites for using a semi-structured interview	This step evaluates if a semi-structured interview is a proper method for the selected research question.

Retrieving and utilizing the previous knowledge	This step aims to identify previous knowledge of the subject, such as an *a priori conducted* literature review, and other empirical complements, such as other practitioners' expertise.
Formulating the preliminary interview guide	This step aims to formulate an interview guide for data collection (questionnaire). It might cover the content to answer the research questions.
Pilot Testing	This step aims to confirm the coverage and relevance of the content of the preliminary guide created in Step 3. Potential external experts assess the guide, test it with colleagues, and refine it (questionnaire).
Presenting the complete interview guide	This step aims to produce a clear and logical guide for data collection after design, testing, and refining in the previous steps.

2.9.2 Semi-structured Interviews for Paper 3 and Paper 4

For Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities", the semi-structured interview guide is presented in section 3.3.7.1 following the steps by Kallio et al. (2016). The interview was conducted with former partners of the Data Science group in Rio de Janeiro city hall (Brazil) called "P3NS4" inside the Smart City intelligence department of the IBM Center of Operations Rio de Janeiro (COR) – https://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html. The semi-structured interview was selected. Kallio et al. (2016) recommended such interviews when necessary for in-depth information and when participants are knowledgeable about the topic. For this dissertation, both conditions hold.

For Paper 4 – Design Principles for Creating Digital Transparency in Government", the semi-structured interview guide followed the steps proposed by Kallio et al. (2016). The interviews were conducted with researchers in OGD and transparency who participate in the EU-funded OpenGovIntelligence (OGI) project. In total, nine designers in charge of creating transparency applications from 4 countries (Belgium, England, Ireland, and the Netherlands) were interviewed about three digital transparency applications in Belgium, England, and Ireland. Semi-structured interviews were selected, taking into consideration the positive aspects of participants expressing their individual opinions and potentially negative aspects, such as the need for a certain level of knowledge about the topic (Kallio et al., 2016).

Step 1: Identifying the prerequisites to use a semi-structured interview

Following Kallio et al. (2016, p. 2959), the first step aims "*to evaluate the appropriateness of the semi-structured interview as a rigorous data collection method in relation to the selected*

research question". After deciding on the research question and the field for these interviews, Kallio et al. (2016, p. 2959) recommended researchers to "*determine some areas of the phenomenon based on previous knowledge before the interview*" and described semi-structured interviews as "*a suitable* [method] *for studying people's perceptions and opinions*", or when "*participants had a low level of awareness of the subject*", or when it is possible to focus on the issues that "*were meaningful for the participant, allowing diverse perceptions to be expressed*" (Kallio et al., 2016, p. 2959).

The first step for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities", identified as the primary objective "*to understand and to support the design of dashboards for creating transparency*" (Ricardo Matheus, Janssen, & Maheshwari, 2018). Dashboards are a type of digital transparency application. These were created within the context of Smart City Rio during the World Cup 2014 and the Olympics Games 2016, both hosted by Brazil and Rio de Janeiro. The first dashboard labelled "Traffic Dashboard", aimed to show the real-time traffic jams in the city using big data from mobile web applications. The second, "Public Transport Dashboard", aimed to show in real-time the buses in the city and the fastest options for citizens take them.

The first step for Paper 4 – Design Principles for Creating Digital Transparency in Government recommended" identified the main objective of using a semi-structured interview, i.e. to check if the 16 design principles are helpful for real-case scenarios in a diversity of countries and systems. This list of 16 design principles (see the sections 3.4.4 and 3.4.8.3) was derived based on an SLR into the barriers to digital transparency (see the sections 3.4.3 and 3.4.8.2).

Step 2: Retrieving and using previous knowledge

In the second step, Kallio et al. (2016, p. 2959) posited its aims "*to gain a comprehensive and adequate understanding of the subject, which required critical appraisal of previous knowledge and the possible need for complementary empirical knowledge*". This previous knowledge might help to create a "*framework for the interview* [and] (…) *create a conceptual basis for the interview*" (Kallio et al., 2016, p. 2959). To identify this previous knowledge, Kallio et al. (2016, p. 2959) recommended carrying out a literature review.

If the literature review is scarce or fragmented, Kallio et al. (2016, p. 2959) suggested the use of "*empirical knowledge to complement and deepen the theoretical background*", using for

example experts in the field to gain empirical knowledge "*to seek understanding of the study phenomenon*" (Kallio et al., 2016, p. 2959). Empirical research can be gathered through focus groups, workshops with team members, and individual interviews (Kallio et al., 2016, p. 2959).

As the second step for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" we conducted a literature review and found Alawadhi et al. (2012) and Paroutis, Bennett, Heracleous, and Change (2014) useful to create the questionnaire described in the next step.

As the second step for Paper 4 – Design Principles for Creating Digital Transparency in Government" we conducted an SLR of the barriers to digital transparency. As we could not find any scientific or practical sources to inspire our research and questions, we created the semi-structured interview questionnaire with a comprehensive overview of design principles and real-case scenarios.

Step 3: Formulating the preliminary semi-structured interview guide

The third step aims "*to formulate an interview guide as a tool for interview data collection, using previous knowledge on structural, logical and coherent forms*" (Kallio et al., 2016, p. 2959). Kallio et al. (2016, p. 2960) described an interview guide as "*a list of questions, which directs the conversation towards the research topic during the interview*". This format of the semi-structured interview can have some positive characteristics, such as "*loose and flexible, which* [allows] *dialogue during an interview* [and] *change the order of the questions and easy movement from question to question*" (Kallio et al., 2016, p. 2960).

Kallio et al. (2016, p. 2960) highlighted that the "*quality of the interview guide affects the implementation of the interview and the analysis of the collected data*". For this reason, they recommend "*well-formulated questions in the guide*", avoiding "*leading*" interviewed people to any conclusion or answer, besides being "*clearly worded, single-faceted and open-ended*" (Kallio et al., 2016, p. 2960). The result would be a guide that "*generates answers from participants that were spontaneous, in-depth, unique and vivid*" (Kallio et al., 2016, p. 2960).

When writing questions, Kallio et al. (2016, p. 2960) list descriptive answers that start with words like "*what, who, where, when or how*" that could be used to gain in-depth knowledge about the question. These questions are divided into the main themes and follow-up questions (Kallio et al., 2016, p. 2960). The former covers "the main content of the research subject",

where all participants are encouraged "*to speak freely about their perceptions and experiences*" (Kallio et al., 2016, p. 2960). The recommended order is progressive or logical, "*a warm-up to break the ice and create a relaxed environment*" (Kallio et al., 2016, p. 2960).

As the third step for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" recommended by Kallio et al. (2016), we created a preliminary semi-structured interview guide. This guide had 12 "main" questions about the dashboards; the first was to identify if the respondents worked with the dashboards. Then, we included questions about the dashboard objectives, which organizations worked with them, which data was collected and used, functionalities included, expected benefits and values, barriers to creating dashboards, and how they were overcome. These questions helped us understand how transparency dashboards were designed in our case study.

As the third step for Paper 4 – Design Principles for Creating Digital Transparency in Government" recommended by Kallio et al. (2016), we created a preliminary semi-structured interview guide, shown in the Annex 3.4.8.4. Initially, we created a list of questions to identify the participants, whether participants followed any design principles when creating transparency applications, and the questions on each of the 16 design principles. We created the "main" questions about design principles and "follow-up" questions for participants to share their opinions about aspects of these design principles. The logic used to organize the questions was progressive, from explaining the principles and checking if participants followed any of them to asking for participants' opinions about them.

Step 4: Pilot testing of the interview guide

The fourth step aims to "*confirm the coverage and relevance of the content of the formulated, preliminary guide*" (Kallio et al., 2016, p. 2960) and, if necessary, "*to identify the possible need to reformulate questions and to test implementation of it*" (Kallio et al., 2016, p. 2960). This step allows researchers to refine the questionnaire and improve the data collection quality, including practising the interview skills. Kallio et al. (2016, p. 2960) suggested that this step can "*produce useful information about research integrity* (…) [and] *research ethics*".

Kallio et al. (2016, p. 2960) recommend three techniques to test the interview guide: internal testing, expert assessment, and field testing. When other types of pilot tests are impossible, internal testing can be conducted with colleagues, bringing "*critical information about the*

interview guide in general" (Kallio et al., 2016, p. 2960). The expert assessment uses specialists outside the research department to check if the interview guide is appropriate and comprehensive (Kallio et al., 2016, p. 2961). The third technique, field testing, uses participants from practice. One of the most used techniques, field testing can assure intelligibility, allow selection of questions or change them to more relevant ones, re-order the questions, check how long an interview can take, identify gaps, flaws, and limitations in the design, and refine the coverage of the interview guide (Kallio et al., 2016, p. 2961).

As the fourth step for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" we tested the interview guide with department colleagues internally. A few tips were provided to change and improve the interview guide: to explain the objectives of this interview better, to include details of data treatment (ethics), to organize the questions starting with "general questions" and going toward more specific topics such as "barriers", "benefits", and "technical perspectives". The last tip was to create some "follow-up" questions. Such questions, e.g. "please give me more details about 'X topic' you commented on in your previous question", were considered unnecessary.

As the fourth step for Paper 4 – Design Principles for Creating Digital Transparency in Government" we tested our interview guide with three experts from Asia, Europe, and South America, using internal testing and expert assessment. This step helped pay attention to explaining the interview, creating an "introduction" section, and complying with the requirement that all data collection be anonymous to participants.

Step 5: Presenting the complete semi-structured interview guide

The fifth step aims to "*produce a clear, finished and logical semi-structured interview guide for data collection*" (Kallio et al., 2016, p. 2961). It includes changes and reflections to consider previous steps. This final interview guide is the final questionnaire to be used by participants.

As the fifth step for Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" the final interview guide was created in the document format (see section 3.3.7.1). After refinements recommended by department colleagues through internal testing, the final guide was used to conduct the semi-structured interviews. Questions were made directly, based on general information about

transparency dashboards developed, databases used, and functionalities, expected public value, benefits, risks and challenges of transparency dashboards by Rio de Janeiro city hall.

As the fifth step for Paper 4 – Design Principles for Creating Digital Transparency in Government recommended" the final interview guide was created. The guide, which can be found in the section 3.4.8.4, consists of four sections. The first section provides an introduction to the participants, including information about ethical compliance. The second requests information about the nature of the OGI project and the participant's role in it. The third aims to determine if any design principle or enterprise architecture were followed when creating the digital transparency applications. The fourth and last section asked 13 questions about each of the 16 design principles, e.g., if the principle was followed when creating the application, how important it was, whether it was easy to implement, its impact on the organization, etc.

2.10 Survey Data Collection Method

This section presents the survey data collection method. The survey methodology is used in Paper 2 – Digital Transparency and the Usefulness for Open Government".

Andrews, Nonnecke, and Preece (2003) explained surveys as "*vehicles for collecting data*". Callegaro, Manfreda, and Vehovar (2015, p. 4) defined a survey as "*a method of systematic data collection, where we ask people questions by using standardized questionnaires for the purpose of quantitatively analysing some target population*" or the entire population, named as "*census*". Callegaro et al. (2015) suggested a series of steps divided into three phases: pre-fielding, fielding, and post-fielding. These steps are summarized in Figure 15, and discussed in the context of Paper 2 in the remainder of this section.

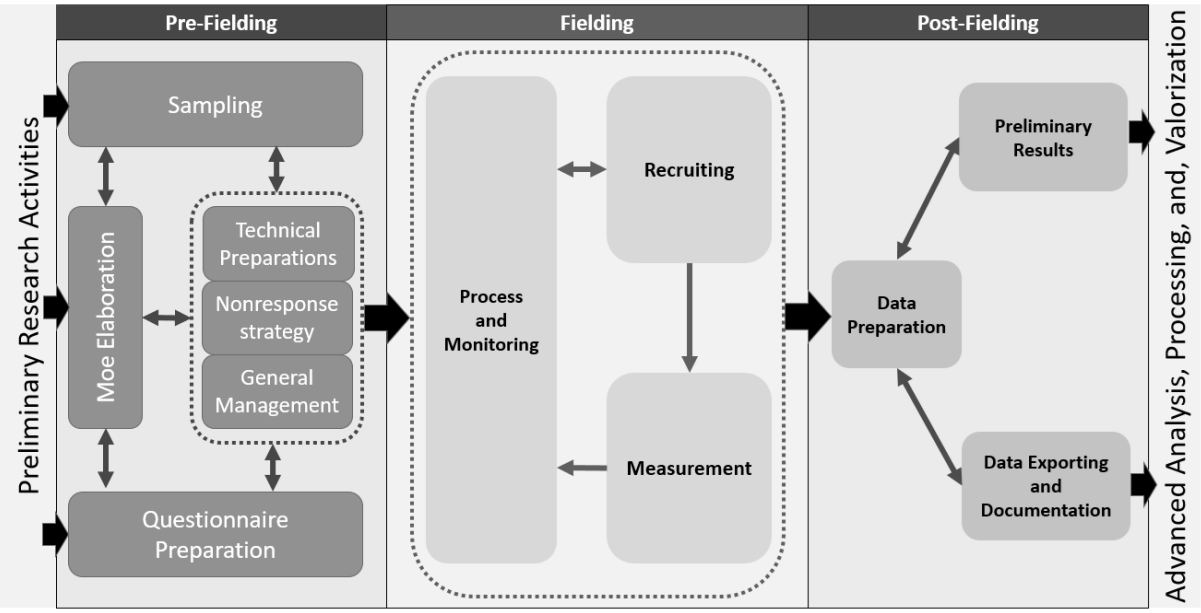


Figure 15 – Steps and stages of the web survey process adapted from Callegaro et al. (2015, p. 11)

2.10.1 Pre-fielding Survey Methodology Decisions

The pre-fielding step consists of deciding on the web survey mode, sampling and defining populations, sampling techniques, sample size, questionnaire preparation, technical preparations, and non-response strategy. These elements are discussed as follows.

Selecting the web survey mode – a web survey using a link to collect the data was used because we had a sample population spread over diverse European countries.

Sampling and defining populations – a sample of transparency dashboard users of OGI transparency applications was chosen in Belgium, England, and Ireland.

Sampling techniques – email was used to disseminate a link to the web survey in a Google Form to the audience – users of OGI transparency applications – in a simple random sample mode. There was no incentive given for people to fill out the survey.

Sample size – we aimed for a sample size of 180, which is a minimum recommended by Hooper, Coughlan, and Mullen (2008). The sample size is described in the section 2.11.2 regarding Structural Equation Modelling (SEM) methodology decisions.

Questionnaire preparations – we created and tested the questionnaire with department colleagues to identify missing questions or errors before it was sent to the OGI transparency applications users. This test conducted with seven colleagues is described in the section 3.2.3.

Technical preparations – we opted to use the Google suite: create the questionnaire in Google Docs, collect the answers in Google Forms, and use Google Sheets to present the data in tabular mode. We organized metadata to simplify data analysis. We did not collect any personal data.

Non-response strategy – creating and using a non-response strategy was unnecessary since we achieved the target sample size of 180 people.

2.10.2 Fielding Survey Methodology Decisions

We discuss this step by covering the recruiting, measurement, process and monitoring stages.

Recruiting stage – we opted to recruit the users of three digital transparency applications in the EU project OpenGovIntelligence – OGI. More than 180 people were invited, achieving the minimum number necessary to run the Structural Equation Modelling method (section 2.11). Since we achieved the minimum number, we did not need other techniques recommended by (Callegaro et al., 2015), such as sending reminders or finding alternative targeted people to respond to the survey. We conducted English grammar and technical checks on Google Forms. Nothing abnormal was uncovered, and we opened the survey for respondents to fill.

Measurement stage – we used Google Forms, and users had no issues responding to the forms. We did not create any Key Performance Indicators or dashboards to monitor the respondents, but we paid attention if the survey had more than 180 people surveyed to comply with the minimum requirements of the SEM statistical method, as suggested by Hooper et al. (2008).

Process and Monitoring stage – the research had no data quality issues and achieved the 180 respondents needed for the SEM method. Besides that, it was not necessary to follow the recommendations provided by (Callegaro et al., 2015); no responses for target periods or socio-demographics. The convenience sample helped to find all targeted respondents easily.

2.10.3 Post-fielding Methodology Decisions

Post-fielding consists of data preparation, preliminary results, data exporting and documentation.

Data preparation – we used Google Forms in the measurement stage, and Google Sheets to extract the data from the platform. In this stage, five tasks are recommended by Callegaro et al. (2015): clarification of response status, editing and validations, imputation, weighting, and coding and recoding tasks. The response status was clarified in Google Sheets, checking for abnormal entries. No personal or technical information was collected, no issue regarding entries was noted, and no respondent abandoned the survey. Editing and validations were conducted in Google Sheets. No issue was identified since most questions were closed, i.e. 5-point Likert scale from strongly disagree to strongly agree. The imputation and weighting were not needed. Coding and recoding translated the text "strongly disagree", "disagree", etc. to numbers 1 to 5 to facilitate machine reading for statistical method SEM. We also renamed all questions in codes, simplifying the visualization and working with data in the spreadsheet.

Preliminary results – we created graphical visualizations with mean and average per question using the 5-point Likert scale. These statistics provided an overview of the trends between countries and questions. The visualization facilitated the dissemination of the results to the European Commission, which is in charge of evaluating OGI project results.

Data exporting and documentation – the final database was exported from Google Sheets (web Google spreadsheet service) to the MS Excel Spreadsheet format (XLS). After that, in the first tab of the MS Excel spreadsheet, we documented all questions and their codes, e.g., Transp_1 and Transp_2 instead of full questions about transparency.

2.11 Structural Equation Modelling (SEM) Method

This section describes the Structural Equation Modelling (SEM) method, which is used to identify what mechanisms create digital transparency, translated into models with factors, and what factors influence the level of digital transparency (RQ2 and RQ3). We also describe the SEM methodology decisions for this dissertation, used in Paper 2 (see section 3.2).

SEM is the right approach for this research since we aimed to create and test a model for digital transparency. SEM can test statistically complex theories proposed by researchers. While the dissertation could use linear or multi-linear regression as a research method, the models would be simple, and measurement errors would not be included in linear regression. For these reasons, SEM is considered a better fit for this research. More reasons why SEM was selected as the research method are presented in the section 2.11.1.

2.11.1 What is SEM and Why Select It?

Following (McDonald and Ho, 2002, p. 64), Structural Equation Modelling (SEM) is a method for representing dependency relations in multivariate data in the behavioural and social sciences, also known as path analysis with latent variables. SEM is a statistical technique that measures and analyses the relationships between observed and latent variables, comprehending basic statistics, correlations, and regression analysis (Schumacker & Lomax, 2004, p. 1). According to Urbach and Ahlemann (2010), SEM is a widespread and popular method used in information systems, and "*an adequate choice, if the research problem meets certain characteristics and the technique, is properly used*" (Urbach and Ahlemann, 2010, p. 1). Hooper et al. (2008, p. 1) state that SEM has become "a 'must' for researchers in the social sciences", aiming to create models in the areas with theoretical disagreements. Schumacker and Lomax (2004, p. 6) give four reasons for selecting SEM as a research method: it can test statistically complex theories proposed by researchers, it explicitly includes measurement error in the statistical analysis of latent and observed variables, it has matured in the past decades, and user-friendly SEM software applications enable it.

All the reasons provided by Schumacker and Lomax (2004, p. 6) are helpful for our research. First, it was important to test a digital transparency model using appropriate statistical methods, improving upon other papers that do not provide a model with factors or explain what is influencing what. SEM allows us to observe multiple factors simultaneously, not only the traditional linear regression approach but also a few independent and one dependent variable. Second, SEM allows us to include error measurement in the statistical analysis. Third, SEM is becoming widely used by researchers. Fourth, easy-to-use software, such as AMOS and IBM SPSS facilitate this research (Schumacker & Lomax, 2004, p. 7).

2.11.2 SEM Methodology Decisions

After deciding on SEM as the research method, we used Schumacker and Lomax's (2004) accessible SEM guidelines (see Figure 16). We applied all steps proposed by Schumacker and Lomax (2004) except Step 10 – SEM Model Cross-Validation, because we did not have the time and funds to run another survey to collect a data sample for another SEM round and compare results. We also did not select the bootstrapping technique for cross-validation of the SEM results because the survey sample size approached the acceptance limit. This is one of the limitations of this research described in the section **Error! Reference source not found.**.

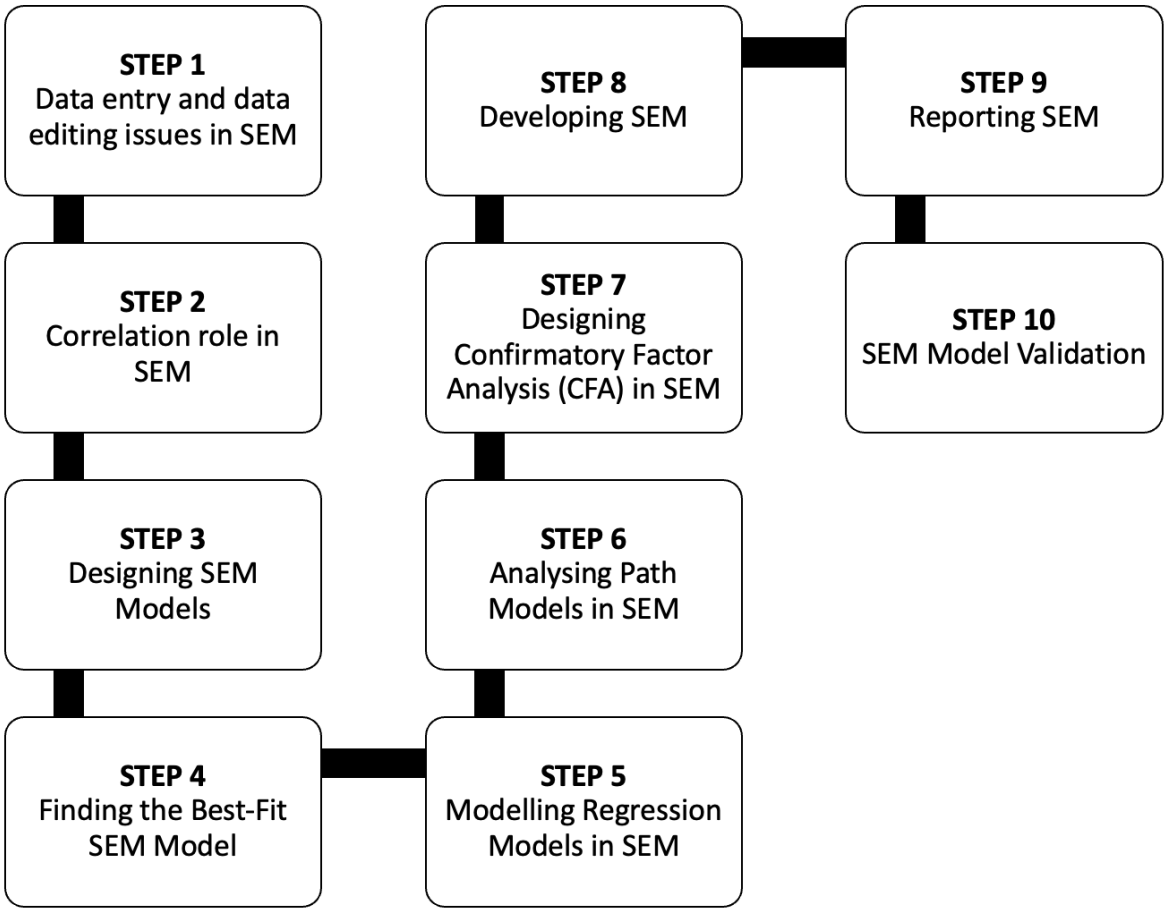


Figure 16 – Overview of the steps SEM approach adapted from Schumacker and Lomax (2004)

Hereafter, we describe each step proposed by Schumacker and Lomax (2004):

1. Step 1, "Data entry and data editing issues in SEM", deals with the entry and, if necessary, editing of data when any issues, such as missing data or outliers, are found. In this research, we collected data using a web survey. We noted common issues with data according to Schumacker and Lomax (2004) but could not find any specific issues, e.g. the measurement scale of variables, missing data values, outliers, linearity, and non-normality. This step is connected to the research method described in the section 2.10 – survey data collection.

2. Step 2, "Correlation role in SEM", describes correlation's important role when conducting SEM statistical method. In this research, we used the standards recommended by Hooper et al. (2008), which explains the importance of the sample size to SEM. The more complex the model (more latent variables, more observed variables), the more respondents are

necessary for SEM to work properly. In addition, Wolf, Harrington, Clark, and Miller (2013, pp. 10-11) also recommended that 180 respondents are a minimum for SEM statistical analysis and model evaluation. We used the website https://www.danielsoper.com/statcalc/calculator.aspx?id=89 to calculate the minimum number of respondents for SEM, and the calculator recommended 198 respondents with the following characteristics:

- Anticipated effect size: 0.2
- Desired statistical power level: 0.7
- Number of latent variables: 3
- Number of observed variables: 10
- Probability level: 0.1

3. Step 3, "Designing SEM models", provides a logical sequence of five processes to design and run the SEM analysis.
4. Step 4, "Finding the best-fit SEM model", finds the SEM model where data fits statistically well with the model designed in Step 3.
5. Step 5, "Modelling Regression Models in SEM", demonstrates how to compute parameter estimates in multiple regression and what model-fit criteria are reported.
6. Step 6, "Analysing Path Models in SEM", aims to illustrate how to analyse multiple regression equations using observed variables and test theoretical relationships.
7. Step 7, "Designing Confirmatory Factor Analysis (CFA) in SEM" includes taking measurement error into account and checking the validity and reliability of the SEM model.
8. Step 8, "Developing SEM", indicates tasks when conducting the SEM statistical analysis.
9. Step 9, "Reporting SEM", expresses how researchers should report the SEM statistical analysis to permit cross-cultural research, replication, validation, or comparison by others.
10. Step 10, "SEM Model Validation", performs the validation of a theoretical confirmatory factor model using two data samples.

Steps 3 to 9 are somehow connected and aim at finding the final theoretical model to be tested by SEM. We started finding the recommended final theoretical model, by excluding variables that did not meet the minimum threshold from the initial theoretical model. The exclusion threshold is r-squared. The r-squared results show the percentage of the variance for a dependent variable explained by an independent variable. All the R-squared (R^2) lower than 0,3 were excluded from the initial model, being considered none or weak effect size. This resulted in the exploratory model of 0.576 (see section 3.2.4.4.1, Paper 2 – Digital Transparency and the Usefulness for Open Government).

After arriving at our initial model, we opted to use the final theoretical model presented in Figure 22 – Conceptual model of digital transparency and usefulness for open government" of Paper 2 – Digital Transparency and the Usefulness for Open Government", named "tested structural model". After that, we conducted the following statistical analysis as fit-indexes:

- Cronbach Alpha (CA) – Range from 0.684 to 0.772 (see results in Table 15 – Construct reliability and validity);
- Composite Reliability (CR) – Range from 0.836 to 0.868 (see results in Table 15 – Construct reliability and validity);
- Item loadings – (see results in Table 14 – Cross loadings); and
- Average Variance Extracted (AVE) – Range from 0.669 to 0.767 (see results in Table 15 – Construct reliability and validity).

For CA and CR, we considered 0,6 as the minimum threshold. For AVE, we considered the minimum threshold higher than 0.600. After conducting the Exploratory Factor Analysis, all indicators with factor loadings below 0,500 were excluded. The final results are presented in Table 14 – Cross loadings", of Paper 2 – Digital Transparency and the Usefulness for Open Government". The software used was AMOS 25 and SPSS 24.

3 Papers Overview

This thesis is paper-based. First, section 1 introduces the topics discussed in this thesis, such as digital transparency, transparency-by-design, full transparency, and the limitations of digital transparency. After that, section 2 shows the research approach, including the research question and the methods overview of all papers. The list of papers is summarized in Table 9, and all papers are presented in section 3. Some of the text is slightly different from the original papers to ensure consistent wording between the papers.

Table 9 – Published papers included in this dissertation

Paper number and URL	Paper Title	Journal
Paper 1	A systematic literature study to unravel transparency enabled by open government data: The Window Theory	Public Performance & Management Review (PPMR)
Paper 2	Digital transparency and the usefulness for open government	International Journal of Information Management (IJIM)
Paper 3	Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities	Government Information Quarterly (GIQ)
Paper 4	Design principles for creating digital transparency in government	Government Information Quarterly (GIQ)

3.1 Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory

This chapter is based on the paper by Matheus, R., & Janssen, M. (2020) entitled "A systematic literature study to unravel transparency enabled by open government data: The window theory". Public Performance & Management Review, 43(3), 503-534. The first author conducted the literature review, designed and conducted the data collection and analysis process, and wrote the manuscript. The co-authors provided feedback on the data collection and analysis process and earlier manuscript versions. We thank the anonymous reviewers for their peer review, insightful comments, and recommendations. The full paper can be found and

freely downloaded at: https://www.tandfonline.com/doi/full/10.1080/15309576.2019.1691025.

Abstract

Data opening has been credited for improving transparency and providing a window on government functioning. Although this relationship is intuitively apparent, it is complex, and the mere opening of data might not yield transparency. In this paper, a comprehensive model of determinants that enable or impede transparency enabled by open government data and the expected effects have been derived by surveying public administration and information systems literature. Public administration literature tends to focus on factors such as participation and trust, whereas information systems literature focuses on user interface, user experience, and data quality. Digital government literature attempts to bridge these elements. The window theory is introduced to unify existing models by integrating a broad range of factors within a single model. The Window Theory can be used to develop context-dependent comprehensive and parsimonious models.

3.1.1 Introduction

Transparency is driven by the opening of government data (Bertot et al., 2010; McDermott, 2010p. 45). Transparency should result in benefits, including better government decisions (Puron-Cid, Gil-Garcia, & Luna-Reyes, 2012) and accountability (Helbig, Styrin, Canestraro, & Pardo, 2010). Expectations of the benefits of ICT-enabled transparency are high (see for example, European_Commission, 2010) and may even be exorbitantly high (Bannister & Connolly, 2011a).

Transparency is about creating insight for someone who is not involved. There is an information asymmetry between the government and the public, as those involved in government processes generally possess more information about their activities than the public (Yoon, Zo, & Ciganek, 2011). Transparency is aimed at overcoming this discrepancy in information, thus enabling the public to view what is happening within the government. This is often viewed in the context of the principle-agent theory (Eisenhardt, 1989). In digital government, transparency portals have created a window to view government functioning to overcome the information asymmetry between the government and the public.

Although transparency might be intuitively an appealing concept, people are likely to give different meanings to the concept, resulting in ambiguity. In some cases, the mere opening of data is viewed as a form of transparency, whereas others suggest that there is only transparency if the public is able to understand and interpret the data (Zyl, 2014). Furthermore, transparency is subjective. What is transparent to one person might not be transparent to another.

For example, one person might possess statistical knowledge and be able to interpret the data, whereas another person might lack the required knowledge. Transparency is sometimes even perceived as a "magical concept for everything" (Ward, 2014, p. 45). Despite the ambiguity surrounding this concept, there is an agreement that transparency is an important factor for creating open government and that government data should be opened for this purpose (Zuiderwijk, Helbig, Gil-García, & Janssen, 2014).

However, disclosing more information can result in less understanding, confusion and less trust (Bannister & Connolly, 2011a) and might not necessarily result in transparency. Therefore, it is important to understand the elements that influence transparency.

OGD initiatives are often focused on creating transparency (Bertot et al., 2010; McDermott, 2010. p. 45), however, there is no insight in enabling or impeding factors. Understanding these factors is of paramount importance to advice efforts to create transparency enabled by OGD. Furthermore, there are many factors influencing OGD-enabled transparency. They range from the technical aspects of data such as metadata (Strathern, 2000) and system characteristics, such as usability (Bertot et al., 2014) to organisational factors such as internal resistance to transparency (Navarro-Galera et al., 2016).

Although there are general models for transparency, no model exists for OGD-enabled transparency. Researchers focus on different aspects of transparency, but there is no overall overview of the determinants influencing transparency.

Also, there are various effects of transparency, ranging from accountability (Gandía, Marrahí, & Huguet, 2016) to trust and credibility (Gregory Michener, 2015). This diversity of factors suggests that transparency is multi-faceted. The effects of transparency enabled by OGD are not known. The multi-faceted nature of transparency poses difficulties for measurement, theory building and testing (Hollyer, Rosendorff, & Vreeland 2014).

Our research questions are 1) what are the determinants influencing OGD-enabled transparency and 2) what are the possible effects of OGD-enabled transparency? By answering these questions, this paper aims to develop a comprehensive model of the determinants that enable or impede transparency and to understand the possible effects of transparency in the field of Open Government Data (OGD). The resulting model is labelled the Window Theory. These determinants can help policy-makers and designers of OGD efforts to truly create transparency. Researchers can use this model to select factors that are relevant to them. This paper is structured as follows. In the next section 3.1.2, the structured literature review (SLR) approach is presented, followed by section 3.1.3 with an overview of determinants and effects of transparency. In section 3.1.4, the determinants and effects found are used to create the model of the Window Theory. The findings and further research are discussed in section 3.1.5. Finally, our conclusions are drawn.

3.1.2 Research Method

We opted to conduct a literature review, as there is a large body of work available in the field of transparency. SLR is a "method of making sense of large bodies of information" (Petticrew & Roberts, 2008, p. 2).

3.1.2.1 Literature Review Method

The literature review method was based on the Petticrew and Roberts (2008) approach to literature review, which was extended using the Hsieh and Shannon (2005) approach to content analysis. The latter was used to identify the determinants of transparency and the expected effects. We adopted an instrumental view in which transparency is positioned between determinants and effects. The initial model, which positions transparency in the centre, is shown schematically in Figure 17 . Determinants enable or impede transparency, whereas expected effects refer to variables showing the intended and unintended consequences of transparency.

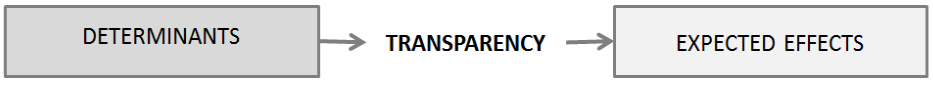


Figure 17 – Basic model of transparency

The literature surveyed included the top 25 journals with an average impact factor higher than 1,0 in the fields of Public Administration (PA) and Information Systems (IS), based on the 2016 Scientific Journal Rank (SJR – Scimago/Scopus). The keywords used for searching were kept broad by including "transparency" and "government" to avoid missing any relevant papers. The terms were only searched in the following fields: title, abstract and keywords.

The SLR was conducted in April 2017. Papers published in 2017 were excluded to avoid any confusion about which subset of papers from 2017 should be included. Searching on "transparency" and "government" returned a large number of papers, as transparency and government are also commonly used words outside the OGD field.

For example, chemistry and environmental science use these words within a different context, such as the transparency of glass or water. Some journals did not have a search function, and in these cases, the search was performed manually by accessing each paper and searching for the keywords using the search feature for an article. The results show that a lot of work is available in the financial domain. Budget transparency is based on the use of well-defined measures and financial reports and is guided by regulations. Articles that focused on budget transparency were excluded, as these do not represent the typical challenges faced by OGD (Conradie & Choenni, 2014; Harrison et al., 2012; Janssen et al., 2012).

This resulted in an initial selection of 173 papers. The number of papers was further reduced to 73 papers by only including papers published between 2007 and 2017. This enabled us to focus on the papers relevant to the OGD domain.

3.1.2.2 Content Analysis Method

After identifying the papers, content analysis was used to analyze the data. Content analysis can be defined as a "research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon, 2005, p. 1278). Content analysis classifies large amounts of text into a number of categories. In each paper, determinants and effects were identified by reading the entire paper and marking each determinant and each effect.

Both authors conducted the content analysis, and the results were discussed. Initially, all determinants and effects were marked to keep the list broad. This resulted in a long list of factors, for which synonyms were used in different papers to depict the same determinant or

effect. It was often impossible to determine whether the authors of one paper used definitions similar to those of a determinant or effect to those of other authors.

Many papers lack clear definitions and conceptualisation, and this complicates our SLR. It was assumed that they would use the same definitions in such situations. On several occasions, the same wording was used in reference to slightly different situations. For example, primary data can refer to raw data that has not yet been processed but also to data collected from the source and made ready for use. Instead of trying to define each element, we would label them both as primary data. In further research, we suggest that the determinants and effects should be defined in detail.

In many situations where the literature suggested a causal or correlated relationship, the explanatory mechanisms were poorly described or could be challenged. For example, it is often stated that access should be given to raw data. Although publishing raw data enables finding new insights and avoids predefining views, it might not per se result in transparency, as raw data might not easy to understand and use for most people. As such, the mechanisms for creating transparency might not work in all circumstances.

In further research, we suggest that empirical research should be conducted to explain the relationship between determinants, transparency and effects and to conduct experiments to investigate the causality of relationships.

3.1.2.3 Model Construction

Hsieh and Shannon (2005) suggest limiting the number of categories to ensure that they are easier to understand. This can be done by merging similar codes. The researchers merged the codes independently and then discussed them while comparing them to existing models in the literature. After studying the models, the content analysis resulted in 4 clusters of determinants and 1 cluster of expected effects of transparency.

Hsieh and Shannon (2005) recommend creating relationships between categories and subcategories based on their concurrence, influence, and consequences. Determinants and effects were cataloged accordingly, including whether a determinant had a positive or negative relationship with transparency. Factors having similar semantics were clustered. After reading the papers, a total of 42 determinants and 8 expected effects were found. Hsieh and Shannon suggest creating a diagram showing the relationships between the clusters and factors. A tree

diagram was created using the software Visual Understanding Environment (http://vue.tufts.edu/). This diagram is presented in figure 5 and was named 'The Window Theory', reflecting the determinants of transparency and their possible effects. The Window Theory is not a tested theory but merely a comprehensive overview of relationships suggested by the papers that were surveyed.

Our SLR depends on previously published research and the availability of these studies. The selection/exclusion criteria limit the number of surveyed papers in order to make it manageable. Some papers that might have been relevant were missed because they were not published in the outlets surveyed. We excluded literature concerning budgetary data as this is well-structured and well-defined, whereas this is not the situation for most OGD.

The authors had to interpret the work of others and to derive what they found to be relevant. Although a lot of work is available about transparency, there is no uniformity in the measurement and use of constructs. Furthermore, many studies did not provide a definition, or simply mentioned that determinants of effects were relevant. The studies surveyed are diverse and contain different interpretations of determinants and effects. The authors had to interpret the studies and make decisions on how to aggregate them. Sometimes, there were discussions about whether determinants should be merged into one or kept separate. We made this decision based on the relevance for OGD, which might not apply in other situations.

The Window Theory aggregates determinants and effects into a single mode but does not consider contextual factors. Some factors might only be relevant to a certain situation. In addition, the relevance and significance of factors could not be aggregated. There is only a small number of quantitative studies that employ different constructs, rendering it impossible to determine.

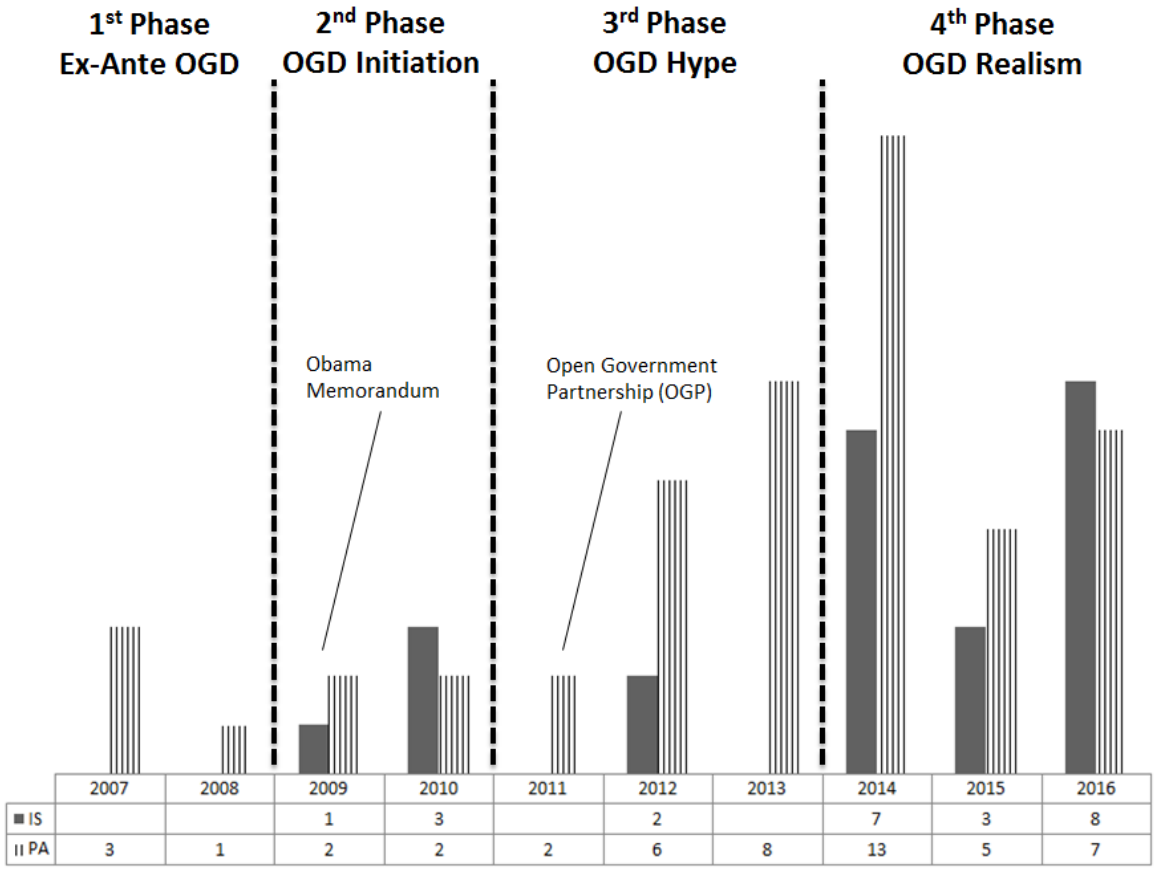
3.1.3 Findings

A content analysis was conducted on 73 papers to identify the determinants and effects of transparency. Of these 73 papers, 41 (56% of all 73 papers) used a quantitative research method. Only 16 papers include an explicit model associating determinants with transparency and its effects. None of these models contain more than 16 determinants and 3 effects. This confirms our original premise of the need for a unified model. The analyzes of papers revealed

4 stages of OGD for transparency development, which will be presented next. A presentation of the determinants and effects of transparency will follow this.

3.1.3.1 The evolution of OGD-driven transparency

Figure 18 shows the number of identified papers per year. The literature survey shows a steady increase in the number of publications. Most manuscripts were published in public administration (PA) literature (49 papers), accounting for 67% of the total papers, whereas 33% of papers were published in the field of information systems (IS) (24 papers). Digital government papers can be found in both the PA and IS domains, although journals such as Government Information Quarterly are classified in the IS domain.



	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
IS			1	3		2		7	3	8
PA	3	1	2	2	2	6	8	13	5	7

Figure 18 – Frequency of publications per year and area

From the SLR and content analysis, four phases of transparency research in OGD were identified by looking at discontinuity. Discontinuity signifies some form of observable change driven by an event, whereas continuity implies maintaining the status quo. Discontinuity helps

to find the boundaries for groups of things that are alike (Cook, 1996). The events resulting in discontinuity are derived from the content analysis of the papers, and their influence becomes visible only after a while.

The first discontinuity was the initiation of Obama's memorandum of understanding in 2009, which resulted in the rise of literature from mainly Western countries. Many papers referring to this memorandum of understanding were published in subsequent years. The second discontinuity is driven by the creation of Open Government Partnership in 2011. The final discontinuity is less visible as an event but clearer from the number of papers published. There is an increase in the number of papers in this field that are focussed on viewing possible benefits from OGD from a realistic perspective.

This phase is less explicit in the literature. The phase is found in papers showing the impediments and disadvantages, whereas the simple statement of advantages still continues.

1) Ex-ante phase (before 2009)

2) Initiation phase (2009-2011)

3) Hype phase (2011- 2013)

4) Realism phase (2014-now)

In the first phase, ex-ante open government data, researchers gave regular attention to transparency. OGD and transparency had no momentum and had gained only limited political attention. Many countries have Freedom of Information Acts (FOIAs) in place, through which citizens can request information (Petticrew & Roberts, 2008). FOIAs allow the public to ask for (partial) disclosure of unreleased information.

The second phase, OGD initiation, is hallmarked by Obama's Memorandum on 'Transparency and Open Government' published in 2009. Before the Memorandum, there was limited attention for OGD, even though transparency has always been an important topic for researchers in the field of government and public administration. The Memorandum encouraged active disclosure of public data. This Memorandum resulted in the development of the OGD portals (www.opendata.gov) which offered open data to the public.

This is followed by the OGD hype phase, which is characterised by an unbridled enthusiasm that resulted in many followers. Many local and central governments initiated open data portals to publish their data. In 2011, several countries worldwide joined the Open Government Partnership (OGP) to create Open Data Plans and increase the proactive disclosure of public data (Petticrew & Roberts, 2008). In 2013, the G8 Charter also declared its objective of opening up high-value data from the government (Hsieh & Shannon, 2005, p. 1279). This reflected a change from waiting for a specific request for papers or electronic forms (FOIA) to the pro-active opening of data, resulting in massive amounts of data sets in a friendly format (open data, linked data) with free and open access, with the intention of contributing to transparency.

In the fourth phase, named realism, a more feasible ambition has been set, and research has been initiated to advance OGD efforts to create transparency. After the initial excitement, scholars identified that data disclosure per se would not lead to proper levels of transparency, accountability, anti-corruption and the other expected effects.

The fourth phase was driven by improving practices and the need for research to deal with limited use. The massive disclosure of data raised the question of whether the opened data had fulfilled the expected promises, for example, and whether transparency enhanced trust, accountability transparency and engagement (Obama, 2009). The number of papers increased substantially, including papers from IS that had a technical perspective. At the same time the term ‘big data’ started to emerge, aimed at using data for the purpose of value creation (Piotrowski & Van Ryzin, 2007).

This research broadened the research in transparency across the board from technical to societal issues. The more technical papers demanded data in standardised, electronic formats using web services following Tim Berners-Lee’s 5 Star model (Obama, 2009). At the same time, the indiscriminate opening of data resulted in privacy concerns Coglianese (2009) and concerns about the proper use of data (Harrison, Pardo, & Cook, 2012). This more technical focus resulted in new concepts such as “transparency-by-design”, in which the opening of data is integrated within the design of information systems (Janssen, et al., 2017).

3.1.3.2 Determinants of Transparency

The content analysis revealed many diverse determinants of transparency for OGD. For each of the determinants, it was described whether the factor enabled or impeded transparency and

the explanatory mechanism, as shown in Table 10. The first column of Table 10 shows the total number of studies in which the determinants were found. The second column gives this number for only the quantitative studies. This is followed by the determinant name, the description of the determinant, and the description of the effect(s). The final two columns contain the references to the studies reviewed.

Table 10 – Types of possible effects of transparency

Expected Effect	**Description**	**Type of Effect**
Accountability	Transparency is expected to encourage governments to account better for their actions. In practice, accomplishing this can be difficult. Sometimes transparency is viewed as a sine qua non for accountability, although there is no agreement on this.	Positive
Trust and credibility	Transparency can result in more trust, trustworthiness, credibility and legitimacy. Occasionally, a decrease in trust is mentioned, resulting from this openness and seeing how the government works or does not work.	Positive and Negative
Civic engagement and Participation	Transparency should result in civic engagement as OGD should enable citizens to engage and participate by providing information. In the literature, there is discussion about the limited level of engagement.	Positive
Efficiency	Transparency can help to create efficiency and reduce costs. For example, an improvement in public procurement, the allocation of budget, and healthcare. Conversely, investments are needed to help create transparency.	Positive
Governance and political turnout	Transparency should result in better governance and political turnout (votes). This is heavily debated in recent literature (fake news).	Positive
Anti-corruption and fraud	Transparency should result in less corruption by reducing bias and the misuse of political power and public funds.	Positive
Overcoming information asymmetry	Transparency should result in overcoming information asymmetry, but some data might not be accurate (low information quality) or manipulated.	Positive And Negative
Privacy	Transparency can affect privacy and even violate data protection legislation. For example, the European data protection act requires that civil servants' identities should not be revealed. This imposes limits on transparency.	Negative

Figure 19 shows the determinants and the number of papers mentioning the determinants. From this figure, it becomes clear that some factors were only mentioned a few times, such as urban area and types of organisation, while open and free access were mentioned most frequently. Some factors are viewed as conditions for creating transparency, whereas others are enablers or inhibiters of transparency. For example, a condition for transparency is completeness, as

missing information results in less transparency. Metadata is a typical enabler for the creation of transparency, whereas lack of usability inhibits transparency. Making a distinction between quantitative and qualitative studies enabled us to compare whether different determinants were found in both kinds of studies. Only 'technology Literacy' was found in 2 qualitative studies, but not in the quantitative study, whereas all the other determinants were found at least once. Although, there are some variations in the number of occurrences, the differences between quantitative and qualitative studies are limited.

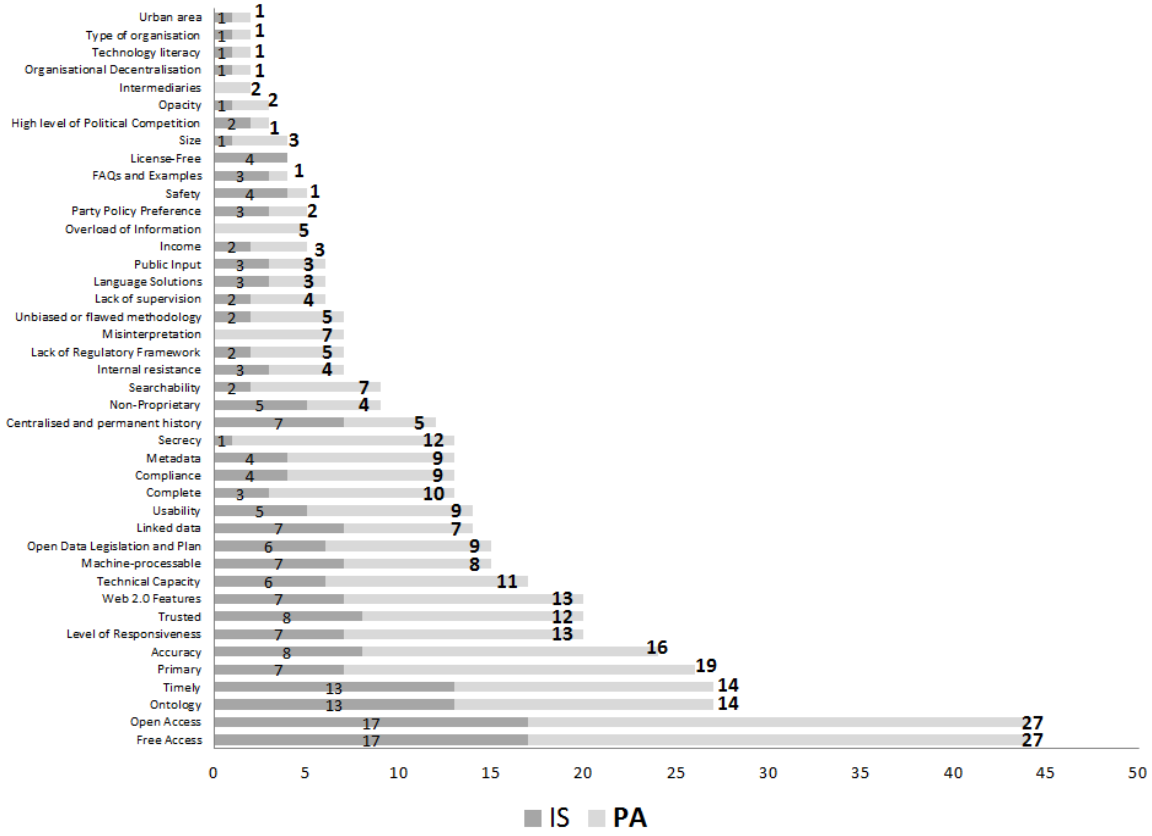


Figure 19 – Frequency of determinants per area

The variety and diversity of determinants also show the complexity of creating transparency. It is both challenging and costly to satisfy all factors and may even be unnecessary (Roberts, 2002).

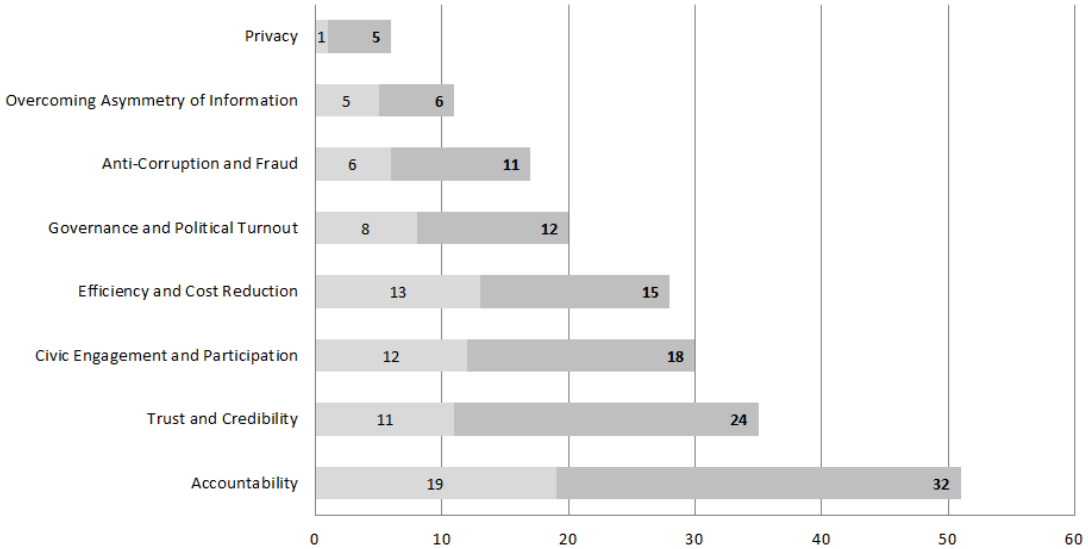
The question remains: which of these factors are conditions that have to be satisfied, which are most influential, and which are nice to have. The SLR does not shed a lot of light on this, as

there is no consensus in the literature. Different authors suggest that different factors are of importance.

Determinants were found to have different influences on the desired effect. For example, 35 determinants out of 42 were found to have an effect on anti-corruption and fraud. In contrast, only 27 determinants were found to be related to accountability. The SLR shows that creating transparency in order to reduce fraud is influenced by other determinants than creating transparency for accountability. This suggests that the form of transparency needed is dependent on the intended effects of transparency.

3.1.3.3 Expected Effects of Transparency

Transparency can offer many benefits but also some unintended or even adverse effects, as listed in Figure 20. In the literature, positive effects are emphasized, while negative effects have been given less attention. Examples of negative effects include violation of privacy due to the release of data containing privacy-sensitive information (T. M. Harrison et al., 2012), or the fact that the release of large amounts of OGD can result in information overload and errors (Grimmelikhuijsen, 2012a). The achievement of either negative or positive effects is highly dependent on how information is released and shared, as shown in Table 11 – List of determinants of OGD-enabled transparency.



	Accountability	Trust and Credibility	Civic Engagement and Participation	Efficiency and Cost Reduction	Governance and Political Turnout	Anti-Corruption and Fraud	Overcoming Asymmetry of Information	Privacy
IS	19	11	12	13	8	6	5	1
PA	32	24	18	15	12	11	6	5

Figure 20 – Overview of expected effects and areas

Table 11 – List of determinants of OGD-enabled transparency

Freq. (total)	Freq. for quantitative studies only	Determinant name	Determinant description	Description of effect	Sources all studies (Papers ID can be found in Table 12)	Sources for Quantitative Studies (Papers ID can be found in Table 12)
44	24	Free Access	Data sets can be directly accessed at no cost.	Data availability is a condition for transparency.	1, 3, 4, 5, 9, 10, 11, 20, 21, 23, 24, 25, 27, 29, 30, 31, 32, 35, 36, 37, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 60, 62, 63, 64, 65, 66, 67, 72, 73.	1, 3, 4, 5, 11, 21, 24, 31, 32, 37, 42, 44, 46, 47, 50, 53, 54, 55, 62, 63, 65, 66, 67, 73
44	24	Open Access	Data sets can be used, re-used and distributed without any restrictions.	The easier the use of datasets the higher the levels of transparency that can be created.	1, 3, 4, 5, 9, 10, 11, 20, 21, 23, 24, 25, 27, 29, 30, 31, 32, 35, 36, 37, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 60, 62, 63, 64, 65, 66, 67, 72, 73.	1, 3, 4, 5, 11, 21, 24, 31, 32, 37, 42, 44, 46, 47, 50, 53, 54, 55, 62, 63, 65, 66, 67, 73.
27	10	Ontology	Definitions and relationships between data elements to ease usage by machines and the	The better datasets are described the higher the level of transparency.	3, 4, 8, 15, 27, 28, 29, 30, 33, 34, 35, 44, 48, 49, 51, 52, 53, 54, 56, 60, 63, 64, 65, 66, 68, 70.	3, 4, 33, 44, 53, 54, 63, 65, 66, 68.

			understanding of users.			
27	16	Timeliness	Data sets are up-to-date.	Up-to-date data is a condition for transparency.	3, 4, 14, 16, 25, 28, 30 ,31, 35, 38, 44, 47, 51, 52, 53, 55, 57, 59, 60, 61, 62, 63, 66, 71, 72, 73.	3, 4, 14, 16, 31, 44, 47, 53, 55, 59, 61, 62, 63, 66, 71, 73.
26	13	Primary	Data sets are disclosed in non-aggregated form or are preselected.	Access to raw data is a condition for transparency.	4, 10, 14, 18, 21, 22, 26, 28 ,29, 30, 35, 39, 44, 45, 46, 48, 51, 52, 54, 57, 59, 63, 64, 71, 73.	4, 14, 21, 22, 26, 39, 44, 46, 54, 59, 63, 71, 73.
24	14	Accuracy	The correctness of datasets.	More accurate data will result in higher levels of transparency.	4, 8, 11, 15, 27, 29, 38, 39, 42, 44, 45, 47, 48, 52, 53, 54, 56, 59, 61, 63, 71, 72, 73.	4, 11, 38, 39, 42, 44, 47, 53, 54, 59, 61, 63, 71, 73
20	11	Trusted	The conviction among users that data sets are reliable and accurate.	Higher levels of trust in data will result in the use of data for creating transparency.	8, 17, 18, 26, 32, 38, 40, 44, 45, 47, 51, 52, 53, 55, 59, 60, 61, 71, 72, 73.	26, 32, 38, 40, 44, 47, 55, 59, 61, 71, 73.
20	13	Web 2.0 Features	Inclusion of features enabled by the change from static websites to dynamic (e.g., videos, audio, animations) and user content generation (e.g., social media such	Web 2.0 features will make the use of data for transparency more attractive.	16, 17, 19, 20, 22, 23, 30, 37, 45, 50, 55, 57, 59, 61, 62, 63, 66, 67, 72, 73.	16, 19, 22, 37, 50, 55, 59, 61, 62, 63, 66, 67, 73.

			as Facebook and Twitter).			
19	14	Openness appetite	Governmental level of desire to disclose and structure OGD.	The more data is opened the higher the level of transparency.	37, 39, 44, 45, 47, 48, 50, 55, 56, 57, 61, 62, 63, 64, 66, 67, 68, 69, 72.	37, 39, 44, 47, 50, 55, 61, 62, 63, 66, 67,68, 69, 72.
17	13	Technical Capacity	Having the expertise and skills to disclose data for creating transparency.	Expertise and skills to disclose datasets is a condition for releasing data for creating transparency.	33, 37, 38, 39, 42, 47, 49, 54, 55, 56, 57, 59, 61, 62, 66, 67, 72.	33, 37, 38, 39, 42, 47, 54, 55, 59, 61, 62, 66, 67.
15	5	Machine-processable	Data sets are structured in such a way that they can be read and processed by machines.	Automatically processable data will result in higher levels of transparency.	15, 21, 27, 28, 30, 45, 48, 50, 51, 53, 57, 59, 60, 64, 65.	21, 50, 53, 59, 65.
15	6	Open Data Legislation and Plan	A set of legislation and plans to create, maintain and update the OGD.	The more openness is enabled by legislation the higher the level of transparency.	2, 3, 10, 31, 43, 45, 49, 57, 62, 64, 66, 67, 68, 70.	3, 31, 62, 66, 67, 68.
14	7	Usability	The degree to which OGD portals are able to be used or are fit for use by citizens. Also described as user experience and	Higher usability will result in higher levels of transparency.	4, 9, 11, 14, 27, 29, 37, 45, 49, 52, 62, 63, 72, 73.	4, 11, 14, 37, 62, 63, 73.

			user interface needs.			
14	4	Linked Data	Method that structures data to be interconnectable.	Linking datasets will result in higher levels of transparency.	28, 30, 35, 44, 45, 48, 49, 51, 53, 60, 63, 64, 65, 70.	44, 53, 63, 65.
13	9	Complete	Data sets should contain all elements.	Complete datasets will result in higher levels of transparency.	14, 22, 26, 27, 29, 39, 45, 46, 59, 63, 64, 71, 73.	14, 22, 26, 39, 46, 59, 63, 71, 73.
13	8	Compliance	Adherence of systems to standards.	Use of standards will result in higher levels of transparency.	2, 3, 31, 43, 52, 54, 57, 62, 64, 67, 68, 69, 71.	3, 31, 54, 62, 67, 68, 69, 71.
13	4	Metadata	Metadata are the data that explain the meaning of the data within a dataset.	Datasets described using metadata will result in higher levels of transparency.	17, 28, 29, 48, 49, 51, 54, 56, 60, 64, 65, 71, 73.	54, 65, 71, 73.
13	5	Secrecy	Not disclosing information.	Keeping data secret will block transparency.	6, 12, 27, 29, 36, 38, 43, 46, 47, 48, 54, 70, 71.	38, 46, 47, 54, 71.
12	6	Centralised and permanent history	Having a centralised, single OGD portal with permanent history to ensure that data from the	Access to historical data and experiences with data will result in higher levels	33, 35, 39, 47, 48, 51, 54, 57, 59, 60, 64, 66.	33, 39, 47, 54, 59, 66.

			past can still be accessed.	of transparency.		
9	5	Searchability	A feature that allows citizens to find what they are looking for.	The ability to search for datasets in portals will result in higher levels of transparency.	4, 9, 12, 19, 21, 28, 37, 49, 73.	4, 19, 21, 37, 73.
8	1	Non-proprietary	Data formats are not the property of any institution (as this might limit use).	Using non-proprietary data will result in higher levels of transparency.	28, 45, 48, 51, 52, 53, 57, 60.	53.
7	1	Internal resistance	Bureaucratic and political reluctance to open data.	Higher levels of resistance will result in releasing less data for creating transparency.	9, 10, 36, 43, 70, 71, 72.	71.
7	6	Lack of Regulatory Framework	Legislation guiding or requiring the disclosing of data.	Lack of legislation for opening data will result in less datasets for creating transparency.	56, 62, 63, 67, 68, 69, 71.	62, 63, 67, 68, 69, 71.
7	2	Misinterpretation	Having a wrong understanding of the meaning of the data.	Inability to interpret or wrong interpretation of data will result in lower levels of transparency.	7, 17, 18, 19, 40, 56, 57.	7, 40.

7	5	Unbiased or flawed methodology	Biased sensors or flawed statistical methodologies providing incorrect data or outcomes.	Biased and flawed data will result in lower levels of transparency.	4, 47, 52, 54, 61, 68, 72.	4, 47, 54, 61, 68.
6	6	Lack of supervision	Management supervision over the handling of disclosure to ensure opening of data.	A lack of management supervision when opening data will result in lower levels of transparency.	62, 63, 67, 68, 69, 71.	62, 63, 67, 68, 69, 71.
6	4	Language Solutions	Supporting different levels of language proficiency or using ideograms to ease understanding.	Language support will result in higher levels of transparency.	29, 31, 35, 55, 65, 67.	31, 55, 65, 67.
6	2	Public Input	Different types of access to support a variety of purposes.	Providing different ways to access data will result in higher levels of transparency.	13, 20, 29, 51, 53, 60.	13, 53.
5	4	Government budget	Bigger budgets allow more opportunity to create, maintain and update an OGD public policy.	Higher budgets for opening data will result in higher levels of transparency.	25, 33, 66, 67, 68.	33, 66, 67, 68.
5	3	Information Overload	Offering much more data than people can process.	Opening non-relevant data will result in lower levels	7, 18, 19, 28, 40.	7, 19, 40.

				of transparency.		
5	3	Party Policy Preference	The political preferences regarding OGD data policy.	Ruling politicians favouring open data will result in higher levels of transparency.	46, 56, 63, 64, 66.	46, 63, 66.
5	3	Security	Data should be protected from manipulation.	Higher levels of security will result in higher levels of transparency.	3, 51, 53, 59, 60.	3, 53, 59.
4	2	FAQs and Examples	Frequently Asked Questions about recurring problems and examples to illustrate good practices.	Support will result in higher levels of transparency.	3, 36, 59, 64.	3, 59.
4	1	License-Free	No license restricting open data use.	Non-restricting licenses for using datasets will result in higher levels of transparency.	51, 52, 53, 60.	53.
4	3	Size	The number of employees of a public organisation.	The larger the public organisation the more datasets will be released and the higher	25, 37, 67, 68.	37, 67, 68.

				the level of transparency.		
3	3	Political pressure	Influence of other countries on initiating OGD.	Political pressure from other countries will result in higher levels of transparency.	32, 33, 67.	32, 33, 67.
3	1	Opacity	Disclosing data in such a way that it cannot be used.	Opacity will result in lower levels of transparency.	46, 49, 60.	46.
2	1	Intermediaries	Intermediaries processing open data and making them available using software applications and websites.	The more intermediaries who are involved in processing open data the higher the levels of transparency.	45, 68.	68.
2	2	Organisational Decentralisation	The level of (de)centralisation of decision-making.	Centrally organised organisations will release more data, resulting in higher levels of transparency.	32, 67.	32, 67.
2	0	Technology Literacy	The ability to understand and use OGD.	The more literate citizens are, the higher the levels of transparency.	9, 29.	-

2	1	Type of organisation	Governments are diverse and range from small to large. Also activities might be outsourced to private parties, who have no obligation to create transparency.	The more complex the administration is, the lower the level of transparency.	32, 56.	32
2	2	Urban area	Geographical location of a government.	Governments in densely populated areas will create higher levels of transparency.	33, 69.	33, 69

Additionally, discussions can be found in the literature about what happens when inaccurate data is shared and the issue of who is accountable for decisions based on inaccurate data (French, 2011). Finally, some work mentions both positive and negative effects of transparency. Table 12 summarises the descriptions and the types of effects identified in the SLR of transparency.

The most mentioned effects of transparency are headed by accountability (51 papers), followed by trust and credibility (35 papers), civic engagement and participation (30 papers), efficiency and reduction of costs (28 papers), governance and political turnout (20 papers), anti-corruption efforts (17 papers), error and asymmetry of information (11 papers) and privacy (6 papers). The papers often mention more than one expected effect. For example, Janssen, Matheus, and Zuiderwijk (2015) found that transparency can increase accountability and can help anti-corruption efforts. That an effect is mentioned does not mean that the effect is also achieved. There is still discussion about whether transparency results in accountability (Park & Gil-Garcia, 2017).

Most of the literature is focused on the positive effects of transparency. The number of papers mentioning negative effects is low: we found that only 16% of the papers (12 out of 73) mentioned negative effects. The average of expected effects is in the surveyed PA literature 62% higher than in IS. In PA there is a wider discussion of the effects of transparency, whereas in IS literature, transparency is often viewed as the outcome (dependent variable). Not surprisingly, discussions in IS are more focused on technical solutions and the determinants of transparency.

Table 12 – The overall classification of the papers

ID	Reference	Area	Year	ID	Reference	Area	Year
1	Piotrowski and Van Ryzin (2007)	PA	2007	38	S. G. Grimmelikhuijsen and Meijer (2014)	PA	2014
2	Blomgren (2007)	PA	2007	39	Hollyer, Rosendorff, and Vreeland (2014)	PA	2014
3	Pina, Torres, and Royo (2007)	PA	2007	40	de Fine Licht (2014)	PA	2014
4	Rawlins (2008)	PA	2008	41	De Fine Licht, Naurin, Esaiasson, and Gilljam (2014)	PA	2014

5	Relly and Sabharwal (2009)	IS	2009	42	Bauhr and Grimes (2014)	PA	2014
6	Coglianese (2009)	PA	2009	43	Gregory Michener (2015)	PA	2014
7	Grimmelikhuijsen (2009)	PA	2009	44	Tan (2014)	PA	2014
8	Furtado et al. (2010)	IS	2010	45	Zyl (2014)	PA	2014
9	Bertot et al. (2010)	IS	2010	46	Cross (2014)	PA	2014
10	Jaeger and Bertot (2010)	IS	2010	47	Auger (2014)	PA	2014
11	McNeal and Hale (2010)	PA	2010	48	Bertot et al. (2014)	PA	2014
12	Worthy (2010)	PA	2010	49	Barry and Bannister (2014)	PA	2014
13	French (2011)	PA	2011	50	Ganapati and Reddick (2014)	PA	2014
14	S. Grimmelikhuijsen (2011)	PA	2011	51	Janssen and van den Hoven (2015)	IS	2015
15	Y.-C. Chen (2012)	IS	2012	52	Rui Pedro Lourenço (2015)	IS	2015
16	Bonsón, Torres, Royo, and Flores (2012)	IS	2012	53	Abu-Shanab (2015)	IS	2015
17	Evans and Campos (2013)	PA	2012	54	Kashin, King, and Soneji (2015)	PA	2015
18	Grimmelikhuijsen (2012a)	PA	2012	55	Porumbescu (2015)	PA	2015
19	Bridges, Appel, and Grossklags (2012)	PA	2012	56	Reynaers and Grimmelikhuijsen (2015)	PA	2015
20	Chun and Cho (2012)	PA	2012	57	B. Worthy (2015)	PA	2015
21	Munson et al. (2012)	PA	2012	58	Grimmelikhuijsen and Klijn (2015)	PA	2015
22	Van Ryzin and Lavena (2013)	PA	2013	59	Venkatesh, Thong, Chan, and Hu (2016)	IS	2016
23	Meijer (2013)	PA	2013	60	Shaikh and Vaast (2016)	IS	2016
24	Themudo (2013)	PA	2013	61	Z.-J. Chen, Vogel, and Wang (2016)	IS	2016
25	Rodríguez Bolívar, Alcaide Muñoz, and López Hernández (2013)	PA	2013	62	Navarro-Galera et al. (2016)	IS	2016
26	Deng, Peng, and Wang (2013)	PA	2013	63	Gandía et al. (2016)	IS	2016
27	Greg Michener and Bersch (2013)	PA	2013	64	Muñoz-Soro, Esteban, Corcho, and Serón (2016)	IS	2016
28	Rui Pedro Lourenço (2013)	PA	2013	65	Höffner, Martin, and Lehmann (2016)	IS	2016

29	Owen, Cooke, and Matthews (2013)	PA	2013	66	Guillamón, Ríos, Gesuele, and Metallo (2016)	IS	2016
30	Halachmi and Greiling (2013)	PA	2013	67	S. G. Grimmelikhuijsen and Feeney (2016)	PA	2016
31	Harrison and Sayogo (2014)	IS	2014	68	David-Barrett and Okamura (2016)	PA	2016
32	Cuadrado-Ballesteros (2014)	IS	2014	69	Peiffer and Alvarez (2016)	PA	2016
33	Angst, Agarwal, Gao, Khuntia, and McCullough (2014)	IS	2014	70	Graham, Gooden, and Martin (2016)	PA	2016
34	Packer, Diochnos, Rovatsos, Gal, and Moreau (2014)	IS	2014	71	Ríos, Bastida, and Benito (2016)	PA	2016
35	Martin, Stadler, Frischmuth, and Lehmann (2014)	IS	2014	72	Reddick, Chatfield, and Brajawidagda (2016)	PA	2016
36	Villeneuve (2014)	IS	2014	73	Song and Lee (2016)	PA	2016
37	Yavuz and Welch (2014)	IS	2014				

3.1.4 The Window Theory

The disclosure of OGD should result in transparency by creating a window on governmental functioning. Our aim is to develop a comprehensive model of possible determinants that enable or impede transparency for OGD and the efficacy of transparency. Hence, the Window Theory is not a tested theory but a comprehensive account of the determinants and effects of transparency. We label it as a theory as it is a supposition of factors aimed at explaining how transparency can be created and what its effects are. The SLR provides a long list of factors but gives little insight into which factors are conditions, which would be good to include, which are nice to have, and which have hardly any influence. Nor does literature give insight into the significance of factors or whether they are antecedents or moderators. Therefore, we opted to create a comprehensive model that contains all determinants deemed to be relevant, as shown in Figure 21. The model shows the relationship between 42 determinants and 8 possible effects and unifies various views on transparency within a single model. Some relationships are likely to be strong, whereas others will be weak. The determinants are grouped into 4 clusters, and the effects are grouped into 1 cluster based on the content analysis.

The resulting model is named ‘The Window Theory’ as OGD should provide a window to see what has happened or what is happening within the government. The frame of the window

determines if it can be used. This is expressed by the ‘organisational characteristics’ and ‘system quality’ clusters. In the window, the ‘data quality’ cluster determines whether the data becomes transparent. Governments are creating a window to the outside world by disclosing OGD in portals. The determinants show that simply disclosing OGD is insufficient and that many determinants need to be addressed. Addressing the determinants can help governments arrive at OGD portals that create transparency.

As a metaphor, a window is very effective during the day, but nothing can be viewed at night without artificial light. The model shows that the night will remain dark unless users are guided towards the data. The same data that creates greater transparency for the expert offers less transparency for someone with different access conditions and a lack of knowledge about how to use the data. This aspect is addressed by the ‘individual characteristics’ cluster.

The many determinants and their relationship suggest that transparency should be viewed as a complex, continuous construct rather than a dichotomous construct. Transparency is hard to express on a single scale, ranging from zero to full transparency. Furthermore, the two extremes (zero or full transparency) are less likely to occur as there will always be some degree of transparency.

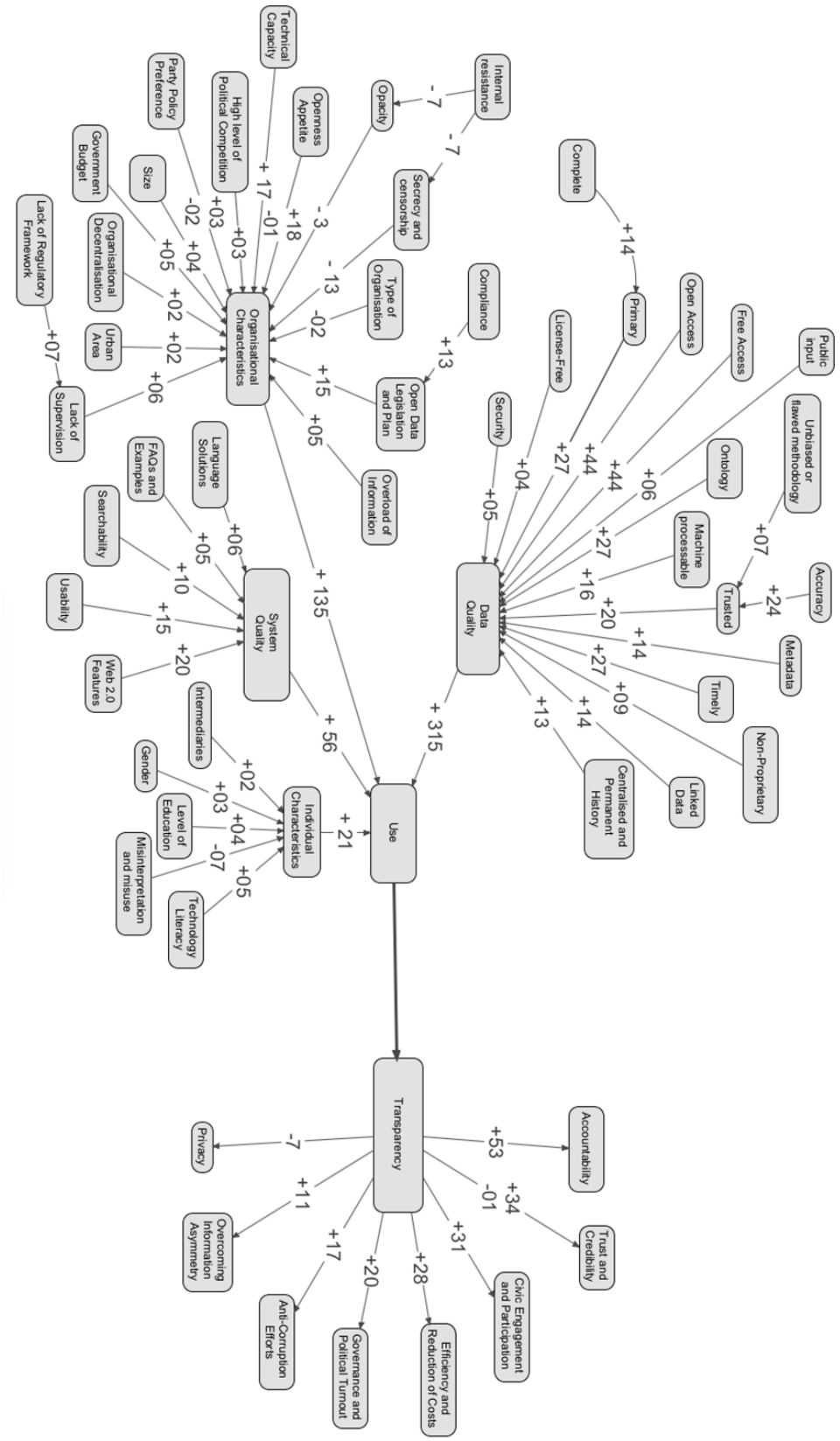


Figure 21 – The Window Theory: Determinants and expected effects of transparency enabled by open government data

Four main clusters of determinants were identified, as shown in Figure 5. In the figure in parentheses is the total count of the individual times that the groups of factors are mentioned:

- **Data quality** (315): The aspects determining the quality of the OGD published
- **System Quality** (56): Characteristics of systems that are used for publishing data
- **Organisational characteristics** (135): A variety of organisations that publish the
- **OGD Individual characteristics** (21): Variety of users of the OGD

Data quality determinants were found most frequently. In almost all the IS papers data quality is mentioned. One of the explanations for this is that data quality plays a major role in the Delone and McLean (2003) success model for information systems, one of the most cited papers in the IS field. This model also includes system quality dimensions, but this is less evident in the literature surveyed. In most IS literature, system quality is not viewed as being directly connected to transparency, and in the first three phases (figure 2), system quality factors are rarely mentioned in the literature. Only recent literature, mostly after 2014, acknowledges that system quality also influences transparency. The simple release of data was not found to be sufficient, and characteristics such as usability, performance (for real-time data provision), and comparability are found to be important for creating transparency (Bertot et al., 2014; Janssen & Kuk, 2015; Muñoz-Soro et al., 2016).

According to the literature, both organisational and individual characteristics play a role. Organisational characteristics are diverse and range from leadership to geographic (urban) area. Some organisations are better prepared and are better equipped to create transparency, while others lack these capabilities (Barry & Bannister, 2014; Guillamón et al., 2016; Ølnes, 2016). Political orientation also plays a role, as some political movements favour disclosure and transparency, whereas others do not (Gandía et al., 2016; Reynaers & Grimmelikhuijsen, 2015).

Users can be diverse, possessing varying individual characteristics, as depicted in figure 5. Users can be citizens (J.C. Bertot et al., 2010), company employees (David-Barrett & Okamura,

2016), public servants from other organisations (Angst et al., 2014), persons working in public-private organisations delivering public services (Reynaers & Grimmelikhuijsen, 2015), and employees of supra-national organisations (Cross, 2014). User characteristics such as technology literacy (Owen et al., 2013), level of education (S. G. Grimmelikhuijsen & Meijer, 2014), and gender (Bauhr & Grimes, 2014) influence the ability to make use of data. Some users only read news from intermediaries such as journalists, who themselves used OGD and FOI (Piotrowski & Van Ryzin, 2007), or use applications providing a predefined view on the data.

The Window Theory contains a long list of factors, but not all are always relevant for every situation. Our analyses suggest that the determinants depend on the desired effects of transparency and the context. Other factors were deemed relevant in the literature depending on the aim of transparency. Metaphorically speaking, different windows are needed for different purposes. This implies that for a certain situation, the Window theory can be used as the basis, but determinants and factors need to be selected to make them context-specific. This can also explain the many different models and lack of uniformity in the literature review. Depending on the context, user group and objective are important.

3.1.5 Discussion and Further Research

The Window Theory is based on the idea that OGD creates a window through which to observe government and overcome information asymmetry. There are two parties: those who look through the window and those who determine what is showcased in the window. OGD is showcased by governments. What is showcased reflects the intentions of the publisher but might not fit the needs and desires of what the public wants to see. Glass is often set in a window to protect against weather conditions but blurs the view. If not properly done, dust can form on the window. The physical characteristics of glass allow it to transmit, reflect, and refract light. In a similar vein, many transparency efforts only reveal a part of the full picture, reflecting what you want to see and not necessarily resulting in transparency. The quality of glass can be enhanced by cutting and polishing. Policy-makers and designers can use the comprehensive model to enhance transparency by addressing the determinants.

Transparency is multidimensional (Furtado et al., 2010), making it hard to define and create. Opening too much data without proper guiding mechanisms can only result in blurring the window and actually preventing the viewer from seeing what is happening inside the

government. This is labelled as “disclosure overload” in compliance and regulation literature (Hoitash & Hoitash, 2017). In this field, regulators ask for more and more information and reporting, which in turn is hardly used by the agencies in charge of the supervision. The volume of disclosure is not helpful in creating transparency and might even have the opposite effect. Removing the blur and presenting the public with easily understandable information in manageable form requires significant investment. The level of transparency is influenced by the type of window (data, quality) but also by who is looking (stakeholders), from where (context, position in charge), when (time) and how people look through the same window (objective use of data). Research is needed on how to lower the public involvement threshold and avoid disclosure overload.

The most comprehensive model addressing transparency that we identified using SLR defined 16 determinants and a maximum of 3 effects. Our model is focused on transparency enabled by OGD and is the most comprehensive, as the Window Theory lists 42 determinants and 8 expected effects. The disadvantages of our model are related to the many factors involved. Furthermore, the relative importance of each factor is unclear. The number of works on the subject of data quality and transparency was found to be high (+315), however, research is not evenly distributed across all factors.

The findings show that the information systems literature focuses more on determinants such as information quality and system quality dimensions, whereas public administration research focuses more on aspects such as trust and accountability. Give the focus of the domains this comes not as a surprise. However, these fields seem to converge, a development expressed in current work that attempts to bridge the gap between them (Bertot et al., 2014; Janssen & Kuk, 2015; Muñoz-Soro et al., 2016). The work combines elements from the public administration and information systems field and might be best characterised as the ‘digital government’ domain. It is likely that these domains will learn from each other and will start to influence each other. More research is needed to understand the influence and relevance of the factors involved, particularly whether a determinant is a condition, a facilitator or an inhibiter of transparency. The SLR shows that determinants' importance depends on the intended effect. Also, the significance of factors might be dependent on the type of OGD and the policy and institutional context. We excluded literature concerning budgetary data as this is well-structured and, well-defined and guided by regulations, whereas this is not the case for most OGD. A comparison between OGD and budgetary data can be made in further research. We

recommend using the Window Theory in future research to identify the most influential factors and develop a parsimonious model. Also, we recommend the development of contextual models that provide insight into the deeper structure and meaning of determinants and effects.

The complexity and the quantity of determinants challenge the assertion that the opening of data will result in transparency. The simple opening of data is unlikely to result in transparency per se. There is no single recipe for creating transparency. Some determinants might have a larger influence than others, and some might be conditions for creating transparency, but the literature does not provide a clear view of this. Our review suggests that transparency is context dependent and that the creation of transparency depends on what information is supplied, how it is provided and on its users, who need to have the knowledge and the capabilities to understand the open data. Alongside the data, details of the context have to be shared to allow the public to interpret the data.

One of the main ideas behind OGD is to create transparency for the general public. The determinants show that this might be more complicated than initially expected. Individual characteristics hinder the use of OGD to create transparency. Hence, the creation of mass-transparency is not easy to accomplish. The requisite individual characteristics are only found in a relatively small group; consequently, transparency is only created for the happy few. By taking the comprehensive list of determinants into account, the Window Theory can be used to facilitate a much broader audience.

Budget transparency was left out in our SLR, as this is a well-defined domain with a long track record in the standardisation of financial reporting. A subset of determinants and effects are likely relevant to budget transparency. Further research can evaluate whether all factors for budget transparency can be found in the Window Theory and whether a specific subset can be derived for this domain.

The Window Theory started with transparency and investigated both determinants and effects. Use is a central construct influencing transparency. Without the use of OGD there can be no creation of transparency. Use is a complicated construct and often requires participation and interaction among people. Citizen's engagement and participation also resulted from transparency (Harrison & Sayogo, 2014). Although both elements are related to engagement and participation, engagement and interaction are needed as determinants to make sense of the OGD. Once OGD is understood and transparency is created, engagement and participation can

be achieved to influence policy-making and political decisions. As such, these two elements should be separated and not combined. Making sense of OGD by interaction and using analytical capabilities is an area of limited research. We only found a few studies that referred to this aspect, which might be the most important element in creating transparency once OGD is published in a suitable format.

Transparency covers a broad range of concepts and elements. In the SLR we encountered the fact that in the literature different names are used to express similar determinants and effects, although there are sometimes slightly different interpretations. This complicated the aggregating of the determinants and effects and the comparison. For example, completeness and incompleteness refer to a variable showing various degrees of completeness, ranging from zero data to complete data. Another example of using different names for similar concepts is the use of the words ‘trust’ and ‘reliability’ of data. Both can refer to data that are defined clearly enough to arrive at similar results, but reliability can refer to being continuously available and trust can refer to whether the data provider can be trusted. Trust is often viewed as being broader than reliability, with reliability as one of its properties. The terms' use seems to depend on the community: PA prefers to use trust, while IS typically prefers to use the term reliability. A limitation of our work is that we merged some of the concepts into one to keep the number of determinants manageable. Although the determinants provide a solid starting point, their descriptions in table 2 should be refined in further research.

The literature suggests that full transparency is hard to achieve but is also unnecessary (Roberts, 2002). Given the many factors involved, it seems to be better to discuss the level of transparency. However, the question remains of how much transparency is sufficient, which might depend on the situation at hand and the users. We do not need to scrutinise public servants at all times during their work, and the privacy of public servants should be respected. A level of trust is necessary, that the public servant can do his job properly and the prevailing culture should ensure that public services are executed professionally.

Politicians want to create transparency using OGD, whereas public servants still need to realise this. The Window Theory shows that there is a large discrepancy between the intuitive notion and the actual realisation. Many open data portals release OGD, but this does not always create transparency, as determinants of transparency have not been considered or have not been handled in a suitable way. It is too simple to assert that transparency is automatically created

by realising OGD. We suggest to use the determinants as a form of checklist to evaluate transparency initiatives and to use this evaluation to broaden the view on creating transparency. Also, the impact of creating transparency can be more diverse, as the effects show.

3.1.6 Conclusions

The creation of transparency using OGD has become one of the key areas in digital government research. The literature review showed an increase in papers related to open data and transparency over time, but also fragmentation, as authors look at different determinants and effects of transparency. There was no comprehensive overview, making it hard to determine which factors should be considered important. Many studies only take a few actors into account.

We unraveled the concept of transparency for OGD by finding 42 determinants influencing transparency and 8 types of expected effects. The determinants were clustered into four factors: data quality, system quality, organisational characteristics and individual characteristics.

The Window Theory is a unifying model containing these determinants and effects of transparency. By creating an overview, the Window Theory unifies the various works in this area originating from public administration and information system publications. The nature and objectives of transparency are likely to differ per situation, and the model can be used to select the determinants that are appropriate to a given context. Factors might be more or less important depending on the situation.

Our SLR depends on previously published research and the availability of these studies, which limits its outcome. We had to interpret the descriptions in the literature, which were often ill-defined and ambiguous and sometimes suggested relationships that lacked a clear description of the explanatory mechanisms. We were not able to discriminate between antecedents and moderators. Empirical research is needed to develop explanations of the relationship between the determinants of and effects of transparency. In further research, a distinction between factors such as conditions, facilitators, or inhibitors of transparency should be made. The determinants with the highest explanative power can be selected to develop a parsimonious model. Furthermore, we recommend further research to focus more on what data is needed and in what form, and on the interpretation and sense-making of data.

3.2 Paper 2 – Digital Transparency and the Usefulness for Open Government

This chapter is based on Matheus, R., Faber, R., Ismagilova, E., & Janssen, M. (2023). Digital transparency and the usefulness for open government. International Journal of Information Management, 73, 102690.

The first author conducted the literature review, designed and conducted the data collection and analysis process, and wrote the manuscript. The co-authors provided feedback on the data collection and analysis process and earlier versions of the manuscript. We thank the anonymous reviewers for their peer review, which contained many insightful comments and recommendations.

The paper can be accessed and freely downloaded at https://www.sciencedirect.com/science/article/pii/S0268401223000713.

Abstract

Open Government efforts are criticized for providing limited value. Instead of looking at a value, we investigate the usefulness of web-based open government portals and apps. Specifically, we investigated the relationship between digital transparency and usefulness. We analyzed perceived digital transparency and usefulness in a survey of 112 respondents using Partial Least Square (PLS) and Structural Equation Modelling (SEM). The results show that perceived functionality, transparency, and efficiency influence usefulness but that functionality of apps and efficiency are more important than transparency. Usefulness can be created without having high levels of transparency, as the public wants answers to their questions. Apps should be designed for efficient use, as users have limited time and resources. Apps having pre-defined functional views can be useful to provide quick insight but might limit transparency by not offering other views and insights. Opening raw data using portals can provide higher levels of transparency, although more time and effort are needed to analyze. Both portals providing access to raw data and apps having pre-defined views are needed for open government and transparency as they serve other stakeholder groups and purposes.

3.2.1 Introduction

Open government is a global phenomenon driven by the need to boost innovation, create transparency and improve accountability (John Carlo Bertot, Paul T. Jaeger, & Justin M. Grimes, 2010; K. Janssen, 2011). The creation of digital transparency is often viewed as one of the key objectives of open government. By opening data and providing functionalities to manipulate them, transparency can be created (Hossain, Dwivedi, & Rana, 2016). Digital transparency refers to the creation of transparency by opening data and providing functionality for processing the data using all kinds of websites. Digital transparency should result in the use of open government data. However, the use lags behind (Zeleti, Ojo, & Curry, 2016), and open government efforts are often criticized for generating limited value by not opening data that can be useful for the public (Hossain et al., 2016; Janssen et al., 2012).

Open government can be achieved by opening data to the public (Luna-Reyes, Bertot, & Mellouli, 2014). Web-based portals and applications (apps) serve as an interface to the public and create active and passive transparency to an external audience (Rui Pedro Lourenço, 2016; Ricardo Matheus & Janssen, 2013). Open data portals often provide access to raw data, whereas web-based apps are developed for a specific purpose. Yet, how open government can be best implemented is underresearched (Tai, 2021).

Although digital transparency looks appealing and simple, in practice, it is more challenging to achieve (J.C. Bertot et al., 2010). Some areas, like open budgeting, are well developed (R.P. Lourenço, 2023), whereas, for other domains, the creation of transparency is more challenging. Many open government initiatives result in limited usefulness (Luna-Reyes et al., 2014). Some websites are cumbersome to use, whereas others only provide a shiny picture of what the government wants the public to see. An underlying question is an alignment between the website design and what the public wants to see (Ricardo Matheus, Janssen, et al., 2018). Máchová and Lněnička (2017) stress the need for offering all kinds of functionality on websites to create transparency.

Websites and apps should help to create digital transparency in open government (J.C. Bertot et al., 2010). There are many definitions and conceptualizations of transparency. For example, Ward (2014, p. 46) defined transparency as “being able to ‘look into’ these agencies and see how they operate”, whereas Corradini, Polini, Polzonetti, and Re (2010, p. 303) conceptualized transparency as “the ability of the administration to make citizens aware of the delivery process

and of its execution state, improving the citizens' perceived trust in this way". In general, there is an agreement that transparency in open government is the ability to gain insight into the government by the public (Ricardo Matheus & Janssen, 2015a, p. 1). Many apps provide some insight for citizens, but this might not be sufficient as citizens might still not understand what is happening. Citizens have various levels of education, experience, and digital skills (Ricardo Matheus & Janssen, 2020), resulting in different needs for digital transparency. Hence, digital transparency is about seeing what is happening in the government through digital open governmental portals and apps. These portals and apps can enable citizens to understand what is happening inside the government without the need to work in the government or to be present daily in public offices. Therefore, we take a slightly different view in this research than the contemporary public administration literature and define digital transparency as a stakeholder's ability to understand what is happening in the government using portals or apps. Our definition streses that stakeholders are diverse and might have different transparency needs. Furthermore, the definition stresses the focus on usefulness of transparency for the stakeholders. For the sake of brevity, we will use the term transparency to refer to digital transparency.

Usefulness can be enhanced by developing efficient and transparent applications by providing a pre-defined view to the public. Governments can create such applications, but they can also be developed by third parties intermediaries (Shaharudin, van Loenen, & Janssen, 2023). Apps often give some insights from a single perspective, whereas other perspectives might give different insights. For example, if only budget information is shared, then no insight is gained into resource utilization. Therefore it is often advocated that raw data should be opened using open data portals (Attard, Orlandi, Scerri, & Auer, 2015). The use of raw data might consume a lot of time and the use of all kinds of functionality for processing the data. On the other hand, providing pre-defined views can be more efficient to use but might not be useful nor provide the transparency the public is looking for. The actual level of insight needed is often difficult to determine, as which views are appropriate to create transparency is dependent on the needs and might change over time (Cahlikova & Mabillard, 2019). This paper focuses on both portals for opening raw data and applications that create data for their users. Portals provide all kinds of data, whereas the second provides one or more pre-defined views that can be used in an efficient way.

Although there is much research into open government, empirically investigating the usefulness and factors influencing usefulness remains scarce (Tai, 2021). Much of the research

is not actionable and focuses on relationships between openness, accountability, trust, and transparency (Grimmelikhuijsen, 2012b; Ohemeng & Ofosu-Adarkwa, 2014; Welch & Hinnant, 2003). In particular, the functionality needed and the efficiency are typically neglected. This paper addresses this void in the literature. This paper aims to develop a model for open government usefulness and to analyze the relationship between transparency and usefulness. Identifying these factors can help designers of transparency applications and public policy-makers to create open government applications and accompanying policies that are more useful.

This paper is structured as follows. In the following section, the background of our study is outlined, and the research hypotheses are formulated. In section 3.2.3, the research methodology is described. The results are presented in section 3.2.4 and discussed in section 3.2.5. Conclusions are drawn, and suggestions for future research directions are made in section 3.2.6.

3.2.2 Theoretical Background

In this section, we develop the hypothesis underlying our research. Bessa-Vilela, Caramelo-Gomes, and Morais (2017, p. 728) argue that digital portals and applications demand certain functionalities to be useful for a diverse group of people with different skills. The proper design requires to balance transparency and functionality (Bessa-Vilela et al. (2017, p. 734). All kinds of functionality are needed to create transparency for the public (Alexopoulos, Loukis,& Charalabidis, 2014). Portals often have comprehensive functionalities, whereas apps provide simple functionality for the public. Various functionality can help the users of portals and apps to create transparency within a certain time frame. Functionality for visualization is essential to interpret and to create transparency. Our first hypothesis takes into consideration that Functionality positively influences Perceived Transparency.

H1: Functionality positively influences the Perceived Transparency

Most people have limited time to be involved in open government, although there are exceptions. For example, NGOs might have the resources to drill into all kinds of detail and do detailed analyzes (Shaharudin et al., 2023). Jetzek, Avital, and Bjørn-Andersen (2013) emphasize users' limited time and resources and the need for efficiency as a value-creating

mechanism. Máchová and Lněnička (2017) stress the need for processing and integrating Functionality in an efficient manner to create Perceived Transparency. Ready-made apps minimize the time of the public, whereas functionality for the processing of data can accelerate efficient use. Alexopoulos, Loukis, and Charalabidis (2014, p. 67) emphasize that efficiency should be created to improve users' use of open data. This results in our second hypothesis. We hypothesize that Functionalities present in open government data portals and apps positively influence users' Perceived Efficiency in processing open data.

H2: Functionality positively influences the Perceived Efficiency

Portals and apps should help to create transparency. Creating transparency can be a cumbersome task, and functionality can help to create transparency in an efficient manner. Functionality can enable efficient use (Alexopoulos et al., 2014), for example, by providing functionality for data processing or by already providing a pre-defined view to the public. Governments can create such applications, but they can also be developed by third parties intermediaries (Shaharudin et al., 2023). Máchová and Lněnička (2017) argue that functionality is needed to integrate data in an efficient manner to transform the data into a useful format. Our third hypothesis suggests a positive relationship between Perceived Transparency and Perceived Efficiency.

H3: Perceived Transparency positively influences the Perceived Efficiency

Máchová and Lněnička (2017) stress the need for having all kinds of functionality to enable useful open government websites. Zuiderwijk, Janssen, and Parnia (2013) and Alexopoulos et al. (2014) provide a list of functionality requirements to create users' usefulness. Rui Pedro Lourenço (2015) observed that governments should decide properly about what data is released and what functionality is needed to process the data, since the nature of the data being disclosed might influence the level of perceived usefulness by citizens and external users of the opened datasets. Due to the aforementioned, our fourth hypothesis is that Functionality positively influences the Perceived Usefulness of open government data portals and apps.

H4: Functionality positively influences the Perceived Usefulness

The more transparent a website or app, the higher the usefulness for open government (Lean, Zailani, Ramayah, & Fernando, 2009). Scholl and Luna-Reyes (2011) suggested a positive relationship between transparency and usefulness. The creation of transparency results in

usefulness, as transparency helps to find the answer to the questions to reach the desired objectives of users. Data disclosure not always results in greater transparency or usefulness of these datasets opened in open data portals. More data might result in the drowning of data and less usefulness of the portal. The study of Weerakkody, Kapoor, Balta, Irani, and Dwivedi (2017) showed that open data portals might enable citizens to see the usefulness of this data by increasing transparency. Our fifth hypothesis aims to identify if Perceived Transparency influences Perceived Usefulness.

H5: Perceived Transparency influences Perceived Usefulness

Jetzek et al. (2013) emphasize users' limited time and resources of users. Too much data might also take more time to process or even result in data overload. Perceived Usefulness might depend on the Perceived Efficiency of the use of the website or App to find the right answers within a short timeframe. In other fields, this relationship is found, e.g., Dillon, McDowell, Salimian, and Conklin (1998) found that nurses would perceive higher perceived use if bedside-computer systems were efficient for them. A similar conclusion was drawn by Jeng (2005) in the research on digital library users. Efficiency is essential as open data portals often consist of many functionalities, and many activities are needed (Zuiderwijk et al., 2013). Inefficient and cumbersome activities might result in a lack of use. Hence, our sixth hypothesis takes into consideration that transparency applications bring citizens Perceived Efficiency in their work, which will increase their perception of usefulness.

H6: Perceived Efficiency is positively related to Perceived Usefulness

Máchová and Lněnička (2017) argue that all kinds of functionality for creating transparency result in usefulness. There is no empirical research discussing the relationship between functionality and usefulness mediated by Transparency, although there is literature about the relationship between functionality for transparency on the one hand and transparency and perceived usefulness on the other hand. Nilashi, Jannach, bin Ibrahim, Esfahani, and Ahmadi (2016) suggested that functionality and perceived usefulness are somehow influenced by transparency. We expect that functionality helps to create transparency, and in turn, transparency will result in higher levels of perceived usefulness. Considering the novelty of this relationship, our seventh hypothesis aims to identify whether the relationship between Functionality and Perceived Usefulness is mediated by Perceived Transparency.

H7: The relationship between functionality and Perceived Usefulness is mediated by Perceived Transparency

In a similar vein, there is no empirical research testing the relationship between functionality and perceived usefulness mediated by efficiency. Alexopoulos et al. (2014, p. 67) list of functional requirements suggests that the functionality enables to create usefulness in an efficient manner. Máchová and Lněnička (2017) and (Zuiderwijk et al., 2013) argue that all kinds of functionality for creating transparency result in usefulness which requires efficient use. However, there are articles suggesting relationships between functionality and efficiency, and usefulness and efficiency as discussed before when we posed those hypotheses. Considering the novelty of these relationships, we have our last hypothesis, the relationship between functionality and usefulness is mediated by efficiency.

H8: The relationship between Functionality and Perceived Usefulness is mediated by Perceived Efficiency

Our hypothesis result in the model presented in Figure 20. These eight hypotheses will be tested in this research by collecting data from citizens.

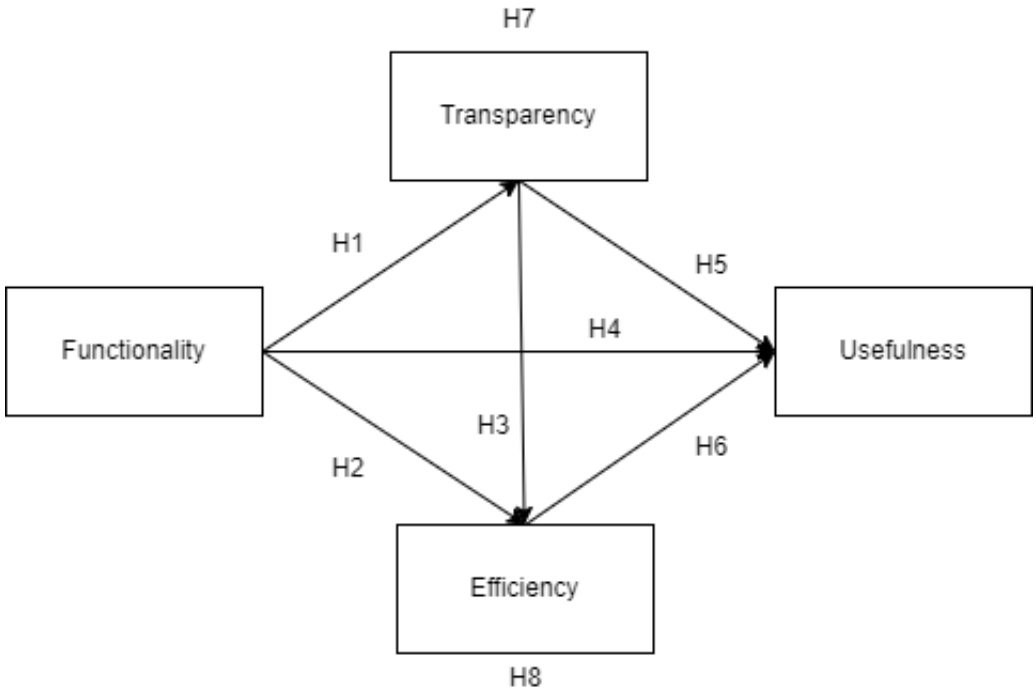


Figure 22 – Conceptual model of digital transparency and usefulness for open government

3.2.3 Research Methodology

3.2.3.1 Data collection

The data used in this study were collected through a survey distributed to users of applications to increase transparency. The questionnaire developed for the survey was pre-tested by seven colleagues experienced in surveys and quantitative research. Their remarks and suggestions were used to improve some formulations, which led to the final version of the questionnaire. We then surveyed users in the Transparency portals and apps conducted in the OpenGovIntelligence (OGI) Project (http://www.opengovintelligence.eu/). OGI was a project funded by the European Commission (EC) within the Horizon 2020 framework (H2020), which developed all kinds of applications for creating transparency. These pilots agreed to email the link to this questionnaire to their users and asked them to fill out the electronic form. In total, 187 valid responses were gathered from Belgium, Estonia, Greece, Lithuania, England, and Ireland from September to November 2019. A reminder was sent after three weeks. The responses are anonymous, and no detailed demographic data was collected. This should ensure that respondents feel comfortable responding, fetching honest answers, and complying with the privacy requirements. The questionnaire was meant to be short and to the point so that it could be distributed to end-users without gathering personally identifiable information in compliance with the ethical guidelines to avoid the collection of personal information if not necessary. Hence, there were no questions to collect demographic data.

3.2.3.2 Data processing

We used Partial Least Square (PLS) as recommended by Pavlou and Gefen (2005). More details of the PLS method is described in section 3.2.4.2. Besides PLS, we followed McDonald and Ho (2002) guidelines, selecting SEM as it is able to analyze the structural relationship between the measured variable and latent constructs. As a SEM step, absolute fit indices are tested to identify whether a priori model fits or does not. McDonald and Ho (2002) stated that "*given the complexity of structural equation modelling, it is not uncommon to find that the fit of a proposed model is poor*".

This paper used SmartPLS 3 and SPSS 24 to conduct the SEM analysis. All our variables were intended to be measured as reflective constructs using multi-item scales, meaning they are

meant to be latent (not directly observed) variables. All items were measured using a 1 to 5 Likert scale.

3.2.4 Results

3.2.4.1 Descriptive statistics

The data originates from Belgium, Estonia, Greece, Lithuania, England, and Ireland. In total 112 people were surveyed, as shown in Table 13. The ethics committee did not recommend collecting demographic data, and no hypotheses were formulated. We kept the questionnaire short and did not collect demographic information, however, the pilot participants primarily filled in the questionnaire. Hence, the questionnaire was filled out by those persons who had an interest in and experience with digital transparency. The persons in charge of the pilots indicated that there were two types of groups. One group consisted of experienced persons who were highly skilled and able to analyze raw data, whereas the other group had hardly any skills and was primarily interested in creating digital transparency. Both groups are included in the sample.

Table 13 – Descriptive key demographic variables

#	City / Country	Number people interviewed
1	Trafford England	28
2	Lithuania	22
3	Estonia	10
4	Belgium	2
5	Ireland	26
6	Greece	24
	TOTAL	**112**

3.2.4.2 Partial Least Square

Partial Lease Square (PLS) was utilized for data analysis using SmartPLS 3.0 software. PLS is appropriate for the analysis of complex models with latent variables and small sample sizes (Pavlou & Gefen, 2005). Previous IS studies successfully applied this technique, which found

that it is an effective method for data analysis (Shirish, Chandra, & Srivastava, 2021; Wamba, 2022). The measurement model was evaluated first by applying the recommended two-stage analytical procedure, then by examining the structural relationships (Hair, Black, Babin, Anderson, & Tatham, 1988).

3.2.4.3 Measurement Model

Three types of validity were tested: content, convergent, and discriminant validity. Content validity, the assessment of the chosen measures' appropriateness in capturing the full domain of constructs (Straub, Boudreau, & Gefen, 2004), was examined by checking for consistency between the measurement items and the existing literature. This was completed at the questionnaire design stage. Convergent validity, which checks for the indicators for a construct correlation with one another in comparison with the indicators of another construct (Petter, Straub, & Rai, 2007) was tested by using factor analysis (Table 14). The output depicts a strong correlation between each item and its corresponding construct, demonstrating convergent validity.

Table 14 – Cross loadings

	Eff	Func	Tr	Usefulness
Eff1	**0.920**	0.44	0.334	0.598
Eff2	**0.829**	0.222	0.336	0.412
Func 3	0.377	**0.88**	0.337	0.617
Func1	0.291	**0.766**	0.071	0.382
Func2	0.288	**0.803**	0.016	0.433
Tr1	0.306	0.238	**0.87**	0.412
Tr2r	0.355	0.128	**0.873**	0.428
Useful1	0.477	0.376	0.414	**0.762**
Useful2	0.484	0.462	0.457	**0.869**
Useful3	0.509	0.647	0.338	**0.854**

Convergent validity was also tested by examining composite reliability (CR) and average variance extracted (AVE) for the indicators (Hair et al., 1988)). As can be seen from Table 15, CR values range from 0.836 to 0.898, which is above the suggested CR threshold of 0.7 (Chin,

1998). AVE values were above the recommended threshold of 0.5 (Fornell & Larcker, 1981) and ranged between 0.688 to 0.767. Cronbach's values ranged from 0.684 to 0.772, fulfilling the threshold criteria (Morgan, Cleave-Hogg, DeSousa, & Tarshis, 2004).

Table 15 – Construct reliability and validity

	M(SD)	**Cronbach's alpha**	**Composite reliability**	**Average variance extracted (AVE)**
Eff				
Eff1	3.76(0.78)			
Eff2	3.49(0.72)	0.704	0.868	0.767
Func				
Func1	3.79(0.73)			
Func2	3.73(0.74)			
Func3	3.85(0.79)	0.760	0.856	0.669
Tr				
Tr1	3.09(0.54)			
Tr2	2.78 (0.61)	0.684	0.863	0.760
Usefulness				
Useful1	3.40(0.66)			
Useful2	3.48(0.74)			
Useful3	3.62 (0.75)	0.772	0.868	0.688

Discriminant validity was verified by checking the square root of the AVE (Fornell & Larcker, 1981). The square root values of AVE are all greater than the corresponding intern construct correlations (Table 16), demonstrating satisfactory discriminant validity.

Table 16 – Correlations and Fornell-Larcker criterion (Discriminant validity)

	Eff	Func	Tr	Usefulness
Eff	**0.876**			
Func	0.397	**0.818**		
Tr	0.379	0.21	**0.872**	
Usefulness	0.591	0.606	0.482	**0.830**

Additionally, Heterotrait-monotrait (HTMT) ratio criterion was also checked (Hair et al., 1988). The HTMT should be lesser than 0.85 to discriminate between factors. The results in Table 17 demonstrated that HTMT is less than 0.85, meeting the HTMT criterion for discriminant validity. As a result, the suggested outputs indicate a satisfactory measurement model.

Table 17 – Heterotrait-monotrait (HTMT) ratio criterion

	Eff	Func	Tr	Usefulness
Eff				
Func	0.504			
Tr	0.549	0.273		
Usefulness	0.781	0.746	0.669	

3.2.4.4 Common Method Bias (CMB)

CMB was accessed by performing the full collinearity variance inflation factors (VIFs) test (Kock, 2015). The degree of common method bias was measured with Harman's single-factor test. Harman's single-factor test was conducted by including all the items in a principal component factor analysis. Based on the analysis, the cumulative variance extracted was 28.84 percent, which is well below the 50 percent threshold (Harman, 1976; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), indicating an absence of common method bias.

The results (Table 18) show that the pf values of the full VIF for each construct are below the recommended threshold of 3.3 (Kock, 2015), suggesting that the proposed research model could be considered free of CMB.

Table 18 – Full collinearity statistics (VIF)

	VIF
Eff2	1.419
Eff3	1.419
Func 4	1.527
Func1	1.492
Func2	1.606
Tr1	1.369

Tr2r	1.369
Useful2	1.399
Useful4	1.906
Useful3	1.737

The hypotheses were tested using SEM. First, we discuss the demographic data, followed by the common method bias. To test the hypothesis, we estimated a measurement model to find whether our items were able to measure the intended concept. Finally, we present the structural model.

3.2.4.4.1 Structural model

We used Smart PLS 3 to assess the hypothesized relationships based on explanatory power (R2), for model quality. Additionally, we followed standardized root mean square residual (SRMR) for model fit (Henseler, Hubona, & Ray, 2016). The SRMR value of the research model is 0.10 which is close to the recommended threshold value of 0.10 (Henseler et al., 2016). In order to test the hypotheses' significance, the bootstrapping re-sampling methods (5000 re-samples) (Hair et al., 2011) and 95% confidence interval (Chin, 1998) was used.

The results of the structural model evaluation are presented in

Table 19. It can be concluded that H1-H6 are strongly supported. In addition to testing direct effects in the proposed research model, various mediating effects were also tested. The table shows that one indirect effect is significant, supporting hypothesis H8.

Table 19 – Hypothesis testing

Hypothesis	Path	β	P values	Results
H1	Func -> Tr	0.210	0.044	Supported
H2	Func -> Eff	0.332	0.001	Supported
H3	Tr -> Eff	0.309	0.000	Supported
H4	Func -> Usefulness	0.421	0.000	Supported
H5	Tr -> Usefulness	0.272	0.000	Supported
H6	Eff -> Usefulness	0.321	0.000	Supported
H7	Func -> Tr -> Usefulness	0.057	0.098	Not supported
H8	Func -> Eff-> Useful	0.106	0.010	Supported
	Q^2 predict		SRMR	
Efficiency	0.131		0.10	

Functionality	0.017	
Usefulness	0.347	

The model exploratory power (R-square) is 0.576.

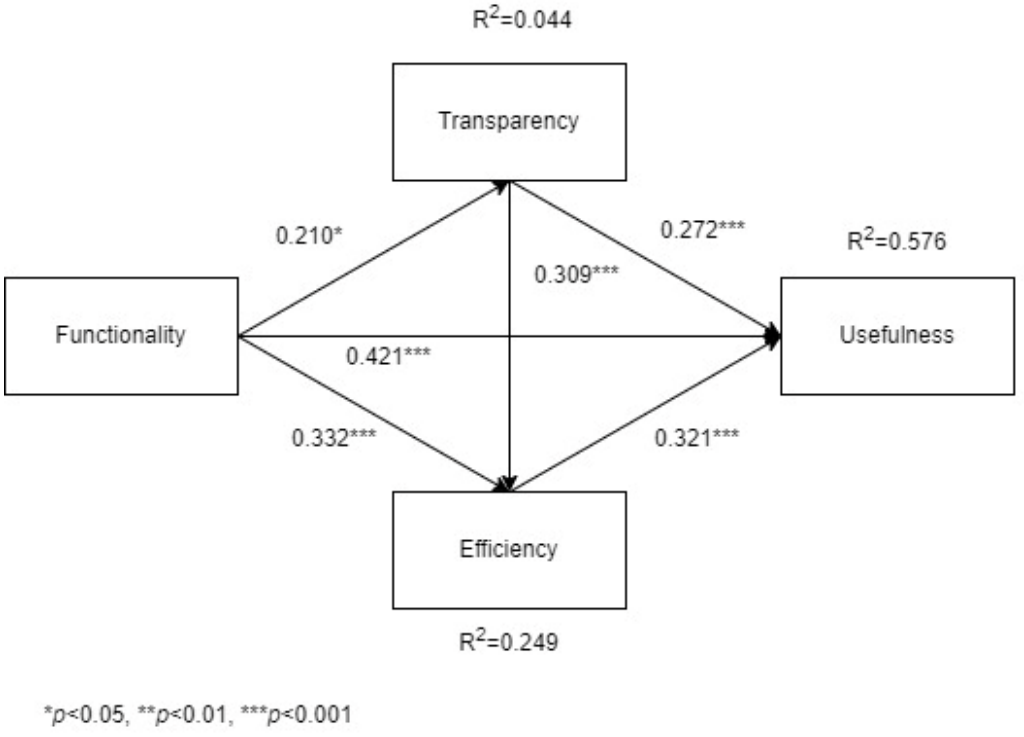


Figure 23 – Results of structural model test

Figure 23 shows the resulting final model. Functionality influences the level of Perceived Transparency (H1) positively, Functionality influences the expected Perceived Efficiency (H2) positively, and Functionality also positively influences the Perceived Usefulness (H4). The level of Transparency influences the Perceived Efficiency (H3), and the level of Perceived Transparency influences the Perceived usefulness (H5). The Perceived Efficiency influences the Perceived Usefulness (H6). Functionality positively influences the expected Perceived Efficiency and the Perceived usefulness. However, functionality does not influence any level of Transparency and expected usefulness (H7). The latter shows the complex relationship between functionality, Perceived transparency, and Perceived Usefulness. More functionality might help some stakeholder groups to increase Perceived Transparency and Usefulness, whereas others are helped by less and simpler functionality.

The survey results show that functionality, Perceived Transparency, and Perceived Efficiency are key considerations for developing useful open government applications. On the one hand, functionality can lead to customized apps having a pre-defined view, being used in an efficient manner, and providing transparency resulting in high levels of Perceived Usefulness. The Functionalities are used for customization but have the disadvantage of having a pre-defined view. On the other hand, diverse Functionalities can be used to process raw data more efficiently, resulting in higher levels of Perceived Transparency and Perceived Usefulness. This requires a wide range of functionalities.

3.2.5 Discussion

Open government initiatives are often criticized for not providing value (Janssen et al., 2012; Jetzek et al., 2013) and being useful (Luna-Reyes et al., 2014), whereas usefulness is essential for creating open government. In the ideal world, the public has the time and capabilities to make sense of open government data. In reality, they have limited time and resources Jetzek et al. (2013), and they need advanced functionality to create transparency Máchová and Lněnička (2017) and to ensure usefulness (Luna-Reyes et al., 2014). Our SEM models suggest that having the right functionalities are the basis for ensuring a useful open government.

Our model provides insight into the value creation by open government. There are two primary types to create open government data. The types differ in the type of functionalities and the way open government data is processed. The first type is to build fancy apps readily for use by the users based on an analysis of their needs. This results in providing them with a pre-defined view that they can be used in an efficient manner and provides the transparency needed to make the results useful. Apps are developed to provide often a single, or only a few views, and are user-friendly and visual. This is a useful approach for repetitive applications, but the disadvantage is that no other views can be created that might result in different or new insights. For some stakeholders, this might not result in the transparency they are looking for (Rowley, 2011). The second type is based on providing raw data using portals. Raw data needs extensive work in understanding, combining, analyzing, and visualizing data, which activities are often time-consuming (Alexopoulos et al., 2014). Functionality is needed to analyze the data in an efficient manner, and then diverse ways of transparency can be created by users to make the results useful. This results in deep insights and higher levels of Perceived Transparency, however, it usually requires much analysis and a lot of work. This might only be useful for

those with the time and ability to make sense of raw data. This is a relatively small group when looking at the whole populations, however, almost half of the population in our sample.

The public should be viewed as a collection of diverse stakeholders having their own interests (Rowley, 2011). Transparency depends on the eye of the beholder and what one is looking for. For one person, open data portals might result in transparency, whereas the same portal might result in no or limited transparency for another person. The public wants answers to their questions, but their questions might be different. This makes the development of digital transparency non-trivial. The results suggest that both apps and releasing raw data are needed to create transparency. Some of the users will prefer the use of apps, whereas others, who have more time, and the capabilities to analyze the data in-depth, will prefer to have access to the raw data. Both ways require different functionalities, which enhance the Perceived Efficiency of use, create higher levels of Perceived Transparency and result in the Perceived Usefulness of open government data.

Functionality can enable efficient use (Alexopoulos et al., 2014). Most portals contain comprehensive functionalities (Zuiderwijk, Janssen, & Davis, 2014) and are less efficient to use. Yet, Perceived Efficiency influences Perceived Usefulness. This suggests that efficiency is important considering when creating and creating portals as this influences the Perceived Usefulness.

Transparency usually requires the inclusion of different views. Higher levels of Perceived Transparency can be created by including diverse functionalities (e.g., filters, maps, graphs, tables) and making raw data available. Then, letting the public analyze the data from their desired perspectives is possible. Although this is more time-consuming and requires understanding how the data is collected, it creates higher levels of transparency. Yet, usage consumes time, and there is a need to focus on efficiency to be able to advance the understanding within a short time frame. The results of complex analysis of open government data can be useful, but the outcomes might not be. The risk is that much time is spent on analysis that might not result in useful analysis. Hence, higher levels of transparency come at a price.

Citizens might perceive transparency differently. The literature has various transparency definitions (Bannister & Connolly, 2011a; J.C. Bertot et al., 2010; Helbig et al., 2010; Luna-Reyes et al., 2014; Ward, 2014). Our findings suggest that transparency is highly contextual and that the portals and apps might have different ways of creating transparency. Apps are

efficient and can show results at a glance having high usefulness. As the context changes, also the influence of the factors changes. Portals should ensure efficient use to be useful. In other words, digital transparency needs to be created in an efficient way. Transparency should always take into account the stakeholder group for whom transparency is created.

3.2.5.1 Research contributions

There is much discussion about the value of open government (Hossain et al., 2016; Jetzek et al., 2013; Luna-Reyes et al., 2014). In contrast to other research, we looked at the usefulness of open government apps and websites, as, in the end, transparency can only be created when websites and apps are used. Usefulness is hardly considered in existing models (Grimmelikhuijsen, 2012b). The scant attention given to perceived usefulness is surprising, as only use can lead to an open government.

Open data portals have many functionalities and activities that are needed to create digital transparency (Zuiderwijk et al., 2013). Perceived efficiency is also given limited attention in research (Kassen, 2013; Zuiderwijk et al., 2013), whereas our research shows that efficiency is important. Users are diverse and have limited time and use open government data in different ways. As such, the public should not be considered a homogenous group in further research. Different groups have different needs, and further research should focus on how to create some level of transparency for different user groups.

Transparency is a complex construct. Functionalities are needed to create transparency. Yet full transparency is often not needed nor required. Apps can be very useful and provide the necessary insight; without needing that the government becomes fully transparent. Raw data can provide more insights but might be less efficient, and the question is if the insights are useful. Some might be, whereas others might not. There is a need for further theorizing to understand better how digital transparency for citizens can be created in different contexts.

3.2.5.2 Practical implications

Open government efforts have often not realized their potential and resulted in disappointing results (Zeleti et al., 2016). Much of the current research has focused on value-creation mechanisms which are hard to bring into practice. Instead, our research shows that the key to open government is creating websites and apps that are useful for the public. Digital Transparency should be viewed from diverse stakeholder's point of view. Hence, Digital

Transparency should be and views as the ability of a stakeholder to understand what is happening in the government using portals or apps. By taking this view, the practical contribution of a portal for the users is stressed.

Digital transparency is an ambiguous concept that is hard to define. By emphasizing the usefulness for stakeholders, the focus becomes more clear. The point of view of stakeholders should bet taken to bring digital transparency into practice.

Our research provides fresh insights and shows that functionalities and efficiency are key for contributing to perceived usefulness. All too often, the open data portals are complex and they cannot be used in an efficient manner (Kassen, 2013). Open government initiatives should focus either on efficiency by showing a single view or a limited number of views or on creating higher levels of transparency by releasing the raw data (Ricardo Matheus, 2017). Apps can be efficient and provide quick insight but have pre-defined views determined by their developers, limiting transparency to these pre-defined views, whereas opening raw data can provide higher levels of transparency by enabling the creation of additional views by the public but need more functionalities and have a longer time to use. A trade-off between these aspects is required, and for open government policy-makers and designers, this implies that both easy-to-use apps and comprehensive portals are needed. Both serve different purposes and create different types of transparency. Apps are often focused on the general public having limited knowledge about statistics and manipulating data (Janssen, Matheus, et al., 2017). Data is put in context to make it easy to understand and manipulate within a limited time. This approach's disadvantages are that a pre-defined view is given and that not all manipulations are possible. This does not result in complete transparency (Cukierman, 2009; Fung et al., 2007), however, it helps to create some level of transparency and can be useful. In contrast, data portals can be used by citizens having the expertise and the time to analyze data (Ricardo Matheus, Janssen, et al., 2018). This can provide greater insight resulting in higher levels of transparency, however, this is only feasible for experts and the usefulness of the results can vary. Hence our findings stress the need to consider the diversity of stakeholders. Furthermore, the findings highlight the need to develop different websites and apps for different stakeholders.

3.2.5.1 Limitations and Future work

Although the current study provides some useful analysis of factors influencing digital transparency from the citizen's perspective, there are some limitations. The sample consists of

citizens who are often familiar with open data, or at least that open data is available for their needs. In retrospect, we found that persons who were not familiar with open data were hardly included in the sample. As such, our sample is only representative for experienced users. This might not be surprising, as only those who are familiar with the use of open data for transparency are probably interested in filling in the survey.

We followed Grimmelikhuijsen (2012b) by developing a realistic view of transparency. We refrained from including typical public administration constructs like trust and openness. For further research, we recommend developing models that integrate constructs like trust, accountability and openness, which are often used in public administration research.

Digital transparency is not easy to realize. There is a need for further theorizing to understand better how datasets and functionality can be used to create digital transparency for citizens. Governments should not just assume that apps or portals achieve transparency. We recommend developing design methods for supporting governments to create digital transparency. Citizens can be involved in the design process of creating transparency to understand their needs better. Governments should explore a variety of means to create transparency and not focus on a single way. Different strategies need to be researched to fulfill the need of citizens.

3.2.6 Conclusions

Transparency will only be created if open data is useful for the public. This paper is one of the first papers investigating the usefulness of open government initiatives from a user perspective. The SEM model shows that having the right functionalities for the apps and websites is the basis for increasing the perceived usefulness. Functionalities can increase transparency, resulting in higher Perceived Efficiency and Perceived Usefulness. Dedicated attention should be paid to the citizens' different needs and ensure efficiency as the time of citizens and other users is limited. We recommend classifying different types of users in further research and testing models which can use constructs like trust, accountability, and openness.

Transparency is a complex and ambiguous construct. Transparency can increase the credibility of open government, but usefulness might not always need transparency. Apps result in higher efficiency for the users, but only provide insight from one or a few pre-defined views. The level of transparency is limited, as other views are not covered. Opening raw data requires many functionalities that are less efficient than pre-defined apps. However, these enable users

to create their own views and find insights that are not pre-defined, resulting in higher levels of Perceived Transparency, but this might not result in higher Perceived Usefulness. Whereas open government apps do not create complete transparency, they can be used by a broad public and their Perceived Usefulness can be higher. Raw data is perceived as efficient to use but can enable higher Perceived Transparency levels, which is only feasible for a limited number of persons. Hence, higher levels of transparency come at a price. In further research, we recommend creating a classification for transparency initiatives considering the context variations. Some types of initiates are likely to be affected by other factors. For example, the functionality of open data portals for raw data will likely differ from ready-for-use apps. These influence the Perceived Efficiency, level of Perceived Transparency, and, ultimately, the Perceived Usefulness.

Acknowledgment

Part of this work is funded by the European Commission within the H2020 Program in the context of the project OpenGovIntelligence www.opengovintelligence.eu under grant agreement [No. 693849].

Part of this work was funded by the European Commission within the ERASMUS+ Programme in the context of the project CAP4CITY www.cap4city.eu under grant agreement [No. 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP]. The authors would like to cordially thank all the experts who participated in the study.

3.2.7 Appendices

3.2.7.1 Appendix A – Overview of Latent Constructs, Item Questions and Sources

Table 20 – Overview of latent constructs, item questions and sources

Latent Constructs	Measurements (Questions)	Source
Usefulness	The visualisations provided by the Apps makes better interpretation of data	Romi (2013)
	The Apps help me to make better decisions	Romi (2013)
	The Apps are useful to me	Romi (2013) Delone and McLean (2003)
	The Apps helps me to achieve my goals	Romi (2013) Delone and McLean (2003)
Ease of Use	I think the Apps are easy to use	Delone and McLean (2003)
	I have sufficient skills to use the Apps	Delone and McLean (2003)
	The Apps do not require high-level technical knowledge	Ricardo Matheus and Janssen (2013)
	I found the Apps are difficult to use	Ricardo Matheus and Janssen (2013)
Efficiency	The Apps will increase the efficiency of my work	Delone and McLean (2003)
	The Apps reduce time spent looking for information	Delone and McLean (2003)
	The Apps reduce the costs to find information	Delone and McLean (2003)
Data Quality	The data in the App are accurate	Romi (2013) Lin (2007)
	I am satisfied with the quality of the datasets provided by the App	Delone and McLean (2003)

	The data in the App are accessible	Acton, Halonen, Conboy, and Golden (2009)
	The data in the App are incomplete	Delone and McLean (2003)
System Functionality	All functions in the App works properly	Ojo (2017)
	I found the various functions in the apps are well integrated	Ojo (2017)
	The app provides all the functions I am interested in	Ojo (2017)
Transparency	The Apps result in an increase of transparency	Ricardo Matheus and Janssen (2013)
	More functions in the Apps are needed to create transparency to support decision making	Ricardo Matheus and Janssen (2013)

3.3 Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities

This chapter is based on: Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly, 37(3), 101284.

The first author conducted the literature review, designed and conducted the data collection and analysis process, and wrote the manuscript. The co-authors provided feedback on the data collection and analysis process and earlier versions of the manuscript. We would like to thank the anonymous reviewers for their reviews.

The paper can be accessed and freely downloaded at: https://www.sciencedirect.com/science/article/pii/S0740624X18300303

Abstract

Dashboards visualize a consolidated set data for a certain purpose, enabling users to see what is happening and initiate actions. Governments can use dashboards to support their decision-making and policy processes or communicate and interact with the public. This paper aims to understand and support the design of dashboards for creating transparency and accountability. Two smart city cases are investigated, showing that dashboards can improve transparency and accountability; however, realizing these benefits was cumbersome, and various risks and challenges were encountered. Challenges include insufficient data quality, lack of understanding of data, poor analysis, wrong interpretation, confusion about the outcomes, and imposing a pre-defined view. These challenges can easily result in misconceptions wrong decision-making, creating a blurred picture, resulting in less transparency and accountability and even less trust in the government. Principles guiding the design of dashboards are presented. Dashboards need to be complemented by mechanisms supporting citizens' engagement, data interpretation, governance, and institutional arrangements.

3.3.1 Introduction

Data science in government deals with the extraction and interpretation of insights from unstructured and structured data that can be either closed or opened. Cleveland (2001) argues that data science consists of multidisciplinary investigations, models and methods for data, computing with data, pedagogy, tool evaluation, and theory. Government data scientists need in-depth knowledge of statistics and data analytics for analyzing data, as well as knowledge on the use of techniques and instruments for predictive purposes. By combining disciplines, new insights and applications can be created.

Data science is an essential area for governments, as they have a lot of internal information in various areas (geographical, traffic, social security, energy, etc.) that can be combined or enriched with data from smart devices and other sources such as discussion forums, social media, and private sector data (Janssen, Matheus, & Zuiderwijk, 2015). Making sound decisions depends on using high-quality data (Chengalur-Smith, Ballou, & Pazer, 1999). Data might be an enabler for creating new innovative applications (Marsh, Pane, & Hamilton, 2006), which should result in the creation of public values like security, safety and transparency.

In data science, the sharing, use and interpretation of data are key aspects in bridging the gap between the government and the public. The use of data and the accompanying instruments will likely influence government decision-making, resulting in new applications, but can also impact its interaction with the public (Ganapati, 2011a). Dashboards can be used to release information for decision-makers (Maheshwari & Janssen, 2014) and the public. Dashboards should help to facilitate transparency, governance, trustworthiness and enable citizens to participate in decision-making in smart cities (Allio, 2012).

Few (2006, p-34) defines dashboards as “a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance.” Recently, dashboards have gained more and more attention in the public sector. For example, in 2009, the US federal government developed dashboards with federal stimulus funding (www.recovery.gov) aiming for transparency and accountability of national economic recovery policy (Ganapati, 2011b). In particular, with open government, dashboards are an important means of communicating with the public.

Dashboards are often part of public organizations' 'open government' efforts to create transparency and stimulate engagement with citizens and businesses. Open government efforts aim for a more democratic, efficient, and effective government and result in accountability and trust in the government. Employing dashboards is often a difficult endeavor. Data is often context-specific, and interpretation will likely be wrong without in-depth knowledge of the context in which the data is collected (Matheus & Janssen, 2013). As such, data science in government requires in-depth skills and knowledge about the inner workings of the government and its environment (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012).

This paper aims to identify the benefits, risks, and principles of designing dashboards. This is done by reviewing the literature about dashboards and complementing the literature by investigating two case studies in detail. This paper is structured as follows. We investigated this question by reviewing the literature about dashboards and demonstrating the value of dashboards in a case study. Section 3.3.2 presents the case study descriptions. Literature is used to identify principles for designing dashboards, presented in the subsequent section 3.3.3 with the benefits, risks, and challenges of dashboards and the data science cycle for dashboards. Finally, conclusions are drawn in section 3.3.6.

3.3.2 Dashboards in Practice

The Smart City of Rio de Janeiro, Brazil, in South America, has developed an infrastructure, a dashboard, and a data portal with more than three thousand datasets and seven APIs for real-time data use (www.data.rio). The Smart City dashboards in Rio de Janeiro were created to solve problems related to public transportation and traffic. Due to the economic boom and access to credit in Brazil, there has been an increase in the number of cars (3 million cars for 6.5 million inhabitants). This has resulted in huge traffic jams, even outside of rush hours.

The IBM Center of Operations Rio (COR) is a four-floor building that reunites almost 30 secretariats and public and private enterprises to identify and solve in real-time issues on the city. COR collects around 4 Gigabytes (GB) of data in transit every day. This includes data about bus stops, car accidents, construction works, and accidents like a tree falling down on the streets. This excludes real-time GPS (geographical location collected using the Global Positioning System – GPS) data coming from buses driving their routes. Every 2 minutes, data is collected from the position of over 8000 buses. This results in another dataset that amounts to 12 GB per day. With data from Center Operations and the work of PENSA – Ideas Room,

dashboards are created with the objective of putting key organizational elements into a consolidated format using several visualization tools, gauges, graphs, charts, and pictograms. The PENSA, a group of data analysts at Rio City Hall, enabled the Center of Operations to visualize the data in a structured, integrated, and organized manner at a glance. The process of data analysis used by PENSA in Rio de Janeiro is based on questions made by the political decision-makers (mayor and secretariats) to create the Public Sector Dashboards.

3.3.2.1 Case 1 – Traffic Jam Dashboard

The first case represents a partnership between the Smart City of Rio de Janeiro and the Social GPS Smartphone application, Waze. This app allows citizens to send the city real-time information about traffic conditions and accidents. The Center of Operations Rio uses the Waze application to send its citizens real-time information about route changes, flood routes, traffic jams, and car accidents. The result is a combination of open data from the Rio de Janeiro City Hall, and the Big Data collected through Waze about Rio de Janeiro's 7 million inhabitants, presented in real-time using electronic panels positioned all over the city. The dashboard enables citizens to select the best route between downtown and peripheral regions to avoid delays due to traffic jams. The dashboards were planned according to the Big Data analysis from PENSA, considering data from several internal databases (GPS data from buses, traffic jams, speed of traffic, car accidents, etc.) and the Waze application.

3.3.2.2 Case 2: Buses Dashboard

The second case studied was a partnership with the MOOVIT social application for buses. This partnership aimed to improve the quality of public transportation and transparency by showing in real time what was happening with the city's bus system. MOOVIT empowers citizens to choose the best route (faster, closer, etc.) in accordance with real-time information on events (traffic jams, accidents, broken-down buses, etc.) presented via the duration and distance to get from their starting to their finishing location. This is normally between peripheral regions and downtown and vice versa for commuter traffic. This application used the API to access the GPS data from buses and based on this data calculates the estimated time of arrival (ETA) of buses at a certain bus stop and the distance between the starting and finishing locations.

3.3.3 Principles for Designing Dashboards

The cases show that the design of a dashboard is dependent on many factors, including the information available and its purpose. One of the main purposes of opening up government data is to create transparency (Bertot, Jaeger, & Grimes, 2010; Dawes & Helbig, 2010). Although there are other purposes, dashboards are often aimed at empowering the public by creating transparency and accountability. Dashboards are an instrument for reducing information asymmetry (Bugaric, 2004). Information asymmetry is when one party has more information than another (Michael C. Jensen & Meckling, 1976). Reducing the information asymmetry from the government to the public can improve trustworthiness (Abelson, Gauvin, MacKinnon, & Watling, 2004) and governance (B. G. Peters & Pierre, 1998).

The design of dashboards should accomplish the goal of creating transparency and accountability. Our literature review of dashboards shows that there are only a few papers about dashboards in the public sector, whereas much more information is available about private sector dashboards. There is no recipe for creating dashboards because the creation of a dashboard is a complex problem. Therefore, we opted to identify principles as these are useful for guiding solving ill-structured or 'complex' problems (Simon, 1996). Gibb (1997) defines principles as "rules of thumb that guide the choices and actions of engineers." We used both public and private sector literature to derive the design principles presented in Table 21. The cases informed us about new principles concerning the use of real-time data. The principles can guide the designing of more effective public sector dashboards.

Table 21 – Principles and literature background

Principle	***Description***	***Literature Background***
1. Present correct and precise data	Governments must give the most correct and precise information, to prevent users from being unable to understand the data and being misled. Incorrect information in the dashboard can result in bad decisions.	(Abelson et al., 2004; Obama, 2009), case studies
2. Customize views	Dashboards should not be merely simple or generic visualizations; dashboards should contain customized views to show the problem at hand. In this way, decision makers and users can gain insight. Customized views can help them	(Eckerson, 2010; Johnston & Pongatichat, 2008; Kaplan & Norton, 2001; Little, 2004), case studies

	understand the situation. The design requires an understanding of organizational strategies, viewpoints, business processes, indirect effects, decision support systems, and priorities. In the cases found in the literature review, separate apps are developed for each purpose to enable a clear view of the problem at hand.	
3. Clear presentation	Dashboards enable the use of charts, graphs, pictograms, bars, and numbers, etc., to visualize information for monitoring and analyzing performance. Dashboards should visualize appropriate, relevant, and precise information in simple ways. In our cases, the simplicity of the dashboards enabled their use by a broad public.	(Baskett, LeRouge, & Tremblay, 2008; Few, 2006; Velcu-Laitinen & Yigitbasioglu, 2012), case studies
4. Offer decision-making support	Relationships between performance metrics and organizational desires must be transparent and clear. Dashboards can be used to evaluate 'what if' scenarios to offer decision support. This can help provide more insight into the situation and help decision-makers. By providing insight into possible alternatives, the effect of choosing an alternative can be predicted in our cases.	(Ganapati, 2011b; Velcu-Laitinen & Yigitbasioglu, 2012), case studies
5. Interaction support	Static dashboards often provide limited insight. More insight can be gained by interacting with dashboards. Real-time information was key for supporting the decisions in the cases studied.	(Chen, Chiang, & Storey, 2012; Mayer-Schönberger & Cukier, 2013; McAfee et al., 2012; R. M. Peters, Janssen, & Engers, 2004; Svensson, Saeverhagen, & Bouillouta, 2015), case studies
6. Provide overview and details	Dashboards should be able to deal with an enormous volume of big and open data. The large volume of analyzed data was presented in a simple visual display that provided an overview and the opportunity to zoom in on details.	(Lohr, 2012; Marz & Warren, 2015), case studies
7. Focus on creating added value	Merely visualizing data has limited use if this is not suitable for creating added value. Dashboards are difficult to develop, especially for big and open data. Often, a business case is required to determine the added value. Dashboards should be designed to create public value.	(Chen et al., 2012; Dietrich, Plachy, & Norton, 2014; Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012)

8. Ensure real-time updates of data	The majority of dashboards are not based on real-time data. Governments present dashboards of what has happened in the past and can use this to predict future events.	Case studies
9. Ensure institutions for supporting accountability	The creation of transparency results in the detection of incorrect data or behavior. For citizens it is hard to take any action if there is a suspicion of fraud or corruption. It should be possible for citizens to report to an independent and trusted agency. There should be institutions to deal with such matters further.	Case studies

These principles were not always followed in the two case studies. This resulted in a number of risks and challenges (section 3.3.3.2). Although these principles do not necessarily need to be followed, they can help guide the design of a dashboard so that its potential and functionalities are maximized. For example, the lack of updated, timely, and precise data contributed to misunderstanding and misinformation among citizens and decision-making managers (Ballou, Heitger, & Donnell, 2010).

3.3.3.1 Benefits of Public Sector Dashboards

Public organizations use dashboards for a variety of purposes, such as transparency, performance monitoring, reporting, planning, and quick decision-making. They can be designed for use by governments (internal) or the public (external). Internal objectives are related to monitoring and analysis for faster and more accurate decision-making, resulting in increased efficiency and effectiveness of operations. External objectives are related to creating transparency for the public, mobilizing external capacity to gain feedback, and facilitating participation by society.

The major benefits of dashboards can be summarized as fast and cheap solutions for people and government, enabling some degree of transparency. People need this information to evaluate, provide feedback, and suggest improvements. Similarly, governments can benefit by understanding people's daily decision-making (planning) and influencing their behavior by displaying information on strategically located electronic panels spread over the city. The use of these panels resulted in more transparency by reducing the information asymmetry between

citizens and public agencies. The transit time of citizens was improved as a result of creating transparency. The list and description of public dashboard benefits are presented in Table 22.

Table 22 – Overview of the main benefits

Benefits	***Description***
Strategic and political benefits	
1. Overview at a glance	Dashboards permit flexibility in creating various types of overviews at a glance. This enables different stakeholders to gain insight easily.
2. Drill into detail	The same flexibility that allows a broad overview also can be used to drill into detail for any event. During traffic monitoring, it is possible to zoom in on accidents.
3. Transparency	Creating a view at a glance can provide transparency about a situation, combined with the ability to drill down to details to really understand the data and conclusions that might be inferred from it.
4. Customer / user-orientated presentation / visualization oriented	Dashboards enable the creation of customized views. In this way, dashboards provide information to people when and where they really need it.
5. Identification of fraud and corruption	Anomalies and patterns in data can be used to detect corruption by showing them in dashboards combined with data analytics.
6. Showing trends	The information presented in dashboards concerns the trends or issues in the city. It is not necessary to know what is still good but rather what is needed to avoid the worst-case scenario.
7. Accountability	The insights can be used to help hold organizations or people accountable for their actions (or inactions). This requires the possibility of following up the results from the data analysis at the institutional level.
Operational benefits	
8. Better and faster decisions	Real-time data can enable operational strategies based on specific situations and immediate decisions.
9. Mobilize external knowledge	External partners such as smartphone applications and media (radio, TV) can use and spread the same information used on dashboards, collaborating for better effectiveness and efficiency (for instance, operational traffic conduction).
10. Improve effectiveness and efficiency of operations	Transparency of relevant information on dashboards in real-time can improve the efficiency of operations with better conduction of many situations, such as reduction of traffic congestion. The efficiency makes people trust the dashboard, improving the effectiveness of operations in the short and long term.

11. Disclose relevant information to people	Dashboards give people relevant information for real-time decision-making. If not used, they can be just an expensive data storage system.
12. Enable participation	Relevant information on dashboards allows people to participate in public decision-making and help improve the wellbeing of others.
13. Public participation in service improvement	Open data and accountable systems allow people and enterprises to create new ideas on data usage. Co-creation allows for the combination of data to develop new applications.

3.3.3.2 Risks and Challenges

Dashboards can provide both an overview and detail at the same time, which is essential for creating transparency. The benefits can only be gained if dashboards are properly designed. Furthermore, it is not only about the technology, but also about how the results created by data scientists will be used. The detection of fraud or identification of strange patterns is useless if there are no means to investigate the use of these by legitimate agencies further. The use of data outcomes requires institutional changes, also to ensure accountability. Only formal authorities, such as ombudsmen, ministries, politicians, and judges, can hold organizations accountable. Romzek and Dubnick (1987) categorize four types of accountability, based on two dimensions: whether the accountable party is internal or external to the organization, and whether the degree of control over agency actions is low or high. Combining the dimensions result in the following four categories:

(1) bureaucratic accountability: high degree of control, accountable party internal to agency;

(2) legal accountability: high control, external;

(3) professional accountability: low control, internal;

(4) political accountability: low control, external.

All of these forms might need to be supported depending on the objectives.

There are many pitfalls that can prevent the achievement of the intended benefits. If data are not properly cleansed in advance, their usage can lead to inappropriate analysis. Higher data quality comes at a price and might not always be necessary. Using statistics analysis can solve

part of the problem of low data quality. Proper vocabularies for metadata (ontology, semantic web, etc.) can solve another issue related to bad data and analytics. The datasets must also have proper timeliness and granularity or options to be accessed based on different formats. The possibility of drilling into the details of time, date, place, and description is as important as having high-level statistics of, for example, car accidents. Both data can be used for the same issue but with different objectives. If high-level statistics can show us the trends of a city traffic jam, the detailed data can give us tips for how to overcome the exact places and times of traffic congestion. They can be visualized in management reports in the form of tables and graphs or, at a glance, in heat maps on the dashboards.

Dashboards can help speed up decision-making for civil servants and the public. In this way dashboards can improve the effectiveness and efficiency of public policy operations to reduce traffic jams and car accidents, for example. This can mobilize external knowledge to use the open government data and combine other datasets to create new dashboards and applications. The external dashboards and applications can give more relevance to data and probably efficiency and effectiveness of public policies. The usage of data on dashboards by the public does not result in transparency and accountability per se. The dashboards also enable public participation in decision-making and the improvement of services. Using electronic panels to show the average time need to get from one place to another on two different routes results in better decision-making for citizens. By creating transparency citizens can choose the best route that is in their own interest of saving time, and is better than government enforcing an option without discussion or transparency.

The development and use of dashboards may involve many risks and challenges as presented in Table 23. One of the main risks is the misunderstanding of information, which could lead to incorrect conclusions about the data. This also raises the question of whether dashboards can result in transparency and accountability. Dashboards might help to improve the level of transparency, but much is dependent on the proper design. An information overload or showing the incorrect or incomplete information might result in less transparency (Matheus & Janssen, 2013). As long as there is information asymmetry, there will be no complete transparency.

Table 23 – Description of the main risks and challenges

Risks and Challenges	***Description***
1. Scarce usage	Scarce usage of dashboards by citizens may not bring the efficiency and effectiveness expected by government and the public. Traffic congestion remained the same and solutions were not working properly in our case study.
2. Fragmentation of responsibility	Many owners can rapidly decrease the quality of data and the information delivery on the dashboards.
3. Limited readiness to adapt to new changes	Politicians and public managers can boycott the new system or new technologies for a variety of reasons, such as the elderly having difficulty using the technology or civil servants avoiding loss of their political power for decision-making on public policies.
4. Limited knowledge	There are not many people with the capabilities needed to design and operate dashboards. As the number of qualified people is low, they are expensive and in high demand by the job market. Hardly any are attracted to the public sector.
5. Interpreting information	Incorrect or inaccurate data and information visualized using dashboards can reduce citizens' trust in government.
6. Wrong decision-making	If people respond or use the dashboard as predicted, the public organization must have the resources to conduct the strategy taken by people or predict plans of action to reduce issues.
7. Consuming a lot of resources	The development of a good dashboard can be expensive as numerous factors must be taken into account during development. This might require many interpretations and continuous updating after completion.
8. No maintenance	Data in a dashboard is not updated for the dashboard interface is no longer current. If dashboards do not perform well, citizens will be less likely to trust them.
9. Pre-defined view	Participatory processes are longer, more difficult and expensive. Showing to people a predefined way should be avoided to avoid bias and not viewing the complete picture. Offering alternatives to customize visualization and showing alternative views can increase insights can improve governance and trustworthiness, avoids bias from governmental perspective.
10. Not being able to adapt to new developments	The use of dashboards can result in new detailed questions, which might require additional data collection, or new plans and strategies to reduce the impact of public decision-making.
11. Data privacy	If data is not properly anonymized, private data can be displayed erroneously.
12. Bad data leads to bad results	If data is not properly cleaned via statistics processes, for example, bad data can lead to bad analytics and give people the wrong message and information. The only thing worse than no information is wrong information!

Dashboards can be expensive and their development and operation might consume a lot of resources. This becomes even worse when considering that specialized human resources are expense and rare. This can lead to a lack of new development and a standardized predefined view in the long term. Having many owners can increase the likelihood of a decrease in quality. There is also a risk of politicians and civil servants boycotting or not using the dashboard due to the difficulty of accepting new technologies, or due to the possible loss of political power in decision-making processes.

Data treatment is an essential part of dashboard development. Implementing the legal requirements to anonymize public data in order to protect people's privacy is both a risk and a challenge (Janssen & van den Hoven, 2015). It is risk because there is possibility to identify people, even when using anonymized data sets. It is challenge as anonymizing data sets requires specific knowledge and skills, and comes at a price. Statistical analysis also helps to reduce outliers and 'bad data', which leads to bad results. The use of low quality data can lead to incorrect or inaccurate decision-making (Lazer, Kennedy, King, & Vespignani, 2014). Even correct data can be misinterpreted by people.

3.3.4 Designing Dashboards

The creation of value, overview of benefits, risks and challenges helps to understand the concept of dashboards. In this section support for the design of dashboards will be derived. First, design principles will be presented, followed by the data cycle.

3.3.4.1 Design Principles

The design of dashboards should accomplish the goal of creating transparency and accountability, but is encountering many risks challenges. Our literature review of dashboards shows that there are only a few papers about dashboards in the public sector, whereas much more information is available about private sector dashboards. However, there exist no support for creating dashboards in the literature. Therefore, we opted to identify design principles as these are useful for guiding solving ill-structured or 'complex' problems (Gilb, 1997; Simon, 1996) views principles as rules of thumb that can be used to guide the designers. Informed by our 2 cases and using both public and private sector literature the design principles presented

in Table 24 were derived. The principles can guide the designing of more effective public sector dashboards.

Table 24 – Overview of design principles for dashboards.

Principles	***Description***	***Source***
1. Collect accurate and precise data	Governments must give the most correct and precise information, to prevent users from being unable to understand the data and being misled. Incorrect information in the dashboard can result in bad decisions.	(Abelson et al., 2004; Obama, 2009), case studies
2. Customize views	Dashboards should not be merely simple or generic visualizations; dashboards should contain customized views for showing the problem at hand. In this way decision makers and users can gain insight. Customized views can help them understand the situation. The design requires understanding of organizational strategies, viewpoints, business processes, indirect effects, decision support systems, and priorities. In the cases found in the literature review, separate apps are developed for each purpose to enable a clear view of the problem at hand.	(Eckerson, 2010; Johnston & Pongatichat, 2008; Kaplan & Norton, 2001; Little, 2004), case studies
3. Support different view	A single view might result in a limited picture on the situation. Different views can avoid bias and improve the understanding. By providing raw data, others can create new views which can result in updating the dashboard and improving usage.	(Eckerson, 2010; Johnston & Pongatichat, 2008; Kaplan & Norton, 2001; Little, 2004), case studies
4. Clear presentation	Dashboards enable the use of charts, graphs, pictograms, bars, and numbers, etc., to visualize information for monitoring and analyzing performance. Dashboards should visualize data in and easy-to-understand manner. In our cases, the simplicity of the dashboards enabled their use by a broad public.	(Baskett, LeRouge, & Tremblay, 2008; Few, 2006; Velcu-Laitinen & Yigitbasioglu, 2012), case studies
5. Offer decision-making support	Relationships between performance metrics and organizational desires must be clear. Dashboards can provide decision-support to evaluate 'what if' scenarios and to use predictive analytics. This can help provide more insight into the situation and help decision makers. By providing insight into possible alternatives, the effect of choosing an alternative can be predicted in our cases.	(Ganapati, 2011b; Velcu-Laitinen & Yigitbasioglu, 2012), case studies

6. Interaction support	Static dashboards often provide limited insight. More insight can be gained by providing interaction features, which enables users to view the data from various perspective, to suggest recommendation based on the data but also to provide feedback to improve the use. Real-time information was a key element for supporting the decisions in the cases.	(Chen, Chiang, & Storey, 2012; Mayer-Schönberger & Cukier, 2013; McAfee et al., 2012; R. M. Peters, Janssen, & Engers, 2004; Svensson, Saeverhagen, & Bouillouta, 2015), case studies
7. Provide overview and details	Dashboards should be able to deal with an enormous volume of big and open data. By providing an overview and the opportunity to zoom in on details, the high-volume of big data analyzed was presented in a simple visual display.	(Lohr, 2012; Marz & Warren, 2015), case studies
8. Focus on creating public values	Merely visualizing data has limited use if this is not suitable for creating added value. Dashboards are difficult to develop, especially for big and open data. Often a business case is required to determine the added value. Dashboards should be designed to create public values like engagement, transparency and accountability and adhere to public values like privacy.	(Chen et al., 2012; Dietrich, Plachy, & Norton, 2014; Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012)
9. Ensure real-time updates of data	The majority of dashboards are not based on real-time data. Governments present dashboards of what has happened in the past and can use this to predict future events.	Case studies
10. Ensure institutional support	The creation of transparency results in the detection of incorrect data or behavior. For citizens, it is hard to take any action if there is a suspicion of fraud or corruption. It should be possible for citizens to report to an independent and trusted agency. There should be institutions to deal with such matters further.	Case studies

These principles presented in table 3were not always followed in the two case studies. This resulted in a number of risks and challenges as reported in the previous section. Although these principles do not necessarily need to be followed, they can help guide the design of a dashboard to improve the creation of transparency and accountability. For example, the lack of updated, timely, and precise data contributed to misunderstanding and misinformation among citizens and decision-making managers (Ballou, Heitger, & Donnell, 2010).

3.3.5 Data Cycle for Dashboards

We abstracted the information flow and information processes from the cases in order to create internal and external dashboards. The information flow and information processes take into consideration the literature review and the framework on principles to evaluate dashboards, and two case studies evaluated according to the framework of principles. The main phases followed by data scientists are data collection, storage, analysis, and visualization. Figure 1 shows the steps and the two main information flows in the cycle. We introduce the new step of 'data usage' to show that data use is important. The main reason for this is because value can only be created from data when it is actually used (Janssen, Estevez, & Janowski, 2014). Data can be used by citizens, governments, or other stakeholder groups. Usage might be stimulated to show the relevance and benefits of the data for the intended user groups. Finally, usage might cover institutional measures and changes to ensure that actions can be taken based on the data. This is particularly relevant in holding organizations accountable.

The first flow is labelled as F1 and is black (line dashed in points) because it is a common flow for the external flow of data for people and the internal flow for the government. F1 has data from sensors, forms, etc., going directly from Stage A (Data Collection) to Stage B (Data Storage) without any treatment, represented by boxes in blue.

Stage B shows the division between the Database with Public Private Data (PPD) and the Database with Public Open Data (POD). PPD cannot be shared with external parties for legal reasons, such as privacy of personal data (names, diseases, etc.) and confidentiality of strategic governmental data (police department guns, etc.). For these reasons, a normalization and standardization was conducted on the PPD to create the POD. In the cases studies, the POD was freely accessible in open format at the Open Data Portal. The data can be accessed by downloading a dataset in the Comma Separated Values (CSV) file format, or by invoking the APIs.

The two databases POD and PPD enable the creation of two different flows with different objectives. While the PPD helps government to create internal strategies for public policy, the POD can be freely used by individuals and enterprises by accessing the Open Data Portal. The governmental flow of information is shown in red and continuous solid line (no dashes) (F2, F4, F6) and the public flow of information is shown in green and dashed with long dashes (F3, F5, F7).

After Stage B, both flows go to Stage C (Data Interpretation). Based on the case studies, normally Stage C includes the Big Data Analytics Processes (BDA). The statistical analysis and geographical analysis enable Stage D (Data Visualization) where the combined results from statistical and geographical analysis are presented. The F4 flow of information enables visualization of management reports and dashboards for civil servants and politicians. The F5 flow enables the creation of public dashboards. The examples found in the cases were public dashboards on streets revealing two options of routes with traffic congestion conditions and average time spent between points of interest in the city.

After the Stage D, this data is or is not used by the visualization options, and the flows of information end at Stage A, being collected again by means of forms and sensors in the city. The F6 flow of information can be used by politicians and the F7 flow of information can be used by the public. This usage can impact decision-making on public policies by politicians and traffic conditions by the public. If used, the city's sensors will probably capture the difference in patterns or civil servant decisions. Both can be identified in Stage C during a second cycle, comparing the history of Big Data Analysis of traffic jams and evaluating whether or not a public policy decision was effective. This can be done by using real-time data on the same day or a history of days, for example.

The bases of public dashboards in governments are presented in the middle of the cycle. The first concerns the legal and institutional requirements to provide big data analysis and open government data. The second concerns Information and Communication Technologies (ICT) providing the infrastructure, the architecture. And the third concerns Human Resources (HR) to conduct the processes of collection, storage, analysis, and visualization.

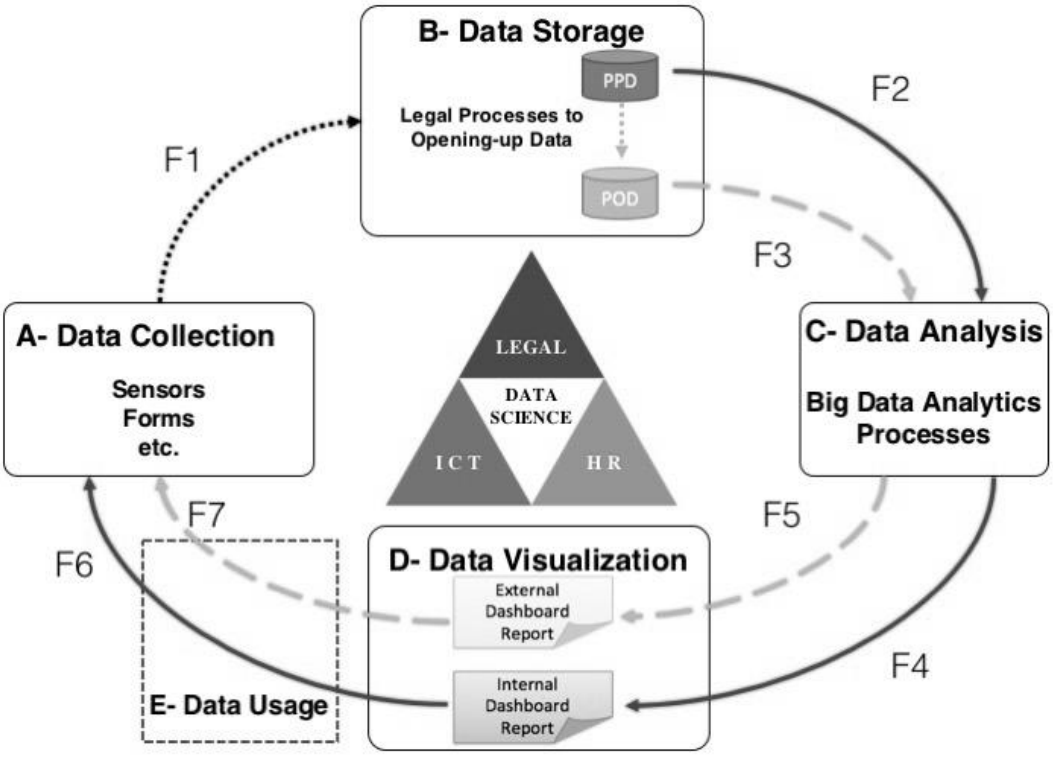


Figure 24 – Data cycle for dashboards

The cycles in Figure 24 suggest that dashboards need to be continuously updated. One cycle opens data for citizens, which should result in higher levels of transparency. For citizens, various views and visualizations can be created. Due to privacy issues and for as data might be sensitive not all data might be indiscriminately shared. The second cycle opens data for a limited group of policy-makers. In this way this data can be used to inform policy-making and other decisions, however, this does not result into transparency for citizens. There is still information asymmetry, as citizens still have less information than governments have. As a result, not being able to open all data might result in less transparency and trust in government.

Dashboards should help citizens to create an understanding of the situation at hand avoid long-lasting search processes and an information overload. Due this reason, the creation of dashboards for citizens need a design focused on relevance and take the citizens' situation into account. Balancing issues such as privacy, information overload and designing an overview at a glance is challenging. A good dashboard is user-centric and provides insight, however, for some citizens, this might not be sufficient. Although dashboards in different areas can have different shapes and forms, there might be users that want to gain access to the raw data. Due

to the diversity of possible users of data, it is paramount that both the societal issues that need to be addressed and the users' needs are elicited.

3.3.6 Conclusions

Government can use dashboards to empower citizens, in this way creating 'smarter' citizens. Data is useless unless it can be used to create public value, and actions can be taken based on the findings of the data analysis. Dashboards are necessary for creating the public values of transparency and accountability. Dashboards are the missing linking between govenrments and the public. Nevertheless, balancing user expectations and needs with issues such as privacy can be challenging when creating dashboards. User needs and the societal challenge addressed should be clear. The design of dashboards is challenging and principles that can help to design dashboards have been identified in the literature. Although these principles do not necessarily need to be followed, they can help guide dashboard design in the right direction.

Data science is about the extraction and finding of new insights through the use of data. Dashboards are instruments for presenting data as a comprehensive single visual display. Dashboards are likely to become more important with the availability of more and diverse data. Many apps are in fact interactive dashboards. Although the benefits of dashboards are intuitively clear, achieving these benefits might be cumbersome. Dashboards can play a crucial role in providing insights into a situation and helping the situation to improve and evolve. The identified benefits of dashboards are related to creating an overview and being able to zoom in on the details. Transparency can be created by overcoming information asymmetry between public organizations and the public. The benefits can only be gained when the dashboards are properly designed. The case studies show that the actual achievement of these benefits might be difficult.

Our findings show that the introduction of dashboards might be useless if their introduction is not accomplished with organizational changes. Finding new insights or detecting corruption is useless if there are no means for follow up actions. Formal authorities that can hold organizations accountable need to be involved or created. Our case study findings also show that, although dashboards are often used for policy evaluation, dashboards can support the complete policy-making cycle, including policy formulation, implementation, and evaluation. Engagement in dashboards, with citizens having the opportunity to provide data and discuss results, plays a crucial role in achieving the benefits. Furthermore, a loosely designed

dashboard might result in misunderstanding of data and can affect the public's trust in the government.

3.3.7 Annexes

3.3.7.1 Annex A – Interview Protocol Form

Introduction

You are selected as respondent of this interview to contribute on the paper "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities". This research aims to understand and to support the design of dashboards for creating transparency and accountability. We argue that to achieve certain level of transparency, dashboards can improve transparency and accountability, however, realizing these benefits was cumbersome and encountered various risks and challenges.

Essentially, this document states that:

(1) all information will be held confidential;

(2) your participation is voluntary, and you may stop at any time if you feel uncomfortable; and

(3) we do not intend to inflict any harm.

Thank you for your agreeing to participate!

We have planned this interview to last about 30 minutes. During this time, we have questions that we would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

QUESTIONS

1. Did you create a Transparency Dashboard?

( ) Yes ( ) No

2. What were the objectives when you create a Transparency Dashboard?

( ) Transparency ( ) Accountability ( ) Accountability

( ) Other: _____

3. What were the expected public values when you create a Transparency Dashboard?

Open Question

4. What were the expected benefits when you create a Transparency Dashboard?

Open Question

5. What were the expected risks and challenges when you create a Transparency Dashboard?

Open Question

6. Which databases did you use in your Transparency Dashboards?

Open Question

7. Can you describe all functionalities in your Transparency Dashboards?

Open Question

8. Explain how your Transparency Dashboards work. Can you draw a flow of these Transparency Dashboards?

Open Question

3.4 Paper 4 – Design Principles for Creating Digital Transparency in Government

This chapter is based on: Matheus, R., Janssen, M., & Janowski, T. (2021). Design principles for creating digital transparency in government. Government Information Quarterly, 38(1), 101550.

The first author conducted the literature review, designed and conducted the data collection and analysis process, and wrote the manuscript. The co-authors provided feedback on the data collection and analysis process and earlier versions of the manuscript. We thank the anonymous reviewers for their constructive comments and recommendations.

This paper was published in the journal Government Information Quarterly (GIQ) and it is freely available: https://www.sciencedirect.com/science/article/pii/S0740624X20303294?dgcid=rss_sd_all.

Abstract

Under pressure to fight corruption, hold public officials accountable, and build trust with citizens, many governments pursue the quest for greater transparency. They publish data about their internal operations, externalize decision-making processes, establish digital inquiry lines to public officials, and employ other forms of transparency using digital means. Despite the presence of many transparency-enhancing digital tools, putting such tools together to achieve the desired level of digital transparency, to design entire government systems for digital transparency, remains challenging. Design principles and other design guides are lacking in this area. This article aims to fill this gap. We identify a set of barriers to digital transparency in government, define 16 design principles to overcome such barriers, and evaluate these principles using three case studies from different countries. Some principles apply to projects, others to systems, yet others to entire organizations. To achieve digital transparency, before building and deploying digital solutions, government organizations should build technological and institutional foundations and use such foundations to organize themselves for transparency. The proposed design principles can help develop and apply such foundations.

3.4.1 Introduction

Lack of transparency in government operations and decision-making processes is often connected to corruption scandals (Harrison & Sayogo, 2014), poor decision-making (Guillamón et al., 2016), lack of accountability of public officials (Rui Pedro Lourenço, 2015), and dysfunctional governance of government organizations (Kosack & Fung, 2014). Transparency is often viewed as one of the critical conditions for good governance and an essential mechanism for balancing power between the government and the public (Janssen & van den Hoven, 2015). Transparency increases the chances that wrongdoings are detected, abuses of power uncovered, and activities scrutinized.

Although easy to grasp intuitively, transparency is hard to define and even harder to realize. Various definitions and conceptualizations of transparency emphasize different aspects and formulate different expectations towards this concept. The latter include improved accountability (Peixoto, 2013), good governance (Ward, 2014), better decision-making (Navarro-Galera et al., 2016), less corruption (J.C. Bertot et al., 2010), and more openness (Frank & Oztoprak, 2015; Ricardo Matheus & Janssen, 2015a). At the same time, an argument is also advanced that the expectations towards digital technology to help create transparency in government are unrealistically high (Bannister & Connolly, 2011a).

Digital transparency refers here to government organizations relying on digital technologies and networks to become more transparent. Digital transparency is often viewed as an effective and low-cost way to create insights into government operations and decisions. Such transparency is part of the broader open government agenda, which purports to improve openness, transparency, and accountability of government decision-making, to increase citizen engagement and trust in government (K. Janssen, 2011; Ubaldi, 2013). A common mechanism for digital transparency is opening government data to the public (Luna-Reyes et al., 2014) through portals, dedicated apps or Application Programming Interfaces (APIs). An open data portal makes raw datasets available for human or machine use. An app provides an interface for exploring, analyzing, and visualizing data in this way, enabling the performance of tightly controlled operations on such data. Big data, data analytics, artificial intelligence (AI), and other data-driven algorithms that process and analyze available data and visualize the outcomes are behind such possibilities.

Despite its merits and the availability of relevant digital tools, full transparency is difficult to achieve (Fung, 2013), and the practical realization of digital transparency is challenging. First, opening government data alone is insufficient (Janssen et al., 2012) as many socio-technical barriers prevent the creation of digital transparency from such data (Conradie & Choenni, 2014). Second, while data can be opened and shared, it could create limited insights into government operations; more data might not automatically lead to more transparency. Third, as those in control commonly lead transparency initiatives, they base their decisions on available data but often fail to consider public needs (Janssen et al., 2012). Fourth, presenting selected and aggregated data, open government data portals might embed their designers' viewpoints (Kitchin, Lauriault, & McArdle, 2015) while suppressing the diversity of views held by different groups in a pluralistic society. Hence, such data might be unsuitable for creating accountability and combating fraud and corruption. Fifth, despite the many tools available to open up aspects of government operations and organization, these tools have their limitations and there is no guidance on how to use them to consistently achieve the desired level of digital transparency across government structures and operations.

Given the challenges above, this article aims to provide guidance for creating digital transparency in government. This guidance is offered through a set of design principles for digital transparency. The principles are intended to overcome the various barriers hindering digital transparency and create a window for the public to view the internal functioning of government. The principles make part of a *Window Theory* (Ricardo Matheus & Janssen, 2019), with many factors relevant to digital transparency and multiple windows offered to realize such transparency. According to Ricardo Matheus and Janssen (2019, p. 3), such a window is required "*to view government functioning, aimed at overcoming the information asymmetry between the government and the public*". The window metaphor captures different influences on who, how, and what we can inspect about government – users, conditions of use, data and system characteristics, etc. The metaphor also captures the fact that transparency goals should inform window design, but that no single window can deliver full transparency by itself.

The rest of this article is structured as follows. Section 3.4.2 presents the research approach. Section 3.4.3 identifies barriers to digital transparency, followed by design principles and how they help overcome the barriers in Section 3.4.4. Section 3.4.5 evaluates the principles using three case studies. A discussion of the principles and their use is carried out in Section 3.4.6. Finally, Section 3.4.7 provides some conclusions.

3.4.2 Design research approach

As our goal is to arrive at a set of design principles for digital transparency, we followed the Design Science Research approach (Chanson et al. (2019). Section 3.4.2.1 presents the Systematic Literature Review method, which is used to derive design principles, followed by the Case Study approach in Section 3.4.2.2, which is used to evaluate the design principles in different practical scenarios.

According to Chanson et al. (2019, p. 1277), the focus of the design science is "on the creation of the artificial and accordingly the rigorous construction and evaluation of innovative artifacts". Using the design science research methodology by Peffers et al. (2007, p. 48), Chanson et al. (2019) created a design cycle to build design principles. The latter "*instantiated by an explicit design feature can be understood as an explanation (design principle) of why a specified piece (design feature) leads to a predefined goal (design requirement)*" (ibid. p. 1279). Chanson et al. (2019) aimed at deriving design principles for a sensor data protection system.

In contrast, the artifacts in our research are digital systems used by government organizations. By following the design principles for digital transparency, a window on government decisions and operations can be created. This set of coherent and generalizable design principles for digital transparency comprises our design theory, which assumes and supplements the Window Theory (Ricardo Matheus & Janssen, 2019).

Whereas most design approaches take an inductive approach to derive general laws from particular instances, we opted for a deductive approach to derive specific instances from general laws. In particular, rather than analyzing concrete government systems to uncover barriers to digital transparency and develop design principles to overcome such barriers, we opted to discover such barriers and principles through literature. This decision was motivated by the many barriers and principles available in literature and their potential for generalizability. For the barriers and principles derived from working systems, achieving such generalizability is difficult. Furthermore, we opted to evaluate the principles using three case studies conducted in different countries and policy areas. The diversity of case studies aims to justify that the proposed design principles can be used to ensure digital transparency for various government organizations and their digital systems.

The research process, depicted in Figure 25 consists of five steps. In Step 1, a Systematic Literature Review (SLR) was conducted to uncover barriers to digital transparency in government organizations. A similar SLR was carried out in Step 2 to identify a set of design principles for overcoming the barriers. The principles were mapped in Step 3 into the Data-Driven Transparency cycle to ensure consistency, facilitate usage and help confirm which principles are relevant (Ricardo Matheus, Janssen, et al., 2018, p. 8). Next, Step 4 demonstrated and tested the principles using three international case studies. Each case study concerned the development of a digital system for a government organization, aimed at making this organization more transparent. Each case study involved conducting semi-structured interviews with experts working on such systems. Finally, Step 5 discussed practical applications of the design principles for digital transparency.

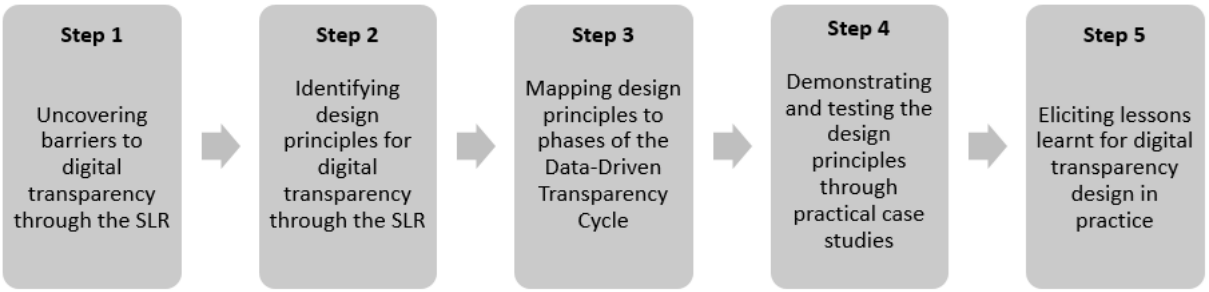


Figure 25 – Overview of the design research approach

3.4.2.1 Systematic Literature Review

According to Fink (2019, p. 6), a Systematic Literature Review is a "*systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars and practitioners*". Fink (2019, p. 6) also recommends conducting SLR through the seven following steps: 1) determine the research question, 2) identify literature sources, 3) define keywords and other search terms, 4) use explicit screening criteria to include or exclude papers, e.g., the papers that are written in specific language or published in particular years, 5) apply the screening criteria methodologically, here to identify the barriers and design principles to build digital systems for transparent government, 6) prepare reliable reviews of all selected articles using standardized

forms to ensure consistency and replication, and 7) synthesize the result into the lists of barriers and design principles.

The SLR for the first step of this research was conducted using the search term: ("big data" OR "open data") AND "barriers" AND "transparency" in four scientific databases – Scopus, JSTOR, SpringerLink and Web of Science – serving as the literature sources. As the inclusion criterion, we limited the search to the top 25 journals in the fields of Public Administration (PA) and Information Systems (IS) with an average impact factor above 1.0 based on the Scientific Journal Rank (SJR – Scimago/Scopus) calculated in 2016. We also limited the publication years to the period between 2007 and 2018.

The result of the SLR, which was conducted between 1 April and 31 May 2019, is a list of 50 relevant articles that helped uncover 364 barriers to digital transparency. The articles are listed in Table 31 and the barriers in section 3.4.3, the latter after categorizing them into political, economic, human and social, and technological areas.

Subsequently, another SLR was carried out to identify design principles that could be applied to build systems for digital transparency and thus overcome the barriers identified earlier. This SLR used the same literature sources and inclusion criteria but involved a different search term: "transparency" AND ("design" OR "architecture" OR "principle")

This search resulted in 29 articles, 22 of which proved to be relevant to this research. In particular, the papers documenting the results of biological or medical research were excluded. The 22 remaining articles were each independently read by two researchers to identify candidates for design principles.

3.4.2.2 Evaluating design principles through cases studies

Three international case studies from Belgium, Ireland and the UK were developed to evaluate the design principles. According to Yin (2013), a case study is an approach to answer questions about events outside the control of an investigator. They focus on contemporary phenomena within a real-life context.

Each case study demonstrated the development of digital systems using the design principles and their deployment within government organizations to make them more transparent. The

case study from Belgium concerned the development of the linked data app for the Flemish Environment Agency. The case study from Ireland discussed the development of the Irish National Tide Gauge Network by the Marine Institute. The UK's case study examined the story of the OpenGovIntelligence pilot for Trafford, a metropolitan borough of Greater Manchester, by the Trafford's Innovation and Intelligence Lab. As part of the case studies, policy-makers, information architects, data analysts, software engineers, and other stakeholders involved in development were interviewed about the use of the proposed design principles. The interview protocol applied in all case studies is presented in 3.4.5.

3.4.3 Barriers of Transparency

Many governments around the world are striving to employ digital means to become more transparent. In the process, they are confronted with different barriers, many of them related to the design of open data portals and applications (Philip Chen and Zhang (2014); Fan, Han, and Liu (2014); and Hu, Wen, Chua, and Li (2014)). Such barriers may result in the recalculation of costs and benefits, as well as lowering expectations towards the use of digital technology for increasing transparency (B. Worthy, 2010).

The aim of this section is to presents the barriers to digital transparency identified by the Systematic Literature Review outlined in Section 2.2. The 42 identified barriers were grouped into data quality barriers, economic barriers, ethical barriers, human barriers, political and legal barriers, organizational barriers, technical barriers, and usage barriers. The barriers, with categories and code names, are presented in Table 1 and described as follows:

- **Data quality** barriers include inaccessible or inaccurate data, information sharing or re-identification from combined data sets causing privacy violations, lack of unified ontologies and language misconceptions causing data misinterpretation, lack of centralized databases causing data quality issues, and difficulties of integrating data from heterogeneous sources.
- **Economic** barriers include high costs of maintaining big data infrastructures and tools for big data analysis, lack of reliable Return-on-Investment (ROI) studies, unreliable architecture plans leading to unpredictable cost increases, and limited organizational budgets.
- **Ethical** barriers deal with data bias and the resulting discriminatory decisions by data-driven algorithms as well as privacy issues related to uncovering human habits through mass surveillance, among others.

- **Human** barriers include lack of workforce able to handle big data and related projects, low quality of decision-makers and decision-making using big data analytics, and lack of data-driven and evidence-based work culture.
- **Organizational** barriers include lack of information sharing plans, unclear ownership of data, data quality issues causing mistakes or allowing misconduct by personnel, unavailable data, lack of information sharing policies causing information asymmetry, the opacity of algorithms and the inability to inspect them, and lack of awareness about the benefits of big data.
- **Political and legal** barriers include lack of privacy policies, mass surveillance causing lack of data protection, and lack of stable regulatory frameworks creating legal issues.
- **Technical** barriers include the need to process vast volumes of data; data volumes causing user overload; lack of methods for managing big data systems; difficult integration between big data and legacy technologies; untimely data delivery; underperformance of big data systems caused by bandwidth limitations and the lack of architecture plans; security breaches caused by the leakage or hacking of data; security risks caused by the unavailability of logs to carry out forensic analysis; data silos lowering data quality; problems with data accessibility; and lack of user-friendly big data tools.
- **Usage** barriers include difficulties in adapting visualizations to different audiences, and users' information overload causing data quality issues.

Table 25 – Barriers to digital transparency

Category	Code	Barrier
Data Quality	DQ1	Privacy issues due to information sharing risks
	DQ2	Data quality issues due to the lack of unified area ontologies
	DQ3	Data quality issue due to heterogeneous (structured vs unstructured) data sources
	DQ4	Data quality issue due to the lack of data accuracy
	DQ5	Privacy issue due to re-identification caused by combining data sets
	DQ6	Data quality issue due to the lack of centralized databases
	DQ7	Data quality issue due to language misconceptions, e.g., usage and jargon
Economic	EC1	The high cost of creating and maintaining big data analysis infrastructures
	EC2	Financial issues due to the lack of reliable Return-on-Investment (ROI) studies
	EC3	Lack of low-cost analytical tools to carry out big data analysis
	EC4	Lack of big data system architecture plans leading to unpredictable cost increases
	EC5	Financial issues due to limited organizational budgets
Ethical	ET1	Prejudicial use of algorithms, e.g., discrimination based on ethnicity
	ET2	Privacy issue due to human habits, ethics and culture
Human	HU1	Lack of skilled workforce able to handle big data
	HU2	Low quality of decision-makers and decision-making
	HU3	Lack of data-driven and evidence-based culture
	HU4	Lack of skilled workforce to lead big data projects
Organizational	OR1	Lack of information sharing plans
	OR2	Data quality issue due to unclear ownership
	OR3	Data quality issue leading to mistakes or allowing misconduct by personnel
	OR4	Lack of or limited availability of data
	OR5	Asymmetry of information due to the lack of information sharing policies
	OR6	Lack of openness and constraints on inspecting algorithms
	OR7	Organizational issues due to the lack of awareness about the benefits of data
Political and Legal	PL1	Privacy issues caused by the lack of explicit privacy policies
	PL2	Data protection issues caused by mass surveillance
	PL3	Legal issues due to the lack of stable regulatory frameworks
Technical	TE1	Difficulties in processing vast volumes of data
	TE2	The complexity of the integration between big data and legacy technologies
	TE3	Lack of appropriate methods to deal with modern big data systems
	TE4	Technical issue due to the volumes of big data, causing users' data overload
	TE5	Data quality issues due to the lack of timeliness in data delivery
	TE6	Underperformance due to the lack of big data system architecture plans
	TE7	Performance issues caused by bandwidth limitations

	TE8	Security issues caused by the risk of data leakage or hacking
	TE9	Data quality issues caused by existing data silos
	TE10	Lack of data accessibility
	TE10	Security issues due to the unavailability of logs to carry out forensic analysis
	TE12	Technical issues due to the lack of user-friendly big data tools
Usage	US1	Visualizations that are hard to adapt to different audiences
	US2	Data quality issues due to the users' information overload

3.4.4 Design principles for digital transparency

In this section, we propose a set of design principles that can help government organizations design and adopt digital systems through which they can become more transparent. Specifically, the principles are intended to overcome data quality, organization, and usage barriers, as these categories are central to building digital transparency portals and opening data for digital transparency. Although relevant, we excluded economic, ethical, human, political and legal, and technical barriers as these are not directly related to the organization and creation of digital transparency.

The rest of this section is structured as follows. Section 3.4.4.1 formulates 16 design principles for digital transparency based on the Systematic Literature Review. Section 3.4.4.2 relates the 16 principles identified in Section 3.4.4.1 to the 42 barriers identified in Section 3.4.3. The resulting many-to-many mapping describes which principles help to overcome which barriers. Finally, Section 3.4.4.3 maps the design principles to different phases of the data-driven transparency cycle (Ricardo Matheus & Janssen, 2018; Ricardo Matheus, Janssen, et al., 2018), thus operationalizing the use of the principles in the engineering for data-driven transparency.

3.4.4.1 Deriving design principles

Richardson, Jackson, and Dickson (1990, p. 388) described design principles as "beliefs upon which the enterprise is created and the bases of its decisions". Bharosa et al. (2011, p. 1) defined design principles as a means "*to guide stakeholders in proactively dealing with some of the transformation issues*" that organizations might encounter.

The Open Group Architecture Framework (TOGAF (2009, p. 1) prescribed that such principles should be easy to understand complete, consistent, stable, and enduring. To support sound decision-making, they should also be robust and precise. According to the TOGAF template – a standard way of defining design principles, each principle should have a name, statement, rationale and implications. The inclusion of the rationale and implications promotes the understanding and acceptance of the design principles throughout the organization (TOGAF, 2009).

The design principles derived in this section aim at creating digital transparency. They are intended to help organizations make the right decisions when realizing digital transparency. As such, they should be generalizable to different situations in which such decisions have to be made. The principles are described using the TOGAF template in Table 33 and summarized in Table 26 below.

Table 26 – Design principles for digital transparency

Code	Name	Short Name
P1	Separating privacy-sensitive and -insensitive data at the source	Privacy
P2	The openness of processes and actors	Openness
P3	Feedback mechanisms for improving transparency	Feedback Mechanisms
P4	Various levels of abstraction for data access	Data Abstraction
P5	Avoid any jargon or terms that the public does not understand	Comprehension
P6	Checking and rating data quality	Data Quality Rating
P7	Visualization of different views	Visualization
P8	Data access in different protocols	Data Access
P9	Use of standardized formats	Standardized Formats
P10	Ensuring that data is unaltered and its history can be traced	Data Persistency
P11	Data and system interoperability	Interoperability
P12	Include metadata for data comprehension	Metadata
P13	Transparency-by-design (automatically opening data)	Transparency-by-Design
P14	Opening of raw data	Opening of Raw Data
P15	Assigning stewards responsible for digital transparency	Stewardship
P16	Supporting views with different level of details	Gradation of Detail

3.4.4.2 Relating principles to barriers

The design principles for digital transparency, as described in 3.4.4.1, should help overcome the barriers to digital transparency, as described in Table 25. The matrix describing which principles address which barriers is presented in Table 27. According to this Table 27, most principles help overcome several barriers, and most barriers are addressed using multiple principles, which demonstrates the complexity involved with organizing and designing for digital transparency. Ignoring some design principles might limit our capacity to address specific barriers, thus lowering the level of digital transparency overall.

Table 27 – Relationships between barriers and design principles

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
DDQ1	X	X	X	X		X				X			X	X	X	
DDQ2					X	X				X	X	X		X	X	X
DDQ3	X	X		X		X	X	X	X	X	X	X	X	X	X	
DDQ4		X	X		X	X				X	X	X	X			X
DDQ5	X	X	X	X		X				X	X		X	X	X	
DDQ6	X	X	X		X	X		X	X	X	X	X	X	X	X	
DDQ7			X		X	X	X						X			
EEC1		X	X	X		X	X	X	X	X	X	X	X		X	X
EEC2	X	X	X							X	X		X	X	X	
EEC3							X	X	X	X	X	X	X	X	X	
EEC4													X			
EEC5	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X
EET1		X														
EET2		X														
HHR1	X	X	X	X	X					X		X	X	X	X	X
HHR2	X	X	X			X				X		X	X			
HHR3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HHR4		X	X										X		X	
OOR1		X	X	X	X	X		X		X	X	X	X	X	X	X
OOR2	X	X	X			X		X	X	X	X	X	X	X	X	
OOR3		X	X							X	X	X	X	X	X	
OOR4		X	X										X		X	
OOR5								X	X	X	X	X	X		X	
OOR6	X	X	X	X		X	X	X	X	X	X	X	X		X	
OOR7																X

PPL1	X															
PPL2	X	X											X	X	X	
PPL3	X	X	X							X	X	X	X	X	X	
TTE1	X			X		X	X	X	X	X	X	X	X	X		X
TTE2	X		X					X	X	X	X	X	X	X	X	
TTE3							X	X	X	X	X	X	X	X	X	
TTE4			X			X	X						X	X		X
TTE5			X							X	X	X	X	X	X	
TTE6				X			X	X	X	X	X	X	X	X	X	X
TTE7								X			X					
TTE8		X		X		X		X	X	X	X	X	X	X	X	X
TTE9	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
TTE10								X	X	X	X	X	X	X	X	X
TTE11	X	X	X	X		X	X	X	X	X	X	X	X	X	X	
TTE12				X				X	X	X	X	X				
UUS1			X	X	X	X	X	X	X	X	X	X	X	X		X
UUS2	X	X	X									X	X			

3.4.4.3 Transparency cycle enabled by design principles

To operationalize the development for digital transparency and the use of the design principles as part of it, we adopted the *data-driven transparency cycle (Ricardo Matheus & Janssen, 2018; Ricardo Matheus, Janssen, et al., 2018)*. The cycle is depicted in Figure 26, adapted from figure 8 of "OGI Tools and Working Flow" in Ricardo Matheus and Janssen (2018, p. 36). The cycle consists of six phases: eliciting data, collecting data, publishing data, using data, sharing results, and determining actions; and two parts: one on publishing data (light color, dotted outline) and another on using data (dark color, solid outline). In line with the iterative nature of development, the phases are ordered into a cycle.

During different phases of the data-driven transparency cycle, various design principles can be used. The assignment of the principles to phases, also depicted in Figure 26 and elaborated in Table 28, helps decide which principles should be used and when. Every phase has several principles assigned to it, and each principle can be mapped to different phases.

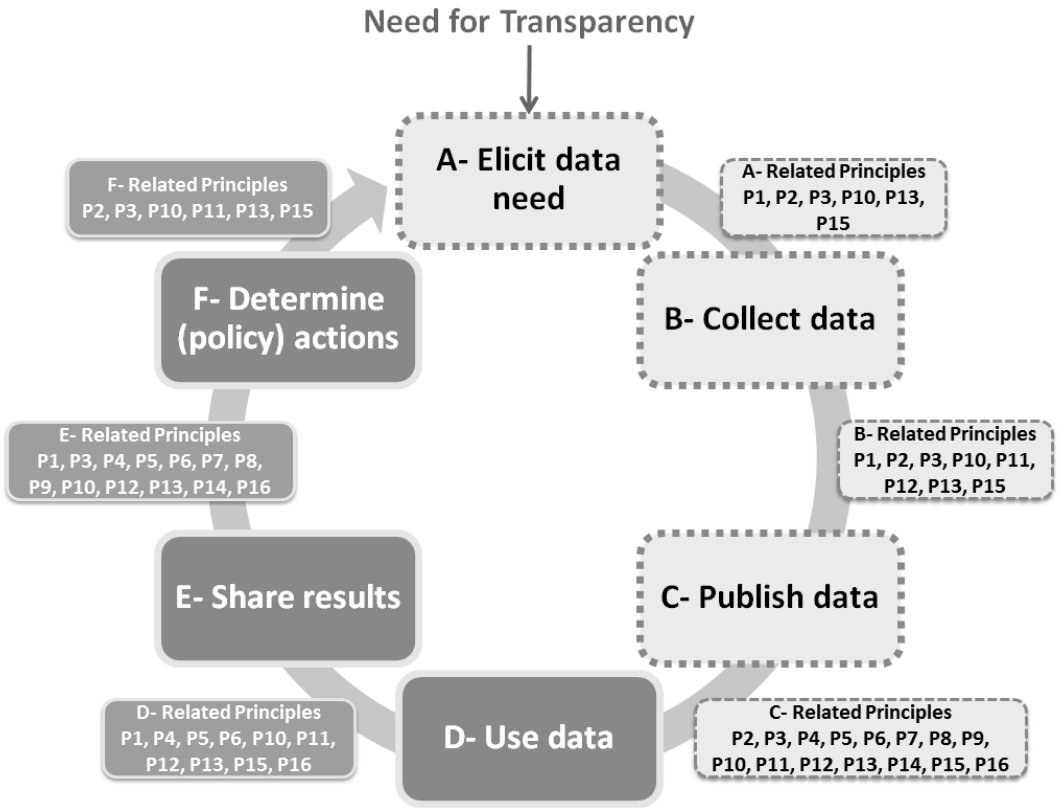


Figure 26 – Data-driven transparency cycle with design principles adapted from Matheus and Janssen (2018, p. 36)

Table 28 – Mapping design principles to the phases of the transparency cycle

#	Phase Name	Description / Justification	Related Principle Codes and Names	
A	**Elicit data need**	Any data created for whatever reason and the disclosure of this data is a transparency action.	P1 P2 P3 P10 P13 P15	Privacy Openness Feedback Mechanism Data Persistency Transparency-by-Design Stewardship
A	**Collect data**	Data must be collected in any form, from manual and physical (e.g., surveys), to automated and digital (e.g., networked sensors).	P1 P2 P3 P10 P11 P12 P13 P15	Privacy Openness Feedback Mechanism Data Persistency Interoperability Metadata Transparency-by-Design Stewardship
C	**Publish data**	A step to become transparent, data must be published (disclosed). Publishing data is at the heart of the Transparency Cycle.	P2 P3 P4 P5 P6 P7 P8 P9 P10 PD11 EP12 P13 P14 P15 P16	Openness Feedback Mechanism Data Abstraction Comprehension Data Quality Rating Visualization Data Access Standardized Formats Data Persistency Interoperability Metadata Transparency-by-Design Opening of Raw Data Stewardship Gradation of Detail
D	**Use data**	Transparency cannot happen if nobody uses data. After disclosure, users must use and create insights from data, as enabled by transparency.	P1 P4 P5 P6 P10 P11 P12	Privacy Data Abstraction Comprehension Data Quality Rating Data Persistency Interoperability Metadata

			P13 P15 P16	Transparency-by-Design Stewardship Gradation of Detail
E	**Share results**	Transparency can happen to only one person. However, the more people use data, the more will have insights enabled by transparency.	P1 P3 P4 P5 P6 P7 P8 P9 P10 P12 P13 P14 P16	Privacy Feedback Mechanism Data Abstraction Comprehension Data Quality Rating Visualization Data Access Standardized Formats Data Persistency Metadata Transparency-by-Design Opening of Raw Data Gradation of Detail
F	**Determine (policy) actions**	After a group of people gained meaningful insights enabled by transparency, policy action can be undertaken.	P2 P3 P10 P11 P13 P15	Openness Feedback Mechanism Data Persistency Interoperability Transparency-by-Design Stewardship

3.4.5 Demonstrating and testing design principles

In order to demonstrate and test their usefulness, the principles were employed in three case studies of government applications that aim at digital transparency. The case studies are outlined in Table 29 , including the responsible organization, application name and purpose, what kind of transparency effect is expected, and who is the target of this effect.

As part of this research, we carried out semi-structured interviews with designers involved in developing the applications, aimed at evaluating the principles. The interviews included questions belonging to different areas: the relevance of the principles; if and how the principles were used in the cases; and to which phase of the transparency cycle each principle belongs.

Table 29 – Overview of the case studies in digital transparency

	Case A	**Case B**	**Case C**
Country	Belgium	England	Ireland
Organization leader	The Flemish Environment Agency	Trafford's Innovation and Intelligence Lab	Marine Institute
Application name	Flemish Environment Agency Linked Data App (FELAP)	OGI – Trafford pilot prototype	Irish National Tide Gauge Network
Application purpose	To enhance environmental policy-making in terms of timely publication of the state of affairs related to the environment, to evaluate the policy of issuing permits, and to develop tools for benchmarking the pollution produced by companies in the same economic domain	To help support decision-making related to unemployment	To enhance the value of the marine data assets for scenario-building purposes by structuring and enriching the data with vocabularies and meanings to aid the extraction of scenario-related requirements
The expected effect of transparency	Accountability	Decision-Making	Co-Creation
Target groups	1. National, regional and local government 2. Enterprises 3. Citizens	1. Department for Work and Pensions 2. Trafford's Economic Growth Team 3. Greater Manchester Combined Authority	1. Civil servants in the Marine Institute 2. Enterprises in the leisure sector 3. Programmers in the maritime sector
Number of respondents	Three designers involved with the case study	Three designers involved with the case study	Three designers involved with the case study

Although all principles were used by at least one person in charge of application development in the case studies, who all found them coherent, the survey showed that the principles were used to various extent. Table 30 summarizes the percentage of the use of different principles by the nine interviewed designers.

All designers used the Privacy (P1) and Metadata (P12) principles; some principles were used occasionally, e.g., Stewardship (P15) at 33%, Comprehension (P5) at 44% or Transparency-by-Design (P13) at 56%; and some were not used at all. Interviews revealed that the reasons for this were that the principles primarily concerned organizational changes, whereas the projects were on application development. This disparity did not make them less relevant; on the contrary, the interviewees suggested that adhering to them is needed to create digital transparency.

Stewardship (P15) refers to the ownership of and responsibility for data quality. Adhering to this principle has considerable organizational consequences and requires organizational changes. An interviewee noted that following this principle would be "major, if well done". Although application designers could hardly use this principle, it was found to be highly relevant. Often strategic projects commence as technical software development, having no mandate to change an organization. This observation suggests that policy-makers and managers need to listen better to their developers to create digital transparency. An interviewee mentioned that it is "easy to allocate responsibilities, but organizational change might be needed". The evaluation even suggested that it is imperative to prepare an organization for transparency before developing systems. Following this suggestion should ensure that data is collected and becomes immediately available at the right quality and in the proper format. Organizing can be viewed as a precondition for creating digital transparency.

Comprehension (P5) is about avoiding jargon or technical terms to ensure that the public can understand them. Removing jargon requires everybody to agree to use the same terms and to provide these terms with the same meaning. However, principle P5 goes beyond the use of jargon. It also covers the harmonization of data collection to ensure that the data is understood and ready to be compared.

Table 30 – Design principles used when building applications

			Usage	
Rank	**Design principles**		**Number of designers**	**Percentage of designers**
1	P1	Privacy	9	100%
2	P12	Metadata	9	100%
3	P8	Data Access	8	89%
4	P9	Standardized Formats	8	89%
5	P11	Interoperability	8	89%
6	P7	Visualization	7	78%
7	P10	Data Persistency	7	78%
8	P14	Opening of Raw Data	7	78%
9	P2	Openness	6	67%
10	P3	Feedback Mechanisms	6	67%
11	P4	Data Abstraction	6	67%
12	P6	Data Quality Rating	6	67%
13	P16	Gradation of Details	6	67%
14	P13	Transparency-by-Design	5	56%
15	P5	Comprehension	4	44%
16	P15	Stewardship	3	33%

Figure 27 plots the 16 design principles on two orthogonal dimensions – ease of use in practice and importance for creating digital transparency. Some principles, particularly Opening of Raw Data (P14), Data Abstraction (P4), Stewardship (P15), Visualization (P7), Data Access (P8), and Feedback Mechanisms (P3) are both essential and easy to use. Thus, organizations could adopt them with little effort and achieve significant progress towards digital transparency. However, to realize stewardship is more than just allocating responsibilities on a drawing board, it has important organizational implications.

In contrast, some principles were found to be less relevant and challenging to use. This category includes Standardized Formats (P9), Openness (P2), Data Quality Rating (P6), Comprehension (P5), Privacy (P1) and Transparency-by-Design (P13), all located in the bottom right quadrant of Figure 3. The interviewees judged them as less important for the projects, difficult to put into practice and requiring much effort to do so. However, for the organizations they can be

essential to ensure that high quality data is automatically opened and can be easily used. Transparency-by-Design (P13), for instance, is essential to create digital transparency and for automating the opening of data, but the projects are focused on patching rather than organizing for Transparency-by-Design. As such, these principles go beyond a single project and might be important for policy-makers. For example, formatting all datasets in a standardized way is vital for comparison but is expensive and time-consuming for a single project. An interviewee pointed out that the ease-of-use is dependent on how data collection and processing are organized: "*if these [formats] are available then it is easy, if they are not then first a standardization process is needed*". Also, Openness (P2) might be hard to adopt. According to one interviewee: "*some agents are very reluctant to be exposed*" and "*it is not always easy to track who has done what*". The latter influences how easy it is to apply this principle in practice.

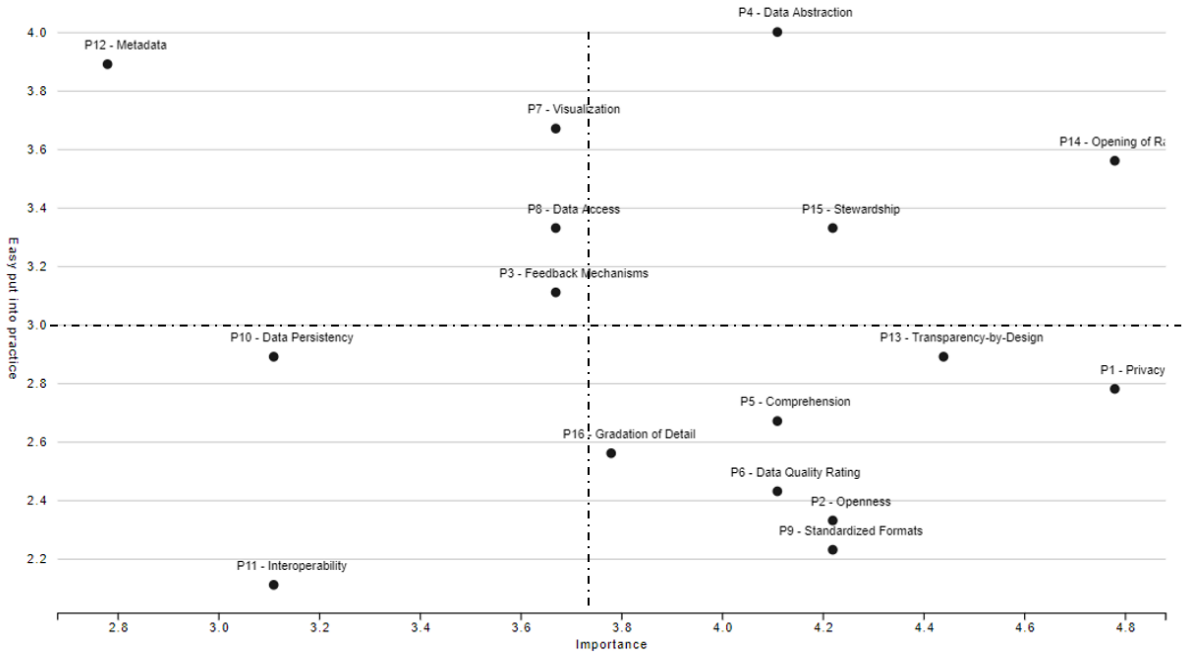


Figure 27 – The ease-of-use and the importance of design principles

Figure 28 plots the design principles against two other dimensions: impact on the organization and importance for achieving digital transparency. The top right quadrant includes all high-importance and high-impact principles, particularly: Privacy (P1), Stewardship (P15), Data Quality Rating (P6), Standardized Formats (P9), Transparency-by-Design (P13), Opening of

Raw Data (P14), Openness (P2), Gradation of Details (P16), Data Access (P8) and Comprehension (P5).

For example, the General Data Protection Regulation (GDPR) was used as the primary motivation by one interviewee for ranking P1 as highly important and having a high impact on the organization. Another interviewee noted: "*If not done properly, credibility is lost and as a result, none or fewer data will be opened*". Similar to P1, an interviewee noted about P6: "*if the transparency portal has no data quality for some datasets, this reduces the trust of people, and they might not use the good quality data in the future. This reduces transparency*".

The bottom-left quadrant in Figure 28 comprises low-impact and low-importance principles, particularly Metadata (P12), Interoperability (P11), Data Persistency (P10), Feedback Mechanisms (P3) and Visualization (P7). It is surprising to see Metadata (P12) in this quadrant, as metadata is often found to be a key contributor. One interviewee pointed out that "*Without proper metadata, it is quite difficult to understand the dataset. Sometimes we have access to data without metadata and is impossible to discover what the variables and observations mean*". This comment is contrasting with another interviewee who recommended following *"ISO 19157 to achieve a high metadata quality*". Various reasons may explain different answers. In some domains, meta-data standards are available; in others, they are not. Another reason for the low scoring of metadata is that digital transparency initiatives generally focus on a few datasets. In contrast, the more datasets are used, the more important metadata becomes to handle them. Concerning Feedback Mechanisms (P3), an interviewee considered this principle of low importance as "*it depends on the data. So sometimes it is essential and sometimes not*", following a quest to monitor "*what is done with the data*". The interviewee comments suggest that the design principles' impact and importance are context-dependent. However, more research is needed to understand and explore this direction.

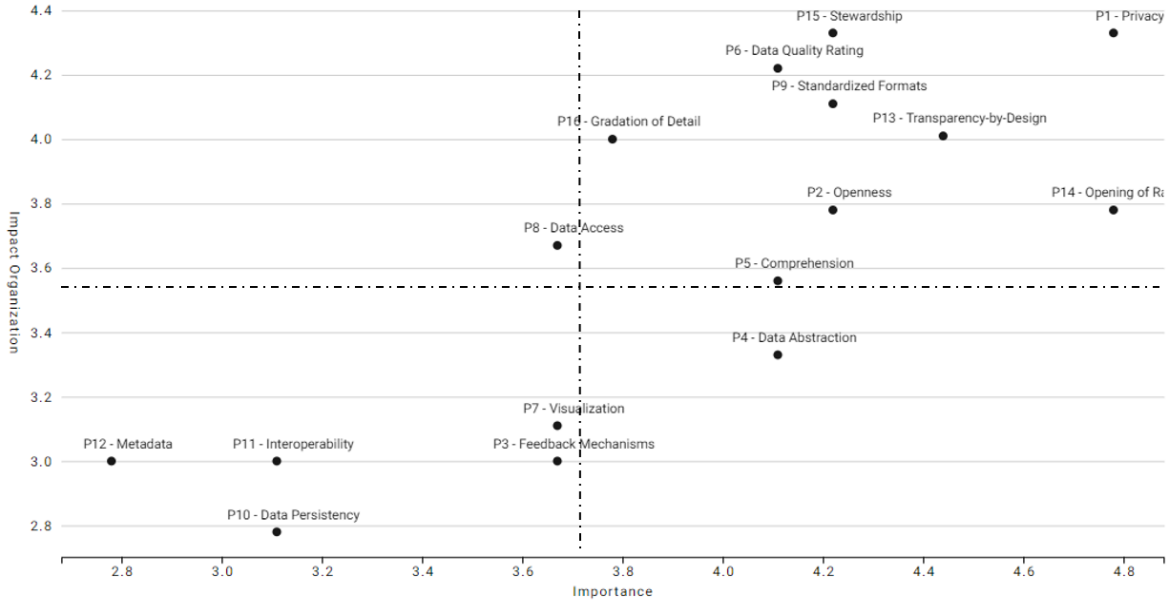


Figure 28 – Organizational impact and the importance of design principles

3.4.6 Discussion

3.4.6.1 Do the design principles always result in digital transparency?

Disclosing data does not by itself result in digital transparency, accountability, or openness (Ricardo Matheus & Janssen, 2015a). Therefore, this article proposes a set of 16 design principles that form a design theory that can help guide the development of systems for digital transparency. To ensure that their contribution to accomplishing digital transparency is well understood, the principles are described in Table 33 using the TOGAF template (TOGAF, 2009).

The principles should be interpreted and used depending on the context, particularly the organizational context. Creating digital transparency is not limited to technical issues associated with developing systems. It also includes organizational changes and creating organizational conditions for digital transparency. For instance, the Privacy (P1) principle of separating privacy-sensitive and -non-sensitive data will influence how personal and non-personal data are separately collected at the source. More research is needed about organizational conditions for digital transparency.

Creating transparency through digital systems can only succeed when such systems are used. While building systems for diverse groups of users consumes money, time, people, and other

resources, it also increases the chances for them to be popular with many users who have different needs and expectations. To build such systems, implementing technical features is necessary. Regular users expect easy navigation, which utilizes the well-designed User Interface (UI) and User Experience (UX), related to Visualization (P7). Experienced users might also want to access data through different protocols related to Data Access (P8) and Standardized Formats (P9). This expectation, however, will influence the back-end organization, which must be ready for including this type of functionality in the front-end.

Adhering to the design principles might be more far-reaching for governments. Openness (P2) and Feedback Mechanisms (P3) connect systems for digital transparency with open data use. Feedback mechanisms will influence the front-ends of transparency portals, to include mailboxes or participation buttons for users to submit criticism and suggestions for improvement. It will also affect the back-end since the organizations must be open and ready to listen to users and promptly respond to complaints and suggestions. As a result, substantive organizational changes will be required.

3.4.6.2 Is full transparency possible or desired?

While full transparency is often viewed as impossible (Fung et al., 2007), it might not be even needed or desirable. To make a decision transparent, we only need to know the information on which the decision is based and the rules applied to reach this decision. Providing other types of information about the decision-making process might not add value and instead can produce an information overload. In order to create the desired level of transparency, it is vital to open the right type of information, in the right way, and to the right audience.

Full transparency might conflict with other public values, like privacy or trust, and might easily result in the released information being used for other purposes than those intended. As a concept, transparency is multidimensional and might be highly subjective. Different users might have different expectations of how transparency should be implemented, with personality, experience, culture, social values, and other structural factors all influencing such expectations. For example, a Chilean case study (González-Zapata and Heeks (2016) showed that previous decisions (experience) play a major role in how transparency initiatives are implemented.

Full transparency can also bring undesirable effects, including opportunities for large-scale surveillance, lack of accountability for the results of consequential decisions made by inscrutable algorithms, bias and discrimination against groups affected by such decisions, etc. To protect users again such effects, our design principles, particularly Privacy (P1), include the protection of personal data. However, when designing systems for public use, such protection might result in trade-offs between transparency and privacy (Janssen & van den Hoven, 2015). Some mechanisms, though, can simultaneously help release data and ensure privacy. Specific design principles for this possibility should be developed.

Another reason why digital transparency can have undesirable effects is the uncertainty about how transparency-generated information will be used. The paradox of digital transparency is that the data opened to make systems and organizations transparent can be used in opaque ways. For example, algorithms might be used to process open data and make decisions that are difficult or impossible to explain (Nograšek & Vintar, 2014), that discriminate certain social groups (Chander, 2016), that draw conclusions that are inaccurate or incorrect. Also, introducing abruptly high levels of transparency in organizations experiencing systemic corruption might destroy trust in them by their constituencies (Bannister & Connolly, 2011a).

3.4.7 Conclusions

Creating digital transparency is a significant challenge faced by governments. Merely opening data does not result in digital transparency and might only result in information overload for those wanting to examine such data. In order to create digital transparency, a transparency window should be designed to enable looking at different aspects and from different perspectives of the organization.

This article proposes a set of 16 design principles for digital transparency, which can help overcome a set of well-recognized barriers to such transparency. The principles, organized into a six-stage transparency cycle to facilitate practical applications, can guide government organizations in how they can improve their levels of transparency by digital means. Some principles are relevant to projects, others to systems, yet others to entire organizations. The latter have long-term implications for the organizations and lay the foundations for their digital transparency.

The case studies provided several lessons about the use of such principles. Although all identified principles proved relevant for digital transparency, some were easier to adhere to than others, some were more important for digital transparency than others, and some had more impact on the organizations than others. All designers interviewed used the principles, like protecting privacy and providing metadata, in all case studies. Other principles, such as the opening of raw data, data abstraction, stewardship, visualization, data access, and incorporation of feedback mechanisms, proved both important and easy to use. Yet, other principles were scarcely used in the projects because they required organizational changes or technical foundations like data standardization and harmonization. This diversity of usage scenarios shows that creating digital transparency should be approached as an organizational rather than a system development challenge only.

The design principles are generic and need to be contextualized for an organization intending to use them. In further research, the principles could be used as a kind of guide or even regulation. Furthermore, the set of principles could be refined by adding new principles and modifying existing ones, as new initiatives will likely create new insights and influences. Although the principles proposed in this article focus on creating data-driven transparency, they could also be used as a basis for creating transparency using Artificial Intelligence (AI) tools. Future research could explore this possibility and refine and extend the principles to AI-driven transparency, considering both public and private sector application scenarios. The principles should also be tested in practice considering different economic, human, political, and legal contexts and barriers that were not considered in this research. Finally, the principles would likely be insufficient for achieving higher levels of digital transparency by themselves. Other factors, like willingness, leadership, capabilities, and resources, play important roles as well.

3.4.8 Annexes

3.4.8.1 Annex A – List of Papers containing barriers to digital transparency

Table 31 – List of papers containing the barriers to transparency

Paper ID	Source	Paper ID	Source
1	Sivarajah et al. (2017)	31	Angrave, Charlwood, Kirkpatrick, Lawrence, and Stuart (2016)
2	Rubinfeld and Gal (2017)	32	Philip Chen and Zhang (2014)
3	O'Connor and Kelly (2017)	33	Dwivedi et al. (2017)
4	Arunachalam, Kumar, and Kawalek (2018)	34	Oussous, Benjelloun, Ait Lahcen, and Belfkih (2017)
5	Alharthi et al. (2017)	35	Lee (2017)
6	Al-Qirim, Tarhini, and Rouibah (2017)	36	Jin, Wah, Cheng, and Wang (2015)
7	Hammond (2017)	37	Rogge, Agasisti, and De Witte (2017)
8	Hardy and Maurushat (2017)	38	Thiago, Victor Diogho Heuer de, and Ana Paula Cabral Seixas (2017)
9	De Laat (2017)	39	Ricardo Matheus, Janssen, et al. (2018)
10	Kourtit and Nijkamp (2018)	40	Pelucchi, Psaila, and Toccu (2017)
11	Wu, Zhu, Wu, and Ding (2014)	41	Cumbley and Church (2013)
12	George, Haas, and Pentland (2014)	42	Janssen and van den Hoven (2015)
13	Bello-Orgaz, Jung, and Camacho (2016)	43	Bertot et al. (2014)
14	Fan et al. (2014)	44	Brayne (2017)
15	Hu et al. (2014)	45	Salonen, Huhtamäki, and Nykänen (2013)
16	Lycett (2013)	46	Joseph and Johnson (2013)
17	Perera, Ranjan, Wang, Khan, and Zomaya (2015)	47	Choudhury, Fishman, McGowan, and Juengst (2014)
18	Schoenherr and Speier-Pero (2015)	48	Amugongo, Nggada, and Sieck (2016)
19	Couldry and Turow (2014)	49	Zicari (2014)
20	Elragal (2014)	50	Wielki (2013)
21	Fairfield and Shtein (2014)		
22	Wang, Liu, Kumar, and Chang (2016)		
23	Mittelstadt and Floridi (2016)		

24	Zakim and Schwab (2015)
25	Roski, Bo-Linn, and Andrews (2014)
26	Nativi et al. (2015)
27	Fernández et al. (2014)
28	Gil and Song (2016)
29	Clarke (2016)
30	Kruse, Goswamy, Raval, and Marawi (2016)

3.4.8.2 Annex B – List of barriers to digital transparency

Table 32 – List of barriers to digital transparency

#	Category	Code Barrier	Barrier Name	Description	Cite Count	Sources
1	Human resources	HR1	Lack of skilled people to work with big data	Organizations face a scarcity of talented people to work with big data, influencing on the transparency.	27	1, 3, 4, 5, 6, 7, 11, 12, 18, 21, 22, 23, 24, 25, 26, 30, 31, 35, 39, 40, 44, 45, 46, 47, 48, 49, 50.
2	Technical	TE1	Difficulties to process a huge amount of data	The huge amount of data is a technical barrier to deal with big data analytics.	25	1, 2, 3, 5, 6, 10, 12, 13, 14, 15, 18, 20, 23, 24, 25, 26, 27, 28, 30, 31, 39, 43, 45, 46, 48
3	Economical	EC1	High cost to create and maintain big data analysis	There is still a high cost to create and maintain big data analysis.	25	1, 2, 4, 5, 11, 13, 14, 15, 18, 21, 22, 23, 24, 25, 26, 28, 30, 31, 32, 33, 34, 37,

						38, 41, 49, 50
4	Technical	TE2	Complex integration between legacy and big data technology	It is hard to combine legacy systems with big data technologies	21	1, 2, 4, 5, 6, 10, 11, 14, 15, 20, 27, 28, 30, 33, 36, 37, 38, 39, 44, 46, 48
5	Data Quality	DQ1	Privacy issue due information sharing risks	Privacy issues due information sharing risks	13	23, 24, 25, 30, 33, 34, 35, 37, 38, 39, 43, 44, 50
6	Human resources	HR2	Low quality of decision-makers	Decision-makers don't perform well when using big data	13	1, 6, 8, 10, 12, 13, 15, 16, 20, 30, 31, 39, 50
7	Data Quality	DQ2	Data Quality issue due lack of unified ontology in the area	There is no unified ontology to reduce data quality issues	11	6, 7, 10, 11, 14, 15, 16, 28, 30, 33, 47
8	Usage	US1	Hard to adapt visualization to wide audience	A wider audience difficult to create transparency on big data projects	11	3, 7, 13, 15, 19, 20, 21, 23, 26, 32, 33
9	Human resources	HR3	Lack of data-driven culture	Lack of data-driven culture influences on big data projects	11	1, 2, 3, 4, 5, 6, 18, 31, 41, 49, 50

10	Data Quality	DQ3	Data Quality issue due multiple types of data sources (unstructured vs structured databases)	Unstructured and structured datasets influencing big data projects	11	1, 5, 30, 37, 38, 40, 41, 45, 46, 49, 50
11	Data Quality	DQ4	Data Quality issue due lack of accuracy	Lack of accuracy influences data quality and big data projects	10	2, 30, 32, 34, 37, 38, 43, 44, 45, 49
12	Economical	EC2	Financial issue due lack of reliable return on investment (ROI) studies	Unclear ROI of big data projects	10	3, 4, 6, 12, 14, 15, 16, 18, 20, 35
13	Data Quality	DQ5	Privacy issue due re-identification combining data sets	Privacy issues when combining different datasets to identify people	10	1, 2, 5, 8, 14, 16, 41, 42, 44, 49
14	Organizational	OR1	Lack of Information sharing plan	Organization has no information sharing plan and/or culture to help transparency and big data projects	9	1, 3, 4,11, 12, 13, 31, 42, 47
15	Organizational	OR2	Data Quality issue due ownership	Private or unclear ownership influences transparency and big data	8	2, 23, 25, 30, 37, 43, 45, 50
16	Data Quality	DQ6	Data Quality issue due lack of centralized databases	Lack of centralized databases influences transparency and big data	8	3, 4, 24, 40, 42, 43, 44, 45
17	Political and Legal	PL1	Privacy issue due lack of privacy policy	There is no privacy policy for transparency and big data projects	7	5, 8, 11, 12, 13, 14, 16
18	Technical	TE3	Lack of appropriated methods to deal	Methods to deal with big data are still on initial stage of development	7	1, 4, 5, 6, 29, 37, 42

			with modern Big Data systems			
19	Political and Legal	PL2	Data protection issues due Mass surveillance	Risk of big data for mass surveillance purposes	7	1, 6, 8, 17, 41, 42, 44
20	Technical	TE4	Technical issue from big data volume creating data overload to user	Huge amount of data leading to data overload	7	7, 11, 13, 14, 15, 37, 42
21	Technical	TE5	Data Quality issue due timely issues	Data is not accessed or published within the desired time	7	4, 30, 32, 37, 42, 43, 45
22	Organizational	OR3	Data quality issue leading to mistakes or misconducts	People make mistakes or misconduct when processing and using data influencing transparency	5	8, 14, 39, 42, 49
23	Economical	EC3	Lack of low cost analytical tools for big data analysis	The market has a few number of free or with low cost analytical tools to deal with big data	5	2, 5, 6, 34, 46
24	Technical	TE6	Lack of Performance due lack of Big Data system architecture plan	Organizations has no big data architecture plan influencing on transparency-by-design	5	5, 6, 42, 43, 45
25	Technical	TE7	Performance issue due bandwidth	There is no bandwidth available to perform big data projects	5	2, 5, 13, 14, 26
26	Technical	TE8	Security issue due chances of leaking and hacking	Organizations are not prepared for leaking and hacking issues	5	2, 5, 13, 14, 26
27	Economical	EC4	Lack of Big Data system architecture plan leading to cost increase not predicted	Lack or not well designed big data architectures lead to increase not predicted costs	4	1, 6, 30, 35

28	Technical	TE9	Data Quality issue due existence of data silos	Data silos influence on big data an transparency	3	2, 42, 50
29	Data Quality	DQ7	Data Quality issue due language barriers such as use or jargons	Language barriers such as jargons influence data quality, big data and transparency	3	3, 30, 42
30	Usage	US2	Data Quality issue due overload of information	Overload of information can lead users to mistake	3	2, 46, 49
31	Organizational	OR4	Lack of data available	There is no data available	3	8, 18, 21
32	Human resources	HR4	Lack of skilled people to lead big data	There is a few people skilled to lead big data projects and create transparency	3	1, 39, 42
33	Political and Legal	PL3	Legal issue due lack of stable regulatory framework	There is no stable regulatory framework for big data and transparency	3	2, 18, 33
34	Organizational	OR5	Asymmetry of information due lack of information sharing policy	Lack of information sharing policy leading to asymmetry of information influencing on big data performance and transparency	2	2, 3
35	Technical	TE10	Lack of accessibility of data	Data has low level of accessibility	2	43, 45
36	Economical	EC5	Financial issue due limited budget of organizations	Organizations have a limited budget for big data and transparency	2	3, 4
37	Organizational	OR6	Lack of openness of algorithms	Algorithms used on big data are not transparent	2	1, 5
38	Organizational	OR7	Organizational issue due lack of awareness about Big Data improvements	People are not aware about benefits big data and transparency can bring to organizations	2	5, 6

39	Ethical	ET1	Prejudice due bad use of algorithms such as discrimination of ethnicity	Algorithms can have prejudice such as discrimination of ethnicity	2	8, 9
40	Ethical	ET2	Privacy issue due human resource habits, ethics and culture	Culture influences bad habits that can lead to privacy issues	2	42, 44
41	Technical	TE10	Security issue due lack of log collection and forensic analysis	Organizations have no log collection to allow forensic analysis	2	1, 2
42	Technical	TE12	Technical issue due lack of user friendly big data tools	Big data tools are not user friendly	2	31, 33

3.4.8.3 Annex C – Design principles for digital transparency

Table 33 – Design principles for digital transparency in TOGAF template

P1	**Name**	**Separating privacy and non-privacy sensitive data at the source**
	Short Name	**Privacy**
	Statement	The essential requirement for transparency is determining the privacy level of data. Without knowing whether the data contains sensitive, personal information, it is risky to open it.
	Rationale	Open data must be balanced with the need to restrict the privacy and sensitivity of data. Private and sensitive data must be protected to prevent improper use and misinterpretation.
	Implications	There should be a process of determining whether the data can be opened without violating privacy. Government and developers should understand the impact of releasing data and find solutions if such data must be opened but is constrained due to its sensitive nature.
	Practical Example	Organizations collect daily a lot of data from users. Part of this data can be collected, stored, and used internally. However, sharing part of this data must comply with the privacy laws such as the General Data Protection Regulation (GDPR). A practical example is given by Chanson et al. (2019) using blockchain cases, where the proper level of transparency is achieved to identify essential aspects of transactions without compromising privacy.
P2	**Name**	**The openness of processes and actors**
	Short Name	**Openness**
	Statement	This principle enables the public to gain information about the operation, structures and decision-making processes of an organization.
	Rationale	If people are aware of how decisions are done, by whom and using which tools, they will be more trustful towards the outcomes of such decisions.
	Implications	In order to be transparent, a public organization must be opened in terms of the process, e.g., the procurement or audit flow, who is responsible for which activities, and which tools were used to make decisions. Any change in those aspects should be documented, and the change process itself must be opened.
	Practical Example	Some processes are unclear, and actors are unwilling to provide details about their actions. A practical example about the openness of processes and actors is the constitution of the United States which aims at reducing corruption and increasing the level of transparency to the public (J.C. Bertot et al. (2010).
P3	**Name**	**Feedback Mechanisms for improving digital transparency**
	Short Name	**Feedback Mechanisms**

	Statement	Feedback mechanisms are critical in understanding the data, which leads to achieving transparency.
	Rationale	Creation of transparency is an ongoing process, a cycle, which requires feedback, especially to improve the data, system and service quality.
	Implications	A transparency platform should provide an interface to allow communication between data users, data providers and policy-makers regarding the quality and use of the released data. Furthermore, data providers and policy-makers should spare some resources (time, dedicated employees, etc.) to interact with data users.
	Practical Example	Communication is based on a two-way process comprising listening and speaking. Giving voice to users is an important factor identified by Rawlins (2008) who recommended to ask for feedback from people to improve information quality, and consequently, transparency.
P4	**Name**	**Various levels of abstraction for data access**
	Short Name	**Data Abstraction**
	Statement	Data is accessible for users based on their needs.
	Rationale	Broader audience leads to different types of user needs and requires various levels of data access.
	Implications	A transparency platform should define different privileges for user access by understanding different uses of data for each group of users against levels of data sensitivity.
	Practical Example	Taking into consideration the needs and levels of users, not everyone should have a similar type of access to data. Due to this, Parnas and Siewiorek (1975) recommend reducing transparency to provide the best user experience. Avoiding exposing the algorithms, e.g., creating queries with search boxes using simple words like in Google Search, will help less knowledgeable users work with systems and data. We can also include practical examples following Privacy (P1) principle because depending on the user level in the hierarchy (managerial, tactical, operational, etc.), users should not have access to all data, avoiding GDPR issues.
P5	**Name**	**Avoiding any types of jargon or terms that the public does not understand**
	Short Name	**Comprehension**
	Statement	Data are presented as simply as possible.
	Rationale	This principle allows a broader audience to understand and interpret data correctly.
	Implications	Data should be checked if regular people can understand and interpret it so that they can use it.
	Practical Example	Jargon and lack of simple language can create barriers to users. As an example, O'Connor and Kelly (2017) recommend using "*bureaucratic language and lack of clarity on specifications, as well as a lack of staff professionalism*" because this can

		reduce transparency when small and medium-size enterprises try to access government funds and services.
P6	**Name**	**Checking and rating data quality**
	Short Name	**Data Quality Rating**
	Statement	Enable ways to provide user features to double-check data quality.
	Rationale	Data quality plays a vital role in the creation of transparency. The use of data depends on its quality.
	Implications	Information regarding data quality must be provided in the metadata. The expected effect of transparency, e.g., accountability, requires enriching data with photos or links to external data sources, e.g., Google maps and crowdsources.
	Practical Example	Disclosed data should have a certain level of accountability to avoid practical issues such as a fear of publishing inaccurate or wrong data leading to misuse or mistakes, e.g., the Australian government example in Hardy and Maurushat (2017), reducing the level of public benefits including transparency.
P7	**Name**	**Visualization of different views**
	Short Name	**Visualization**
	Statement	Different types of data require different types of visualization.
	Rationale	Providing different types of visualizations such as tables, graphs or maps, as well as the options expected by users, enables more usage and insights.
	Implications	The same data can be visualized in different ways based on user preferences or data needs.
	Practical Example	Providing different views on the same data is relevant when working in an interconnected operation. A practical example is given by Ricardo Matheus, Janssen, et al. (2018) using the IBM Center of Operations as an empirical initiative to demonstrate how different departments might use the same data in different ways. A car accident data would be relevant for various departments in a diversity of forms. Traffic managers would be interested in seeing how much traffic jam it is creating and how to reduce its impact. Police would be interested in contacting the closest car and managing the accident locally as a crime scene requiring a forensic officer. Ambulances would like to know what the fastest route to any hospital with the available surgical operating room is.
P8	**Name**	**Data access using different protocols**
	Short Name	**Data Access**
	Statement	Data is accessible based on user preference and expertise.
	Rationale	Providing a different way of access can reach a broader audience.
	Implications	Accessibility involves protocols through which users obtain data. The way data is made available must be sufficiently flexible to satisfy a broader audience and respective access methods. For example, to follow the linked data framework.

	Practical Example	A practical example of the relevance of accessing data using different protocols was made in Finland to monitor the growth of companies (Salonen et al., 2013). Facebook, Twitter and Google are public web portals. To collect data, data scientists can scrape the portals using bots that copy-paste data from the web pages, or access such pages using Application Programming Interfaces (APIs). Depending on the amount of data, the difference between scraping and APIs can be in the magnitude of hours or days. While some people can be satisfied to access Facebook, Twitter and Google web pages, developers would prefer the automated versions using APIs.
P9	**Name**	**Use of standardized formats**
	Short Name	**Standardized Formats**
	Statement	Data is available in different but standardized formats to allow comparison
	Rationale	Different user needs and preferences require different data format types, ranging from human- to machine-readable.
	Implications	The use of data depends on available formats. Data should be available in many formats.
	Practical Example	A defined data standard can shape a sector. Goëta and Davies (2016) give a practical example, where many cities use mobile applications that rely on the General Transit Feed Specification (GTFS) when dealing with traffic data, e.g., Google maps-related features and data. Other examples can be given of data related to Geographical Information Systems (GIS) such as shapefiles, open data standards such as Comma-Separated Value (CSV) or linked data using the Resource Description Framework (RDF). While CSV and RDF are machine-readable and can be easily used by developers, they also enable human reading.
P10	**Name**	**Persistency to ensure that data is not altered and the history can be traced**
	Short Name	**Data Persistency**
	Statement	Keeping the data with the same original characteristics, i.e. content, name, place etc.
	Rationale	The original data characteristics should be maintained to facilitate data comparisons.
	Implications	The implications include applying a consistent place of access, using the same data content and updating metadata.
	Practical Example	A practical example of simultaneously enabling persistency and transparency is made through the blockchain initiatives. For example, Paik, Xu, Bandara, Lee, and Lo (2019) show the traceability of blockchain-based system architectures.
P11	**Name**	**Data and system interoperability**
	Short Name	**Interoperability**
	Statement	Promoting data, application and technology interoperability.
	Rationale	In order to ensure the integration between building blocks and data, interoperability is required.

	Implications	In order to implement system and data standards for interoperability, a process to implement standards, updates and exceptions should also be provided.
	Practical Example	Transparency is a crucial element of Smart Cities, which have different sources of data and various departments using the same data. A functional Smart City architecture has a high level of interoperability. A practical example is given by Pardo, Nam, and Burke (2012) through the interoperability architecture created to share and integrate all systems and data within internal and external organizational boundaries.
P12	**Name**	**Include metadata for understandability of data**
	Short Name	**Metadata**
	Statement	High-quality metadata supports the *understandability* of data.
	Rationale	Provide insights, allow combining and check methodology. High-quality metadata is needed to assess data quality and understand the nature of data for the usage intention.
	Implications	Quality Metadata must be provided, including information about context, supporting multilingualism, and identifying data properties and quality.
	Practical Example	Metadata is a crucial element to understand and describe what the data contains. Practical examples are given by Praditya, Janssen, and Sulastri (2017) and (Praditya, Sulastri, Bharosa, & Janssen, 2016). They describe the importance of including metadata in the eXtensible Business Reporting Language (XBRL) for transparent financial reporting.
P13	**Name**	**Transparency-by-design (automatically opening data)**
	Short Name	**Transparency-by-design**
	Statement	Transparency requirements are satisfied by the very nature of the design, that the outcomes of the design process should meet these requirements.
	Rationale	The software and business processes should be designed to be open and to open up the public sector.
	Implications	Transparency requirements are considered when designing new systems, administrative processes and procedures. The systems should enable the collection of data and metadata from the source and ensure that such data and metadata can be opened for transparency. Also, the systems should facilitate the understanding and interpretation of data.
	Practical Example	A practical example of transparency-by-design is given by Saxena (2017), who describes the open data initiative of the Sri Lankan government. The author explains how transparency should influence and shape all steps of the data cycle, from data collection to data disclosure through open data portals.
P14	**Name**	**Opening of raw data**
	Short Name	**Opening of Raw Data**
	Statement	Transparency requires raw, low-granularity data.

	Rationale	Granularity refers to the level of detail embedded in data. If the data is provided on the aggregate level, the users will have limitations to use the data, including considerations of the privacy and sensitivity of data.
	Implications	For transparency, open data portals should provide several levels of data granularity.
	Practical Example	Disclosing data in raw formats can help people increase the level of transparency by themselves. A practical example is given by Iqbal, Wallach, Khoury, Schully, and Ioannidis (2016). The authors explain why it is essential to have raw data (data at the low granularity level) in the biomedical sector, allowing other researchers to shape their studies and come up with different conclusions.
P15	**Name**	**Assign Stewardship for digital transparency**
	Short Name	**Stewardship**
	Statement	There is a need for an actor who is responsible for maintaining the data and metadata quality. There is also a need to ensure the openness of the process that leads to transparency.
	Rationale	Stewardship refers to the actor role that ensures data and metadata quality. Usually, a database administrator is in charge of system governance to provide proper transparency level. This role should also know about privacy regulations.
	Implications	The transparency steward must be designated. This person must be knowledgeable, trained and experienced in dealing with data and metadata quality.
	Practical Example	An example of a steward influencing transparency is given by Dawes (2010). The author describes the importance of stewards in the governance of data in the USA Census Bureau and the New York Health Department to increase government openness and transparency when disclosing data to people.
P16	**Name**	**Supporting views with different level of details**
	Short Name	**Gradation of Detail**
	Statement	Data should be presented from the overview to the detailed level.
	Rationale	A wide range of users requires different views of data, from the abstract to the detailed level. This requirement is also influenced by various scenarios and needs of using the same data.
	Implications	The system must provide a range of features that enable the customization of different user needs.
	Practical Example	It is highly recommended that a portal provides a variety of features to increase transparency, for example, dashboards for the public and decision-makers by the IBM Center of Operations Rio (Ricardo Matheus, Janssen, et al. (2018). The public has direct and straightforward information about traffic conditions and how to avoid traffic jams, e.g., via mobile apps or public dashboards over streets with high levels of traffic jams. However, traffic managers, police or ambulance should have in-depth access to all data collected in real-time from the city sensors, enabling the best

		decisions possible. For instance, the same map with traffic condition can be shown with few details to the public, but with many details including several layers and filters to government decision-makers.

3.4.8.4 Annex D – Interview Protocol Form

Introduction

You are selected as respondent of this interview to contribute on the creation of transparency portals of the OpenGovIntelligence (OGI) project (www.opengovintelligence.eu). This research aims to synthesize the principles behind the design of transparency portals. We argue that to achieve certain level of transparency, principles should be considered in the design of open data portal.

Essentially, this document states that:

(1) all information will be held confidential;

(2) your participation is voluntary, and you may stop at any time if you feel uncomfortable; and

(3) we do not intend to inflict any harm.

Thank you for your agreeing to participate!

We have planned this interview to last about one hour due to the wide range of the needed information. During this time, we have several questions that we would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

General information

9. What was your pilot?

( ) Belgium ( ) England ( ) Ireland ( ) Other: _____

10. What was your role during OGI project?

( ) Technical ( ) Managerial ( ) Other: _____

A- Following Enterprise Architecture and Principles

11. Normally, I follow the enterprise architecture when creating applications

( ) Strongly disagree

( ) Disagree

( ) Neutral

( ) Agree

( ) Strongly agree

12. Normally, I follow principles when creating applications

( ) Strongly disagree

( ) Disagree

( ) Neutral

( ) Agree

( ) Strongly agree

B- Principle Questions

Table 1 – Description of Principle 1

P1	**Name**	**Separating private and non-private sensitive data at the source**
	Statement	The basic requirement for transparency is determining the privacy level of the data. Without knowing whether the data contains privacy (including non-private but sensitive) data or not, it is risky to opening data.
	Rationale	Open data must be balanced with the need to restrict privacy and sensitive data. Privacy and sensitive data must be protected to prevent improper use and misinterpretation.
	Implications	There should be a process to determine if the data can be opened without violating privacy issues. The government as well as the developers should understand the potential effects of releasing the data, as well as find solution if there is data that have to be opened but constrained of sensitivity of the data.

[1] Do you agree with this Pprinciple Name? If not, please write below your modified Name (not mandatory).

_____

[2] Do you agree with this Principle Statement? If not, please write below your modified Statement (not mandatory)

_____

[3] Do you agree with this Principle Rationale? If not, please write below your modified Rationale (not mandatory)

_____

[4] Do you agree with this Principle Implications/ If not, please write below your modified Implication (not mandatory)

_____

[5] Did you take into account Principle 1 during the development of your OGI application?

( ) Yes ( ) No

[6] What is your rating for Principle 1 in terms of importance?

( ) Low importance

( ) Slightly important

( ) Neutral

( ) Moderately important

( ) Extremely important

[6.1] Do you want to explain your argument? If yes, please explain below (not mandatory):

_____

[7] In practice, it is easy to implement Principle 1

( ) Strongly disagree

( ) Disagree

( ) Neutral

( ) Agree

( ) Strongly agree

[7.1] Do you want to explain your argument about easiness to implement this principle? If yes, please explain below (not mandatory):

_____

[8] What is your rating for Principle 1 in terms of Priority?

( ) Low priority

( ) Somewhat priority

( ) Neutral

( ) High priority

( ) Essential priority

[8.1] Please explain your argument:

_____

[9] What is the impact of Principle 1 on the Organization?

( ) No impact

( ) Minor impact

( ) Neutral

( ) Moderate impact

( ) Major impact

[9.1] Do you want to explain your argument about the impact of this principle on the organization? If yes, please explain below (not mandatory):

_____

4 Epilogue

Although transparency and its achievement with digital means are desirable, they are also challenging. This research contributes to achieving transparency with digital means, i.e. digital transparency. It shows that digital transparency is influenced by many more factors than present in the literature. It also points to the trade-off between transparency and usability.

This chapter concludes the dissertation. It is structured as follows. Section 4.1 answers the research questions, and section 4.2 summarises the scientific contribution. Sections 4.3 and 4.4 discuss the dilemma of principle-based vs rule-based approaches and the non-existence of full transparency. Finally, section 4.5 outlines the limitations of this research, and section 4.6 presents possible directions for future research.

4.1 Answering the Research Questions

4.1.1 RQ1 – What is Digital Transparency?

This research question is addressed in Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory" (see section 3.1).

In this paper, we uncovered a diversity of digital transparency concepts, and we provided our definition in the sections 1.1 and 3.1.3. Answering RQ1:

> Digital transparency is "*any initiative using digital means contributing to the insights of the public into the government*.".

Beyond the definition, we found out that despite its merits and the availability of relevant tools, full transparency is difficult to achieve since even when information is available, it might be challenging to interpret or may not arrive in time to be useful (Fung, 2013, p. 184). However, achieving full transparency is not desirable, as argued in the section 3.4.6.2. We also found out that the practical realisation of digital transparency is challenging. Surprisingly, despite that, scientific publications usually conceptualise digital transparency as an outcome but fail to explain what mechanisms can used to achieve that outcome.

The digital transparency conceptualisation process produced two main findings. First, we identified through the literature review and content analysis the evolution of OGD-driven digital transparency into four phases: the Ex-ante Phase before 2009, the Initiation Phase

between 2009 and 2011, the Hype Phase between 2011 and 2013, and the Realism Phase since 2014. Most scientific manuscripts were published in the public administration field (49 papers or 67%), whereas 24 papers or 33% were published in the information systems field.

Second, we identified the expected effects of digital transparency. The most expected effect is accountability (51 papers), followed by trust and credibility (35 papers), civic engagement and participation (30 papers), efficiency and cost reduction (28 papers), governance and political turnout (20 papers), anti-corruption efforts (17 papers), error and information asymmetry (11 papers), and privacy (6 papers). Typically, the papers mention more than one expected effect. Most of the literature is focused on the positive effects of digital transparency. The adverse effects are mentioned only in 16% of the papers (12 out of 73).

4.1.2 RQ2 – Which Factors Influence the Creation of Digital Transparency?

This research question is answered by Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory" (see section 3.1). In the paper, we have not established a framework to serve as a theoretical and practical model for implementing OGDPs. This prompted using an SLR to identify the factors that influence digital transparency. Such factors are listed in the section 3.1.3.

Answering RQ2, 42 factors were identified, including free access to data, open access to data, ontology, timely access to data, data accuracy, etc. (see Table 11). The factors were divided into four categories: Data Quality, System Quality, Organisational Characteristics, and OGD Individual Characteristics. The aggregation and categorisation of all uncovered factors comprise "The Window Theory" (see section 3.1.4 and Figure 21).

The name "The Window Theory" was chosen to highlight the expectation that OGD provides a window to what has happened or is happening within the government. The window's frame determines if it can be used. Sometimes, the same dataset, seen by different observers, can result in more or less transparency. This thesis cover includes two images of the same frame, one during the day and another at night. Without natural or artificial light, it is hard to provide transparency. The same happens with OGD.

The Window Theory contains a list of factors, but not all are relevant to the situation. The factors depend on the transparency's context and desired effect (see section 3.1.3.3). As a metaphor, we need different windows for different audiences and purposes.

4.1.3 RQ3 – What are the Types of Mechanisms that Results in Transparency?

This question is answered by Paper 4 – Design Principles for Creating Digital Transparency in Government" (see section 3.4). Three case studies were analysed to understand how digital transparency is created using OGD. The factors identified in RQ2 were used to investigate the case studies and understand the mechanism for creating transparency.

The case studies uncovered what users consider as transparent and which mechanisms worked for them to create such transparency. With the level of transparency varying per case study, they demonstrated the importance of considering the context and users. The context determines what is considered transparent and which factors influence transparency.

To create design principles, we conducted an SLR to uncover barriers to digital transparency. 43 barriers were uncovered. They were divided into eight dimensions: Data Quality, Economic, Ethical, Human, Organisational, Political and Legal, Technical, and Usage (see Table 25). The list of barriers is provided in the section 3.4.3.

Starting from these barriers, we derived design principles to overcome them. The list of design principles is presented in the section 3.4.4 and Table 27 shows the relationship between them and the barriers. In total, 17 design principles were defined and used as sources to conduct the case study (see Table 26). Besides identifying the relationship between the barriers and design principles, we created a Digital Transparency Cycle to guide the application of design principles to increase digital transparency (see Figure 26).

The case study was conducted in three European countries with different expectations on the effects of transparency. The first in Belgium within the Flemish Environment Agency and its Environmental Digital Transparency web dashboard. The second is in England, within Trafford's Innovation and Intelligence Lab and its Worklessness (unemployment) Digital Transparency web application. And the third in Ireland with the Marine Institute and its Irish National Tide Gauge Network web portal.

We tested the use of the design principles within all programmers across the three case studies. Table 30 shows the results. For designers, the top five design principles used when creating digital transparency applications were Privacy, Metadata, Data Access, Standardized Formats, and Interoperability. All designers used the Privacy and Metadata principles. Some principles were used occasionally, e.g., Stewardship at 33%, Comprehension at 44% or Transparency-by-Design at 56%. Some were not used at all. The interviews revealed that the principles primarily concerned organisational changes, whereas the case study projects focused on app development. This disparity did not make them less relevant. On the contrary, the interviewees suggested that they must adhere to them to deliver digital transparency.

These results enable the answer to RQ3 (see section 3.4.5):

"*There is a diversity of the types of mechanisms resulting in digital transparency*"

However, design principles should be interpreted and used depending on the context, particularly the organisational context. Creating digital transparency is not limited to technical system development issues. It also includes the issues of organisational changes and creating the right organisational conditions for digital transparency.

This implies that creating digital transparency is a significant challenge to governments. Merely opening data does not automatically produce digital transparency and might only result in information overload for those wanting to examine such data. In order to create digital transparency, a Transparency Window, proposed by The Window Theory, should be designed to enable looking at different aspects of the organisation from different perspectives.

4.1.4 RQ4 – What are the Factors Influencing the Usefulness of Transparency?

This question is answered by Paper 2 – Digital Transparency and the Usefulness for Open Government" (see section 3.2). The paper describes the case studies that provide insights into the mechanisms and factors influencing transparency, which differ from case to case. A parsimonious, testable model was created to generalise the factors found in the case studies.

A survey was conducted among the transparency application and portal end-users. The survey helped answer RQ4, providing a list of factors that influence the usefulness of transparency (see section 3.2.4). Perceived ease of use, usefulness and efficiency were found to be the main

determinants of transparency while creating transparency and efficiency simultaneously was found challenging. Figure 22 and Figure 23 summarise the SEM model and the list of factors.

In the same survey, we found that the digital transparency applications that result in higher usage efficiency often only provide predefined views. They are efficient to use but only provide insights from this view. In contrast, transparency requires the inclusion of many different views. More transparency can be created by making raw data available and letting people analyse it from their perspectives. Although more time-consuming and requires an understanding of how the data is collected, it creates higher levels of transparency.

Our findings suggest that transparency is highly contextual (Janowski, 2015). This explains why a large number of factors have a limited influence. As the context changes, the influence of the factors changes as well. In the Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory", Matheus and Janssen (2019) reconfirmed this finding, as they developed a compressive model consisting of 41 determinants and eight effects and suggested instantiating these for a situation at hand. Searching for a unified model might be challenging, given the variety of options.

The survey models show that transparency and efficiency cannot occur simultaneously; they require a trade-off. Digital transparency applications result in higher efficiency in use but only provide insight from one or a few pre-defined views, bringing advantages and limits to the level of transparency. Digital transparency is created from the raw data, which is inefficient to use but enables users to create their views and find original insights.

4.1.5 RQ5 – What are the risks and benefits when creating Transparency Dashboards?

This research question is answered by Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" (see section 3.3). According to this paper, dashboards can improve transparency and accountability, but realising these benefits is cumbersome and prone to various risks and challenges. Challenges include insufficient data quality, lack of data understanding, wrong interpretation, poor analysis, confusion about the outcomes, and imposition of pre-defined views. These challenges can easily result in misconceptions, wrong decision-making, and a blurred picture, resulting in less transparency, accountability, and even less trust in government.

The answer to RQ5 is in the section 3.3.3.2. Table 23 – Description of the main risks and challenges provides descriptions of 12 risks and challenges identified when creating digital transparency dashboards, e.g. data leading to bad results or wrong decision-making, limited knowledge, and pre-defined views.

4.1.6 RQ6 – What are the design principles for transparency dashboards?

The answer to this question is in Paper 3 – Data Science Empowering the Public: Data-driven Dashboards for Transparent and Accountable Decision-making in Smart Cities" (see section 3.3). The paper identified that design principles guide the design of transparency dashboards and are helpful as guides to solving ill-structured or 'complex' problems, such as digital transparency. Design principles, complemented by citizen engagement, data interpretation, governance, and institutional arrangements, can also guide the design of more effective public sector dashboards.

The answer to RQ6 is in the section 3.3.3. Table 21 – Principles and literature background provides a list of design principles influencing digital transparency dashboards. We uncovered eight such design principles: correct and precise data, customised view, clear presentation, offering decision-making support, interaction support, providing an overview and details, focusing on creating added value, ensuring real-time updates of data, and ensuring that institutions support accountability.

These principles are connected to RQ5 and the 12 risks identified. When creating digital transparency dashboards, the risks can be addressed by following these design principles.

4.1.7 RQ7 – What are the Design Principles for Creating Transparency?

The answer to this question is in Paper 4 – Design Principles for Creating Digital Transparency in Government" (see section 3.4). The paper recognises that digital transparency is a multi-dimensional concept that is hard to implement. Based on the factors influencing the level of transparency, architecture principles are developed to guide developers in creating apps for transparency, depending on the characteristics and desired effects of transparency.

To answer RQ7, section 3.4.4, we derived design principles from a list of barriers to creating digital transparency. Table 26 – Design principles for digital transparency presents all 16 design principles: privacy, openness, feedback mechanisms, data abstraction, comprehension,

data quality rating, visualisation, data access, standardised formats, data persistency, interoperability, metadata, transparency-by-design, opening of raw data, stewardship, and gradation of detail.

Table 27 – Relationships between barriers and design principles presents a relationship between barriers and design principles when creating digital transparency. The table can help designers find design principles to overcome given barriers. Some barriers are related to one or more design principles, but few to all or none.

Finally, Figure 26 – Data-driven transparency cycle with design principles adapted from Matheus and Janssen (2018, p. 36) presents a 6-step design cycle combined with all design principles. The principles are in one or more steps, helping designers double-check if they follow the digital transparency design principles correctly.

4.2 Scientific Contribution

Transparency is researched extensively in public administration (Roberts, 2002). However, translating the results to digital transparency is challenging. In the digital world, we can open a large volume of data, but this volume might also complicate the creation of transparency.

This research produced four identified scientific contributions.

The first contribution is the conceptualisation of digital transparency (Ricardo Matheus & Janssen, 2019). This is important since no major publication explains what digital transparency is. An SLR identified various views and concepts and enabled us to identify what digital transparency is, resulting in Paper 1, presented in the section 3.1.

The second contribution is the collection of factors that impede or enable transparency (Ricardo Matheus, Faber, Ismagilova, & Janssen, 2023). Some scientific papers generically use digital transparency without identifying the factors that influence digital transparency positively or negatively. The abovementioned SLR built the collection of factors impeding or enabling digital transparency. This contribution is included in Paper 1 and presented in the section 3.1.

This contribution also comprises a new theory for digital transparency, the Window Theory (Ricardo Matheus & Janssen, 2019). The Window Theory posits that transparency depends on the *human perception of making sense of the presented data*. It means the same dataset presented on the same website and visited by different persons can result in different levels of

perceived transparency. The same happens when looking through a window. Different persons are triggered by different views (e.g. one looks at the trees and another looks at the buildings), looking through the window at different times and angles. This can result in looking for other elements in the window and different perceptions about the external world. This contribution is included in Paper 2 and presented in the section 3.2.

The third contribution is identifying various ways through which transparency can be helpful. This includes types of data disclosure, data use, as well as views on transparency. This contribution is essential since transparency is often viewed as a condition *sine qua non* for democracy. However, the strategies selected by policymakers leading to transparency, their usage, and the expected views of transparency are not evident in the scientific literature.

For example, a policymaker could decide to have an Internet-based portal (website) as a channel for data disclosure, with an open data format for users to download. Another policymaker would prefer to create a simple website where people can access data in static websites without downloading or following any open data principles but by providing easy-to-use static dashboards with maps, graphs, and tables. Both options provide transparency to citizens. Nonetheless, to identify the "most transparent" option, we should consider users' context and needs (i.e. types of mechanisms) to achieve transparency. This contribution is included in Paper 2, presented in the section 3.2.

The fourth contribution is the creation of design principles that could help OGD website and app designers deliver transparency (Ricardo Matheus et al., 2021). Since there are no guidelines or principles to help create transparent OGD-based portals or applications, this dissertation brings a collection of design principles to help designers create digital transparency. Selecting one or more design principles will influence how transparent portals or apps look.

The literature suggests that full transparency is both hard to achieve and unnecessary (Fung et al., 2007; Roberts, 2002). Following all principles is not required or even possible. Transparency-by-design emerged as a foundation for governments aiming to include transparency in each design step (Janssen, Matheus, et al., 2017).

An example of the balancing principles is that when creating a transparency portal, following the ease-to-use principle, regular citizens might automatically reduce the number of options for the type of formats and access to data (Ricardo Matheus & Janssen, 2015a; Ricardo Matheus,

Janssen, et al., 2018). While OGD programmers and experienced users would prefer an Application Programming Interface (API) to access data or JavaScript Object Notation (JSON) file format to automate machine processing easily, ordinary persons might prefer static dashboards with plain features such as heat maps, coloured graphs and tables for comparing the relative pollution of areas or school locations, to protect family health and plan transport of children to schools. This contribution is included in Paper 3 and presented in the section 3.3.

4.3 Principle- or Rule-based Approaches?

Principles are used by some but not all countries. The majority of the countries have their principles connected with supranational projects such as the Open Government Partnership (OGP), the Organisation for Economic Co-operation and Development (OECD), or the European Union (EU). However, the lack of design principles can also result from differences between a rule-based and principle-based approach (Burgemeestre, Hulstijn, & Tan, 2009). Countries with rule-based systems prefer to prescribe what type of data should be disclosed (transparent) to avoid legal issues, like in Brazil. The focus is on 'what' should be done. In contrast, in countries like the Netherlands, the focus is on the desired effect or results to be achieved instead of explaining in detail what should be done.

Related to the rule-based vs principle-driven orientation is the "*open data washing*". This can be found in the work of Heimstädt et. al. (2017), which classified "how" some governments have been publishing data due to the legislation enforcing active disclosures. However, by following the rules, publishing data logs did not produce transparency for citizens.

As a rule-based country, Brazil has several transparency legislations, following several international and supra-organizational recommendations (Ricardo Matheus & Janssen, 2016a). Part of these were created using a top-down approach, without the proper participation of civil society and public organizations. As an example, Brazil has "Complementary Legislation" (LC) which "refines" the federal constitution, meaning that it is hard to change by the legislature and hard to avoid by the executive because the judiciary tends to hardly respect the federal constitution. The "Tax Liability Law" LC or "*Lei de Responsabilidade Fiscal*" (LRF) in Portuguese (http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp101.htm) requires governments to disclose financial data but offers no guides on how to ensure transparency of such data. This results in lots of published data but not increasing transparency.

The LRF was released in 2000 before the Internet took up. When released, it was not predicted how the world would become. It means that for every technological improvement or any type of legal "upgrade", a hard change is required via the legislature, which takes a lot of time and effort. This might create problems later since technology may emerge after the law's released.

Given the technological developments and uncertainties, a principle-based approach might be more sustainable, but many governments have structures favouring a rule-based approach. That is the reason behind a list of design principles to create transparency portals presented in this dissertation. The list might be helpful for designers and policymakers in charge of transparency, changing the mindset from following the rules to helping achieve digital transparency.

4.4 Why Does Full Transparency Not Exist? Or Should Not be Desired?

Some authors suggest that full transparency is impossible in practice (Fung et al., 2007). Some also claim that full transparency would not be desired by the principal (who publishes data) and agents (who consume data), taking into account the Agency Theory (Eisenhardt, 1989). Full transparency suggests overcoming the information asymmetry, a central problem in the Agency Theory. According to Cukierman (2009), full transparency "*fully and immediately transmits to the public all its private information about the economy, about its objectives and about the internal procedures that underlie the bank's policy decisions*".

Information asymmetry and lack of accountability are connected to human access to data and how people perceive the world. Matheus and Janssen (2015a) suggest that transparency is a multidimensional concept that might be connected with the subjectivity of users. Factors such as the democratic structure and culture might influence how countries have implemented OGD policies and portals. For example, according to González-Zapata and Heeks (2016), a case study in Chile showed that previous decisions and experiences played a big role in how OGD portals and agendas were implemented.

Full transparency can also result in non-desired effects, such as the creation of a surveillance state. When the government collects all kinds of data, including for transport, energy use, social security, etc. making these transparent can result in bias or discrimination. Opening too much data might also result in the reidentification of citizens and violating their privacy.

Aiming to protect people's privacy, a collection of principles should ensure the protection of personal data. In our social contract, e.g., national constitutions, governments have our

permission to collect our data to provide us public services. However, governments have been collecting more and more data about citizens (Janssen & Kuk, 2015). Part of this data complies with legislation and a consensus about their collection (e.g., birth and death). However, some governments collect data indiscriminately, without public consent. Combining huge amounts of data can bring citizens benefits or risks. For example, data can be used to reduce electricity bills or public transportation journeys and commutations, saving people time and money (Matheus, Janssen, et al., 2018). However, the data use can also have non-desired effects, e.g. biased and/or discriminatory algorithms (Chander, 2016) or algorithms that can modulate democracies (Isaak & Hanna, 2018), reducing our privacy and manipulating our choices.

The third non-desired effect is the lack of knowledge about how things work. The paradox is that the data opened to create transparency can be used by the algorithms that are opaque. Part of the algorithms cannot be revealed or explained as they are closed (Nograšek & Vintar, 2014). For example, Paracetamol, used to reduce fever, was first made in 1877 by Morse. Von Mering discovered Paracetamol's positive effects on human health and started to recommend it to patients in 1893. However, only 120 years after its discovery, in 1997, it was discovered how Paracetamol works in the human body (Sharma & Mehta, 2013). This example shows us that relying on closed artefacts (medicines, algorithms, etc.) could bring us benefits. We use such artefacts without any knowledge of how they work in practice.

Social media and big IT companies such as Facebook, Twitter, Google, and Amazon currently use a business model that does not reveal how algorithms work. Rader et al. (2018, pp. 1-2) recommend a set of three actions and mechanisms to increase the transparency of social media algorithms: the more experienced the users, the more they know the rules of the algorithms; algorithms should be audited by external and independent experts; and written explanations should be provided to users on how the algorithms they use work in practice.

Rader et al. (2018, p. 2) explain that the Facebook feed is based on an algorithm that decides what content to show. It means that posts are selected without users' knowledge of how and why they were selected. This might influence people's beliefs and judgment of the social reality. However, the Facebook algorithms are protected by commercial property rights and, as such, the company is not required to disclose them.

The Facebook algorithmic transparency issue and the mass surveillance by governments show how complex and sensitive is the discussion of transparency. On the one hand, people's rights

must be protected and enforced. On the other hand, commercial property rights must also be protected. For these reasons, full transparency might not be possible or even desired in practice. Similar to Rader et al. (2018), Kizilcec (2016) discussed algorithms transparency. The author identified a relationship between how much information should be provided (transparency) to achieve awareness and trust in algorithms. The authors found out that "*expectation violation was a critical moderator of the effect of transparency on trust. If users' expectations were met, interface transparency did not affect trust, as individuals were less likely to examine information thoroughly and more likely to rely on general impressions or their own mood*" (Kizilcec, 2016, p. 2392), and called for a balanced scenario, "*not too little and not too much*" (Kizilcec, 2016, p. 2390). This shows that digital transparency remains a challenging field.

4.5 Research Limitations

This thesis developed a design theory for digital transparency. The theory consists of four elements – transparency, functionality, efficiency, and usefulness (see Figure 22), and targets policymakers and designers. The theory is founded in literature, interviews, and surveys. All of them have several limitations that might impact the theory.

1. **Literature Review**

We reviewed the literature to identify the factors affecting digital transparency and to understand its effect. The review was focused on high-impact papers published until 2016. We might have missed some factors in the paper considered out of the scope, published after 2016, or published in journals excluded from the review.

More insights into factors were created during our research. These factors, such as privacy, metadata, data access, standardized formats, interoperability, visualization, data persistency, openness, data quality, gradation of details, and stewardship must be interpreted. The literature review revealed different views on those factors. We recommend more research into them.

Ignoring the grey literature sought to attribute greater reliability to research results, but publications such as theses and dissertations, which often rely on theoretical insights relevant to contemporary research were also disregarded. Thus, it is possible that such references can elucidate new perspectives within the theme.

The selected keywords may also have led to some deviations or absence of relevant perspectives for analysis. In the case of Paper 1 – A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory", few keywords were considered, and some (e.g., governance), which may correlate with the topic, were not used. On the one hand, the generalization attributed to the literature search implies a broad spectrum of the resulting works. On the other hand, some concepts or "peripheral" (but directly related) topics were excluded. Future analyses on the coherence of such choices could be conducted.

Another limitation relates to the initial selection of the sample references. An analysis of titles, abstracts, and keywords was carried out, but other reasons were also considered for elimination, e.g. lack of access. In this sense, the elimination of such references may imply intrinsic and unavoidable limitations to the method.

2. **Case Study**

The case studies were used to demonstrate and test the design principles described in the section 3.4.5. As the number of cases was limited, we used a convenient sample of three case studies within the context of the H2020 project that the PhD candidate participated in. In the meantime, more cases have emerged which could have enriched the study. It is not known whether the new cases that emerged during this period could imply new insights and additional knowledge to the research, so this point can be understood as a future research path.

No cross-analysis with other case studies was conducted. Such analyses could elucidate the existence or perception of similar, analogous, or divergent characteristics between the cases. This would have broadened the research spectrum and helped in future understanding of the reproducibility and repeatability of concepts, practices, and structures. Considering this point, it is recommended for further research.

3. **Semi-Structured Interviews**

In the same way as the previously described case studies, the semi-structured interviews used SLRs to recall previous knowledge in the construction of the interviews. In this sense, a method was selected that may have the same limitations previously exposed for SLR conducted in the literature review, which is inherent to the necessary choice of a procedure to obtain the results.

The possibility of more in-depth discussions on some of the topics of the semi-structured interviews led each respondent to address the subject in different ways, which was partially

avoided through specific guidance so that they did not go too far beyond the desired topic. Even so, considering the intrinsic subjectivity of the step, new perspectives of analysis could arise from different perspectives on the insights, and the complexity of some questions initially proposed may imply unconsidered bias and, in this sense, new perspectives for the future.

4. Survey Data Collection

The survey focused on a specific public, targeting a group of experts on the subject matter. Thus, the perspectives of those with lower levels of technical knowledge on the subject matter were not considered, nor were the understandings of other groups sought. A convenient sample in this survey was considered as representative.

Another point is the number of surveyed respondents. As in any method focusing on specific groups, restricted selection implies a reduced number of respondents, which limits the method's coverage. Data collection in the present study was also restricted to Europe. Considering that the topic can be seen through different priorities, territories and stakeholders, this can be treated as a limitation to this research, opening up the recommendation for future research that involves data collection in other territories.

Finally, a limitation of the present study is the absence of data collection regarding the demographic and socioeconomic circumstances of the respondents. The research opens the possibility of seeking answers in some locations or social or intellectual realities. However, the selection of respondents based on the technical attributes and prior experience related to the subject matter shows that the focus of the present work was rational and guided.

5. SEM

The first limitation related to the SEM method is the absence of cross-case analysis. More respondents would have allowed us to use some techniques to conduct this cross-analysis. Moreover, a minimum number of respondents could influence statistical results, and bigger surveys might result in better results. Cross-case analysis was not the purpose of the present work, so it remains a future research perspective.

The amount of latent variables is considerably smaller than the total. This drop does not necessarily mean that such variables are not relevant, but rather that they were not significant for the present model, which is mentioned in 3.2.4, when considering the existence of previous

knowledge of the respondents. Therefore, they should not be disregarded, and a suggestion is made for specific studies involving these eliminated variables in future research.

Likewise, the elimination of any hypotheses from the research (section 3.2.3) does not mean that they are not correlated in any context but that the data from the present model do not support conclusions in this regard. Thus, it is important to provide evidence that the model was limited to the data collection method, with the above limitations present.

Even though the model suggested that transparency and efficiency of use could not occur in parallel, it is again emphasized that such a conclusion follows from the model in question. Considering recent research trends and the evolution of the theme, it is desirable to expect that technologies and new structures will bring these topics closer. It must be emphasized that the purpose of the present work was not to discuss these aspects but rather to show that, in the current scenario, the correlation was not identified.

4.6 Recommendations for Further Research

This final section provides some recommendations for future research.

First, we recommend a better understanding of the situational and stakeholder views influencing digital transparency. Our research shows that digital transparency is dependent on the context and the stakeholder view. There is a large number of stakeholders, having different needs for different types of data, resulting in context-dependent views on transparency. The differences in context and stakeholder groups can be analysed in further research.

The Window Theory consists of a large number of context-dependent factors. Some factors have more effect in certain situations than others. We recommend understanding the relationship between the factors in the Window Theory and the expected effects of digital transparency on accountability, anti-corruption, innovation, decision-making, and public policy improvement. For this, we recommend survey research to generalize the findings.

Although we used the design theory in practice, the theory can be further tested and refined using case studies from around the globe. It has not yet been tested if the Window Theory has different results in other contexts than Europe. Culture, institutional systems, and societal differences may all play a role and influence transparency. Besides that, one limitation of this

dissertation was a survey using convenient samples in 6 cases in Europe. We suggest using a random sampling approach with a higher number of respondents.

In September 2021, we conducted a workshop with IT specialists in The Netherlands called "Transparency by Design: Creating Transparency in a black-boxed world", using the case of the EU plan for reducing CO2 pollution from aeroplanes and airports. The results showed that there is a tension between what technical people desire and what managerial people prefer. The first group was focused on creating more transparency and high-quality data, whereas the second focused on sensitivity and possible political consequences of transparency.

Another line of study is AI and digital transparency. AI can help create transparency; it can help analyse data and detect patterns. In contrast, generative AI makes it easy to create fake news and disinformation, which might look like creating transparency. This calls for insight into the correctness of data used to create transparency. What is correct and what is fake might not be easy to distinguish. The paradox is that the rise of AI can be used to create digital transparency but it can also reduce transparency. This suggests another research avenue.

References

Abelson, J., Gauvin, F.-P., MacKinnon, M. P., & Watling, J. (2004). Transparency, trust and citizen engagement. *Ottawa: Canadian Policy Research Networks Inc*.

Abramo, G., D'Angelo, C. A., Di Costa, F., & Solazzi, M. (2011). The role of information asymmetry in the market for university–industry research collaboration. *The Journal of Technology Transfer, 36*(1), 84-100.

Abu-Shanab, E. A. (2015). Reengineering the open government concept: An empirical support for a proposed model. *Government Information Quarterly, 32*(4), 453-463. doi:10.1016/j.giq.2015.07.002

Ackerman, J. M., & Sandoval-Ballesteros, I. E. (2006). The global explosion of freedom of information laws. *Administrative Law Review*, 85-130.

Acton, T., Halonen, R., Conboy, K., & Golden, W. (2009). DeLone & McLean success model as a descriptive tool in evaluating the use of a virtual learning environment.

Adams, M. B. (1994). Agency theory and the internal audit. *Managerial auditing journal*.

Al-Qirim, N., Tarhini, A., & Rouibah, K. (2017). *Determinants of Big Data Adoption and Success.* Paper presented at the Proceedings of the International Conference on Algorithms, Computing and Systems.

Alawadhi, S., Aldama-Nalda, A., Chourabi, H., Gil-Garcia, J. R., Leung, S., Mellouli, S., . . . Walker, S. (2012). *Building understanding of smart city initiatives.* Paper presented at the International conference on electronic government.

Alexander, V. D., Thomas, H., Cronin, A., Fielding, J., & Moran-Ellis, J. (2008). Mixed methods. *Researching social life, 3*, 125-144.

Alexopoulos, C., Loukis, E., & Charalabidis, Y. (2014). A platform for closing the open data feedback loop based on Web2. 0 functionality. *JeDEM-eJournal of eDemocracy and Open Government, 6*(1), 62-68.

Alexopoulos, C., Saxena, S., Janssen, M., Rizun, N., Lnenicka, M., & Matheus, R. (2023). Why do Open Government Data initiatives fail in developing countries? A root cause analysis of the most prevalent barriers and problems. *The Electronic Journal of Information Systems in Developing Countries*, e12297.

Alexopoulos, C., Saxena, S., Rizun, N., Matheus, R., & Janssen, M. (2024). Are creative users more apt in reusing and adopting Open Government Data (OGD)? Gender differences. *Thinking Skills and Creativity*, 101478.

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons, 60*(3), 285-292.

Amagoh, F. (2009). Information asymmetry and the contracting out process. *The Innovation Journal: The Public Sector Innovation Journal, 14*(2), 1-14.

Amugongo, L. M., Nggada, S. N., & Sieck, J. (2016). *Leveraging on open data to solve city challenges: A case study of Windhoek municipality.* Paper presented at the Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on.

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International journal of human-computer interaction, 16*(2), 185-210.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1-11. doi:doi:10.1111/1748-8583.12090

Angst, C., Agarwal, R., Gao, G. G., Khuntia, J., & McCullough, J. S. (2014). Information technology and voluntary quality disclosure by hospitals. *Decision Support Systems, 57*, 367-375.

Ansari, B., Barati, M., & Martin, E. G. (2022). Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research. Government Information Quarterly, 39(1), 101657.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review, 114*, 416-436.

Asante-Korang, A., & Jacobs, J. P. (2016). Big Data and paediatric cardiovascular disease in the era of transparency in healthcare. *Cardiology in the Young, 26*(8), 1597-1602.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), 399-418.

Attila, G. (2012). Agency problems in public sector. *Annals of the University of Oradea, 21*(1).

Auger, G. A. (2014). Trust me, trust me not: An experimental analysis of the effect of transparency on organizations. *Journal of Public Relations Research, 26*(4), 325-343.

Ballou, B., Heitger, D. L., & Donnell, L. (2010). Creating effective dashboards: how companies can improve executive decision making and board oversight. *Strategic Finance, 91*(9), 27-33.

Bannister, F., & Connolly, R. (2011a). The Trouble with Transparency: A Critical Review of Openness in e-Government. *Policy & Internet, 3*(1), 1-30.

Bannister, F., & Connolly, R. (2011b). Trust and transformational government: A proposed framework for research. *Government Information Quarterly, 28*(2), 137-147.

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity, 19*(1, 2), 129-152.

Barzelay, M., & Thompson, F. (2010). Back to the future: Making public administration a design science. *Public Administration Review, 70*, S295-S297.

Baskerville, R. (2008). What design science is not. In: Taylor & Francis.

Bauhr, M., & Grimes, M. (2014). Indignation or resignation: the implications of transparency for societal accountability. *Governance, 27*(2), 291-320.

Beck, R., Weber, S., & Gregory, R. W. J. I. S. F. (2013). Theory-generating design science research. *15*(4), 637-651.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45-59.

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity, 19*(1, 2), 5-16.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). 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.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). 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*, 264–271.

Bertot, J. C., McDermott, P., & Smith, T. (2012). Measurement of Open Government: Metrics and Process. In *Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS-45)* (pp. 2491-2499). Maui, HI, USA: IEEE.

Bessa-Vilela, N., Caramelo-Gomes, J., & Morais, P. (2017). Government transparency: reality or mirage? *Cybernetics and Systems*, 307-309.

Best, N. J., Ribeiro, M. M., Matheus, R., & Vaz, J. C. (2010). Internet e a participação cidadã nas experiências de orçamento participativo digital no Brasil. *Cadernos PPG-AU/FAUFBA, 9*(1).

Bharosa, N., van Wijk, R., Janssen, M., de Winne, N., & Hulstijn, J. (2011). *Managing the transformation to standard business reporting: principles and lessons learned from the Netherlands.* Paper presented at the Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times.

Birchall, C. (2011). Introduction to 'Secrecy and Transparency'The Politics of Opacity and Openness. *Theory, Culture & Society, 28*(7-8), 7-25.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205-227.

Blomgren, M. (2007). The drive for transparency: Organizational field transformations in Swedish healthcare. *Public Administration, 85*(1), 67-82.

Bloom, G., Standing, H., & Lloyd, R. (2008). Markets, information asymmetry and health care: towards new social contracts. *Social science & medicine, 66*(10), 2076-2087.

Bonsón, E., Torres, L., Royo, S., & Flores, F. (2012). Local e-government 2.0: Social media and corporate transparency in municipalities. *Government Information Quarterly, 29*(2), 123-132. doi:10.1016/j.giq.2011.10.001

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review, 82*(5), 977-1008.

Bridges, F., Appel, L., & Grossklags, J. (2012). Young adults' online participation behaviors: An exploratory study of web 2.0 use for political engagement. *Information Polity, 17*(2), 163-176.

Burgemeestre, B., Hulstijn, J., & Tan, Y.-H. (2009). Rule-based versus principle-based regulatory compliance. In *Legal Knowledge and Information Systems* (pp. 37-46): IOS Press.

Cahlikova, T., & Mabillard, V. (2019). Open Data and Transparency: Opportunities and Challenges in the Swiss Context. *Public Performance & Management Review*, 1-25.

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*: Sage.

Castiglia, B., & Turi, D. (2011). The impact of voluntary accountability on the design of higher education assessment. *Academy of Educational Leadership Journal, 15*(3), 119.

Chakraborty, S., Bhatt, V., & Chakravorty, T. (2020). Big-data, iot wearable and mhealth cloud platform integration triads-a logical way to patient-health monitoring. *International Journal of Engineering Advanced Technology, 9*(3), 388-394.

Chander, A. (2016). The racist algorithm. *Mich. L. Rev., 115*, 1023.

Chanson, M., Bogner, A., Bilgeri, D., Fleisch, E., & Wortmann, F. (2019). Blockchain for the IoT: privacy-preserving protection of sensor data. *Journal of the Association for Information Systems, 20*(9), 10.

Chen, Y.-C. (2012). A comparative study of e-government XBRL implementations: The potential of improving information transparency and efficiency. *Government Information Quarterly, 29*(4), 553-563.

Chen, Z.-J., Vogel, D., & Wang, Z.-H. (2016). How to satisfy citizens? Using mobile government to reengineer fair government processes. *Decision Support Systems, 82*, 47-57.

Chilisa, B., & Kawulich, B. (2012). Selecting a research approach: paradigm, methodology and methods. *Doing Social Research, A Global Context. London: McGraw Hill*.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research, 295*(2), 295-336.

Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: lessons learned from genomics. *Frontiers in human neuroscience, 8*, 239.

Chun, S., & Cho, J.-S. (2012). E-participation and transparent policy decision making. *Information Polity, 17*(2), 129-145.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77-90. doi:doi:10.1111/isj.12088

Coglianese, C. (2009). The transparency president? The Obama administration and open government. *Governance, 22*(4), 529-544.

Cole, F. L. (1988). Content analysis: process and application. *Clinical nurse specialist, 2*(1), 53-57.

Collins, A. (1992). Toward a design science of education. In *New directions in educational technology* (pp. 15-22): Springer.

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly, 31, Supplement 1*, S10-S17. doi:http://dx.doi.org/10.1016/j.giq.2014.01.003

Cook, M. A. (1996). *Building Enterprise Information Architectures. Reengineering Information Systems*. New Jersey: Prentice Hall.

Corradini, F., Polini, A., Polzonetti, A., & Re, B. (2010). Business processes verification for e-government service delivery. *Information Systems Management, 27*(4), 293-308.

Couldry, N., & Turow, J. (2014). Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy. *International Journal of Communication, 8*, 1710-1726.

Craig, M. J., & Kopits, M. G. (1998). *Transparency in government operations*: International monetary fund.

Crantschaninov, T. I., Matheus, R., & Silva, V. F. d. (2011). Governança, acesso à informação e conselhos de políticas públicas. *Revista Gestão & Políticas Públicas, 1*(1).

Cretu, V., & Manolea, B. (2013). *The Influence of the Open Government Partnership (OGP) on the Open Data discussions, EPSI*. Retrieved from

Cross, J. P. (2014). The seen and the unseen in legislative politics: Explaining censorship in the Council of Ministers of the European Union. *Journal of European Public Policy, 21*(2), 268-285.

Cuadrado-Ballesteros, B. (2014). The impact of functional decentralization and externalization on local government transparency. *Government Information Quarterly, 31*(2), 265-277.

Cukierman, A. (2009). The limits of transparency. *Economic Notes, 38*(1-2), 1-37.

Cumbley, R., & Church, P. (2013). Is "Big Data" creepy? *Computer Law & Security Review, 29*(5), 601-609. doi:https://doi.org/10.1016/j.clsr.2013.07.007

Curtin, D., & Meijer, A. J. (2006). Does transparency strengthen legitimacy? *Information Polity, 11*(2), 109-122.

Dash, N. K. (2005). Module: Selection of the research paradigm and methodology. *Retrieved August, 9*, 2009.

DataPortals.Org. (2024). List of Open Data Portals. Retrieved from https://dataportals.org/

David-Barrett, E., & Okamura, K. (2016). Norm diffusion and reputation: the rise of the extractive industries transparency initiative. *Governance, 29*(2), 227-246.

Davies, T. (2010). Open data, democracy and public sector reform. *A look at open government data use from data. gov. uk*.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319-340.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly, 27*(4), 377-383.

De Bruijn, H. (2007). *Managing performance in the public sector*: Routledge.

de Fine Licht, J. (2014). Policy area as a potential moderator of transparency effects: An experiment. *Public Administration Review, 74*(3), 361-371.

De Fine Licht, J., Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When does transparency generate legitimacy? Experimenting on a context-bound relationship. *Governance, 27*(1), 111-134.

De Laat, P. B. (2017). Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? *Philosophy & Technology*, 1-17.

de Mingo, A. C., & Cerrillo-i-Martínez, A. (2018). Improving records management to promote transparency and prevent corruption. *International Journal of Information Management, 38*(1), 256-261.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems, 19*(4), 9-30.

Deng, S., Peng, J., & Wang, C. (2013). Fiscal transparency at the Chinese provincial level. *Public Administration, 91*(4), 947-963.

Dillon, T. W., McDowell, D., Salimian, F., & Conklin, D. (1998). Perceived ease of use and usefulness of bedside-computer systems. *Computers in Nursing, 16*(3), 151-156.

Duchemin, R., & Matheus, R. (2021). Forecasting customer churn: Comparing the performance of statistical methods on more than just accuracy. *Journal of Supply Chain Management Science, 2*(3-4), 115-137.

Dwivedi, Y. K., Janssen, M., Slade, E. L., Rana, N. P., Weerakkody, V., Millard, J., . . . Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling. *Information Systems Frontiers, 19*(2), 197-212. doi:10.1007/s10796-016-9675-5

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of management review, 14*(1), 57-74.

Ellison, A. M. (2010). Repeatability and transparency in ecological research. *Ecology, 91*(9), 2536-2539.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing, 62*(1), 107-115.

Elragal, A. (2014). ERP and big data: the inept couple. *Procedia Technology, 16*, 242-249.

European_Commission. (2010). *The European eGovernment Action Plan 2011-2015*. Retrieved from Brussels:

Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management, 32*(1), 172-185.

Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics, 29*(1), 38-51.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review, 1*(2), 293-314.

Fang, J., Zhao, L., & Li, S. (2024). Exploring open government data ecosystems across data, information, and business. Government Information Quarterly, 41(2), 101934.

Faust, J., & Svensson, L. E. (2001). Transparency and credibility: monetary policy with unobservable goals. *International Economic Review, 42*(2), 369-397.

Fenster, M. (2005). The opacity of transparency. *Iowa L. Rev., 91*, 885.

Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M. J., Benítez, J. M., & Herrera, F. (2014). Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4*(5), 380-409.

Fink, A. (2019). *Conducting research literature reviews: From the internet to paper*: Sage publications.

Foerstel, H. N. (1999). *Freedom of information and the right to know: the origins and applications of the Freedom of Information Act*: Greenwood Publishing Group.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Frank, M., & Oztoprak, A. A. (2015). *Concepts of Transparency: Open Data in UK Local Authorities.* Paper presented at the Proceedings of the International Conference for E-Democracy and Open Government 2015 (CeDEM15).

French, P. E. (2011). Enhancing the legitimacy of local government pandemic influenza planning through transparency and public engagement. *Public Administration Review, 71*(2), 253-264.

Fung, A. (2013). Infotopia: Unleashing the democratic power of transparency. *Politics & Society, 41*(2), 183-212.

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*: Cambridge University Press.

Furtado, V., Ayres, L., De Oliveira, M., Vasconcelos, E., Caminha, C., D'Orleans, J., & Belchior, M. (2010). Collective intelligence in law enforcement–The WikiCrimes system. *Information Sciences, 180*(1), 4-17.

Ganapati, S., & Reddick, C. G. (2014). The use of ICT for open government in US municipalities: Perceptions of chief administrative officers. *Public Performance & Management Review, 37*(3), 365-387.

Gandía, J. L., Marrahí, L., & Huguet, D. (2016). Digital transparency and Web 2.0 in Spanish city councils. *Government Information Quarterly, 33*(1), 28-39.

Garcia, G., Vaz, J. C., Matheus, R., & Jayo, M. (2022). Open Government Data and Urban Mobility: Comparative Analysis of Buenos Aires and Rio de Janeiro. *Administração Pública e Gestão Social, 14*(4).

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. In: Academy of Management Briarcliff Manor, NY.

Giddens, A. (2013). *The consequences of modernity*: John Wiley & Sons.

Gil, D., & Song, I.-Y. (2016). Modeling and Management of Big Data: Challenges and opportunities. *Future Generation Computer Systems, 63*, 96-99. doi:https://doi.org/10.1016/j.future.2015.07.019

Gilb, T. (1997). Towards the engineering of requirements. *Requirements engineering, 2*(3), 165-169.

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*.

González-Zapata, F., & Heeks, R. (2016). *The influence of the transparency agenda on open government data in Chile*. Paper presented at the 2016 Conference for E-Democracy and Open Government (CeDEM).

Graham, F. S., Gooden, S. T., & Martin, K. J. (2016). Navigating the Transparency–Privacy Paradox in Public Sector Data Sharing. *The American Review of Public Administration, 46*(5), 569-591.

Gregor, S. (2006). The nature of theory in information systems. 611-642.

Grimmelikhuijsen, S. (2009). Do transparent government agencies strengthen trust? *Information Polity, 14*(3), 173-186.

Grimmelikhuijsen, S. (2011). Being transparent or spinning the message? An experiment into the effects of varying message content on trust in government. *Information Polity, 16*(1), 35-50.

Grimmelikhuijsen, S. (2012a). A good man but a bad wizard. About the limits and future of transparency of democratic governments. *Information Polity, 17*(3, 4), 293-302.

Grimmelikhuijsen, S. (2012b). Linking transparency, knowledge and citizen trust in government: An experiment. *International Review of Administrative Sciences, 78*(1), 50-73.

Grimmelikhuijsen, S., & Klijn, A. (2015). The effects of judicial transparency on public trust: evidence from a field experiment. *Public Administration, 93*(4), 995-1011.

Grimmelikhuijsen, S. G., & Feeney, M. K. (2016). Developing and Testing an Integrative Framework for Open Government Adoption in Local Governments. *Public Administration Review*.

Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory, 24*(1), 137-157.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research, 2*(163-194), 105.

Guillamón, M.-D., Ríos, A.-M., Gesuele, B., & Metallo, C. (2016). Factors influencing social media use in local governments: The case of Italy and Spain. *Government Information Quarterly, 33*(3), 460-471.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1988). *Multivariate data analysis* (5 ed.): Prentice hall

Halachmi, A., & Greiling, D. (2013). Transparency, e-government, and accountability: Some issues and considerations. *Public Performance & Management Review, 36*(4), 562-584.

Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism, 18*(4), 408-424.

Hansson, K., Belkacem, K., & Ekenberg, L. (2015). Open government and democracy: A research review. Social Science Computer Review, 33(5), 540-555.

Hardy, K., & Maurushat, A. (2017). Opening up government data for Big Data analysis and public benefit. *Computer Law & Security Review, 33*(1), 30-37.

Harman, H. H. (1976). *Modern factor analysis*: University of Chicago press.

Harrison, T., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., . . . Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information polity, 17*(2), 83-97.

Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet, 4*(4), 900-928.

Harrison, T. M., & Sayogo, D. S. (2014). Transparency, participation, and accountability practices in open government: A comparative study. *Government Information Quarterly, 31*(4), 513-525.

Heeks, R. (2001). *Understanding e-governance for development*: Institute for Development Policy and Management Manchester.

Heimstädt, M. J. T. f., & change, s. (2017). Openwashing: A decoupling perspective on organizational transparency. *125*, 77-86.

Helbig, N., Styrin, E., Canestraro, D., & Pardo, T. (2010). *Information and transparency: learning from recovery act reporting experiences*. Paper presented at the Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities.

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial management & data systems, 116*(1), 2-20.

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In *Design research in information systems* (pp. 9-22): Springer.

Hildebrandt, M. (2012). The dawn of a critical transparency right for the profiling era. *Digital enlightenment yearbook 2012*, 41-56.

Hildebrandt, M. (2013). Profile transparency by design?: Re-enabling double contingency. *Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology*, 221-246.

Hobbs, J. E. (2004). Information asymmetry and the role of traceability systems. *Agribusiness: An International Journal, 20*(4), 397-415.

Höffner, K., Martin, M., & Lehmann, J. (2016). LinkedSpending: openspending becomes linked open data. *Semantic Web, 7*(1), 95-104.

Hoitash, R., & Hoitash, U. (2017). Measuring Accounting Reporting Complexity with XBRL. *The Accounting Review*. doi:10.2139/ssrn.2433677

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2014). Measuring transparency. *Political analysis*, mpu001.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles*, 2.

Hossain, M. A., Dwivedi, Y. K., & Rana, N. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of organizational computing electronic commerce, 26*(1-2), 14-40.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research, 15*(9), 1277-1288.

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access, 2*, 652-687.

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. (2016). Reproducible research practices and transparency across the biomedical literature. *PLoS biology, 14*(1), e1002333.

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer, 51*(8), 56-59.

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly, 27*(4), 371-376.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly, 28*(4), 446-456.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29*(4), 258-268.

Janssen, M., Estevez, E., & Janowski, T. (2014). Interoperability in Big, Open, and Linked Data--Organizational Maturity, Capabilities, and Data Portfolios. *Computer*(10), 44-49.

Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review, 40*(2), 478-493.

Janssen, M., Janssen, M., Matheus, R., Matheus, R., Longo, J., Longo, J., . . . Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy, 11*(1), 2-8.

Janssen, M., & Kuk, G. (2015). Big and Open Linked Data (BOLD) in Research, Policy and Practice. *Journal of Organizational Computing and Electronic Commerce*(just-accepted).

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy, 11*(1), 2-8.

Janssen, M., Matheus, R., & Zuiderwijk, A. (2015). Big and Open Linked Data (BOLD) to Create Smart Cities and Citizens: Insights from Smart Energy and Mobility Cases. In E. Tambouris, M. Janssen, H. J. Scholl, M. A. Wimmer, K. Tarabanis, M. Gascó, B. Klievink, I. Lindgren, & P. Parycek (Eds.), *Electronic Government* (Vol. 9248, pp. 79-90): Springer International Publishing.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly, 32*(4), 363-368.

Jeng, J. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. *Libri, 55*(2-3), 96-121.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305-360. doi:https://doi.org/10.1016/0304-405X(76)90026-X

Jensen, M. C., & Smith, C. W. (2000). Stockholder, manager, and creditor interests: Applications of agency theory. *Theory of the Firm, 1*(1).

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). The generative mechanisms of open government data.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research, 2*(2), 59-64. doi:https://doi.org/10.1016/j.bdr.2015.01.006

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional, 15*(6), 43-48.

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing, 72*(12), 2954-2965.

Kashin, K., King, G., & Soneji, S. (2015). Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts. *Political Analysis, 23*(3), 336-362.

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly, 30*(4), 508-513. doi:http://dx.doi.org/10.1016/j.giq.2013.05.012

Kierkegaard, S. (2009). Open access to public documents–More secrecy, less transparency! *Computer Law Security Review, 25*(1), 3-27.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*(2004), 1-26.

Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science, 2*(1), 6-28.

Kizilcec, R. F. (2016). *How much information?: Effects of transparency on trust in an algorithmic interface*. Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.

Klein, T. J., Lambertz, C., & Stahl, K. O. (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy, 124*(6), 1677-1713.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1-10.

Kosack, S., & Fung, A. (2014). Does transparency improve governance? *Annual Review of Political Science, 17*, 65-87.

Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities–An experiment on stockholm. *Land use policy, 71*, 24-35.

Kruse, C. S., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and Opportunities of Big Data in Health Care: A Systematic Review. *JMIR Med Inform, 4*(4), e38. doi:10.2196/medinform.5359

Lapão, L. V., Da Silva, M. M., & Gregório, J. (2017). Implementing an online pharmaceutical service using design science research. *BMC medical informatics decision making, 17*(1), 1-14.

Lean, O. K., Zailani, S., Ramayah, T., & Fernando, Y. (2009). Factors influencing intention to use e-government services among citizens in Malaysia. *International Journal of Information Management, 29*(6), 458-475.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293-303. doi:https://doi.org/10.1016/j.bushor.2017.01.004

Liebwald, D. (2015). On transparent law, good legislation and accessibility to legal information: Towards an integrated legal information system. *Artificial Intelligence and Law, 23*(3), 301-314.

Lin, H.-F. (2007). Measuring online learning systems success: Applying the updated DeLone and McLean model. *Cyberpsychology & behavior, 10*(6), 817-820.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research, 4*, 97-128.

Lnenicka, M., Nikiforova, A., Luterek, M., Milic, P., Rudmark, D., Neumaier, S., Santoro, C., Flores, C.C., Janssen, M. & Bolívar, M. P. R. (2024). Identifying patterns and recommendations of and for sustainable open data initiatives: a benchmarking-driven analysis of open government data initiatives among European countries. Government Information Quarterly, 41(1), 101898.

Longhurst, R. J. K. m. i. g. (2003). Semi-structured interviews and focus groups. *3*(2), 143-156.

Louise Barriball, K., & While, A. J. J. o. a. n. (1994). Collecting Data using a semi-structured interview: a discussion paper. *19*(2), 328-335.

Lourenço, R. P. (2013). Data disclosure and transparency for accountability: A strategy and case analysis. *Information Polity, 18*(3), 243-260.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*.

Lourenço, R. P. (2016). Evidence of an Open Government Data Portal Impact on the Public Sphere. *International Journal of Electronic Government Research (IJEGR), 12*(3), 21-36.

Lourenço, R. P. (2023). Government transparency: Monitoring public policy accumulation and administrative overload. *Government Information Quarterly, 40*(1), 101762.

Luna-Reyes, L. F., Bertot, J. C., & Mellouli, S. (2014). Open Government, Open Data and Digital Government. *Government Information Quarterly, 31*(1), 4-5.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. In: Taylor & Francis.

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of theoretical and applied electronic commerce research, 12*(1), 21-41.

Manolea, B., & Cretu, V. (2013). Topic Report No. 2013/10: The influence of the Open Government Partnership (OGP) on the Open Data discussions. In: European Public Sector Information Platform, http://epsiplatform. eu/topicreports.

Martin, M., Stadler, C., Frischmuth, P., & Lehmann, J. (2014). Increasing the financial transparency of european commission project funding. *Semantic Web, 5*(2), 157-164.

Mascharka, D., Tran, P., Soklaski, R., & Majumdar, A. (2018). *Transparency by design: Closing the gap between performance and interpretability in visual reasoning.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Matheus, R. (2007). *A implantação da nota fiscal eletrônica no Município de São Paulo.* Paper presented at the XII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Sto. Domingo, Rep. Dominicana.

Matheus, R. (2009). *Consultas públicas do Governo Federal e agências reguladoras.* Paper presented at the Congresso CONSAD, Brasília.

Matheus, R. (2017). *Designing and Evaluation Transparency in Open Government.* Paper presented at the Conference for E-Democracy and Open Government.

Matheus, R., Alexopoulos, C., Rizun, N., Loukis, E., & Saxena, S. (2024). Impact of information systems (IS) infusion on Open Government Data (OGD) adoption. *Digital Policy, Regulation and Governance, 26*(2), 209-222.

Matheus, R., Angélico, F., & Atoji, M. I. (2014). Dados Abertos no Jornalismo: Os Limites e os Desafios das Estratégias de Uso e Criação de Cadeia de Valor Social incentivando a transparência e controle social na América Latina. OD4D, 2014. In.

Matheus, R., Crantschaninov, T. I., & Nerling, M. A. (2011). *Interlegis program in Brazil: use of web 2.0, collaborative tools and free software in legislative power.* Paper presented at the Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., & de Lemos, M. L. (2014). Administration beyond enterprises: The fundamental role of management in the government. *Revistra Brasileira de Administração Científica*.

Matheus, R., Faber, R., Ismagilova, E., & Janssen, M. (2023). Digital transparency and the usefulness for open government. *International Journal of Information Management, 73*, 102690.

Matheus, R., & Janssen, M. Factors impacting the development of Public Organizational Models: Electronic Government Infrasctructure as the Foundation. *TIC*, 247.

Matheus, R., & Janssen, M. (2013). *Transparency of civil society websites: towards a model for evaluation websites transparency.* Paper presented at the Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., & Janssen, M. (2015a). Transparency Dimensions of Big and Open Linked Data. In M. Janssen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. van Loenen, & A. Zuiderwijk (Eds.), *Open and Big Data Management and Innovation* (Vol. 9373, pp. 236-246): Springer International Publishing.

Matheus, R., & Janssen, M. (2015b). *Transparency dimensions of big and open linked data: Transparency as being synonymous with accountability and openness.* Paper presented at the Open and Big Data Management and Innovation: 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015, Delft, The Netherlands, October 13-15, 2015, Proceedings 14.

Matheus, R., & Janssen, M. (2016a). *Exploitation and Exploration Strategies to Create Data Transparency in the Public Sector*. Paper presented at the Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, Montevideo, Uruguay.

Matheus, R., & Janssen, M. (2016b). *Towards an ambidextrous government: Strategies for balancing exploration and exploitation in open government.* Paper presented at the Proceedings of the 17th International Digital Government Research Conference on Digital Government Research.

Matheus, R., & Janssen, M. (2017a). *An evaluation framework for linked open statistical data in government.* Paper presented at the Electronic Government: 16th IFIP WG 8.5 International Conference, EGOV 2017, St. Petersburg, Russia, September 4-7, 2017, Proceedings 16.

Matheus, R., & Janssen, M. (2017b). *How to Become a Smart City?: Balancing Ambidexterity in Smart Cities.* Paper presented at the Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., & Janssen, M. (2018). *D4.6. Pilots Evaluation results – Third round of OpenGovIntelligence European Commission EC Project*. Retrieved from http://www.opengovintelligence.eu/downloads/deliverables/OGI_D4.6_Pilots%20Evaluation%20Results%20Third%20Round_v1.pdf

Matheus, R., & Janssen, M. (2019). A systematic literature study to unravel transparency enabled by open government data: The Window Theory. *Public Performance & Management Review (PPMR)*.

Matheus, R., & Janssen, M. (2020). A systematic literature study to unravel transparency enabled by open government data: The window theory. *Public Performance & Management Review, 43*(3), 503-534.

Matheus, R., Janssen, M., & Janowski, T. (2021). Design principles for creating digital transparency in government. *Government Information Quarterly, 38*(1), 101550. Retrieved from https://reader.elsevier.com/reader/sd/pii/S0740624X20303294?token=7A611BC40D7D3044D034763EE6E23F7DB49F55A0CE2201123664EB591F406B5A6B7751601DF778E5EB03CAA9F2FD9628

Matheus, R., Janssen, M., & Maheshwari, D. (2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*.

Matheus, R., & Maia Ribeiro, M. (2014a). Case study: open government data in Rio de Janeiro City.

Matheus, R., & Maia Ribeiro, M. (2014b). Open Data in Legislative: the case of São Paulo city council.

Matheus, R., & Ribeiro, M. M. (2009a). *Models for citizen engagement in Latin American: case studies of public digital budgeting.* Paper presented at the Proceedings of the 3rd international conference on Theory and practice of electronic governance.

Matheus, R., & Ribeiro, M. M. (2009b). *Public online consultation of federal ministries and federal regulatory agencies in Brazil.* Paper presented at the Proceedings of the 3rd international conference on Theory and practice of electronic governance.

Matheus, R., & Ribeiro, M. M. (2009c). *Telemedicine in Brazilian public policy management.* Paper presented at the Proceedings of the 3rd international conference on Theory and practice of electronic governance.

Matheus, R., Ribeiro, M. M., & Vaz, J. C. (2012). *New perspectives for electronic government in Brazil: The adoption of open government data in national and subnational governments of Brazil.* Paper presented at the Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., Ribeiro, M. M., & Vaz, J. C. (2015). Brazil Towards Government 2.0: Strategies for Adopting Open Government Data in National and Subnational Governments. In *Case Studies in e-Government 2.0* (pp. 121-138): Springer.

Matheus, R., Ribeiro, M. M., & Vaz, J. C. (2018). Strategies and instruments for the dissemination and promotion of open government data use in Brazil: case study of Rio de Janeiro city hall. *Revista Tecnologia e Sociedade, 14*(33), 172-189.

Matheus, R., Ribeiro, M. M., Vaz, J. C., & de Souza, C. A. (2010a). *Case studies of digital participatory budgeting in Latin America: models for citizen engagement.* Paper presented at the Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., Ribeiro, M. M., Vaz, J. C., & de Souza, C. A. (2010b). *Internet use for social control and participation: what are local governments doing in Latin America?* Paper presented at the Proceedings of the 4th international conference on theory and practice of electronic governance.

Matheus, R., Ribeiro, M. M., Vaz, J. C., & de Souza, C. A. (2010c). *Using internet to promote the transparency and fight corruption: Latin American transparency portals.* Paper presented at the Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., Ribeiro, M. M., Vaz, J. C., & de Souza, C. A. (2012). *Anti-corruption online monitoring systems in Brazil.* Paper presented at the Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance.

Matheus, R., Rodrigues, D., Vaz, J., & Jayo, M. (2016). Analysis of the openness level of governmental data about the brazilian motor vehicle traffic. *Revista Eletrônica de Sistemas de Informação, 15*(2), 1-19.

Matheus, R., Saxena, S., & Alexopoulos, C. (2024). Role of perceived technological innovativeness (PTI) across genders vis-à-vis open government data (OGD) adoption and usage. *The International Journal of Information and Learning Technology*.

Matheus, R., Vaz, J. C., & Ribeiro, M. M. (2010a). *Electronic Government and the right promotion of public management participation: Models of Citizenship participation in Latin America.* Paper presented at the International Conference on Information Systems and Technology Management – CONTECSI, Sao Paulo.

Matheus, R., Vaz, J. C., & Ribeiro, M. M. (2010b). *The Map of Transparency and Social Control of the Electronic Procurement State Legislative Assemblies (ALES): Case Study of Brazil to ALES.* Paper presented at the International Conference on Information Systems and Technology Management.

Matheus, R., Vaz, J. C., & Ribeiro, M. M. (2011). *Anti-Corruption online tools: Online monitoring systems against corruption in Latin America.* Paper presented at the First Global Conference on Transparency Research. Newark, NJ.

Matheus, R., Vaz, J. C., & Ribeiro, M. M. (2014a). *Dados Governamentais Abertos e o ciclo virtuoso de uso de dados e informações públicas para o controle social, melhoria dos serviços públicos e desenvolvimento da economia: Estudo de Caso da Cidade do Rio de Janeiro.* Paper presented at the Workshop de Transparência em Sistemas.

Matheus, R., Vaz, J. C., & Ribeiro, M. M. (2014b). *Open government data and the data usage for improvement of public services in the Rio de Janeiro City.* Paper presented at the Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance.

McBride, K., Matheus, R., Toots, M., Kalvet, T., & Krimmer, R. (2018). *The role of linked open statistical data in public service co-creation.* Paper presented at the Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance.

McCormick, A. C. (2010). Here's looking at you: Transparency, institutional self-presentation, and the public interest. *Change: The Magazine of Higher Learning, 42*(6), 35-43.

McDermott, P. (2010). Building open government. *Government Information Quarterly, 27*(4), 401-413.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods, 7*(1), 64.

McDonnell, L., Scott, S., & Dawson, M. J. Q. R. (2017). A multidimensional view? Evaluating the different and combined contributions of diaries and interviews in an exploration of asexual identities and intimacies. *17*(5), 520-536.

McNeal, R., & Hale, K. (2010). E-disclosure of campaign finance information: Divergent interests in the states. *Journal of Information Technology & Politics, 7*(1), 52-66.

Meijer, A. (2013). Understanding the complex dynamics of transparency. *Public Administration Review, 73*(3), 429-439.

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of theoretical and applied electronic commerce research, 9*(3), 32-44.

Mergel, I. (2012). The social media innovation challenge in the public sector. *Information Polity, 17*(3, 4), 281-292.

Meth, H., Mueller, B., & Maedche, A. (2015). Designing a requirement mining system. *Journal of the Association for Information Systems, 16*(9), 2.

Michener, G. (2015). How cabinet size and legislative control shape the strength of transparency laws. *Governance, 28*(1), 77-94.

Michener, G., & Bersch, K. (2013). Identifying transparency. *Information Polity, 18*(3), 233-242.

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics, 22*(2), 303-341.

Mol, A. P. (2010). The future of transparency: Power, pitfalls and promises. *Global Environmental Politics, 10*(3), 132-143.

Morgan, P., Cleave-Hogg, D., DeSousa, S., & Tarshis, J. (2004). High-fidelity patient simulation: validation of performance checklists. *British Journal of Anaesthesia, 92*(3), 388-392.

Muñoz-Soro, J. F., Esteban, G., Corcho, O., & Serón, F. (2016). PPROC, an ontology for transparency in public procurement. *Semantic Web, 7*(3), 295-309.

Munson, S. A., Avrahami, D., Consolvo, S., Fogarty, J., Friedman, B., & Smith, I. (2012). Sunlight or sunburn: A survey of attitudes toward online availability of US public records. *Information Polity, 17*(2), 99-114.

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big data challenges in building the global earth observation system of systems. *Environmental Modelling & Software, 68*, 1-26.

Navarro-Galera, A., Alcaraz-Quiles, F. J., & Ortiz-Rodríguez, D. (2016). Online dissemination of information on sustainability in regional governments. Effects of technological factors. *Government Information Quarterly, 33*(1), 53-66.

Nilashi, M., Jannach, D., bin Ibrahim, O., Esfahani, M. D., & Ahmadi, H. (2016). Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications, 19*, 70-84.

Nograšek, J., & Vintar, M. (2014). E-government and organisational transformation of government: Black box revisited? *Government Information Quarterly, 31*(1), 108-118.

O'Hara, K. (2012). *Transparency, open data and trust in government: shaping the infosphere.* Paper presented at the Proceedings of the 4th Annual ACM Web Science Conference.

O'Connor, C., & Kelly, S. (2017). Facilitating knowledge management through filtered big data: SME competitiveness in an agri-food sector. *Journal of Knowledge Management, 21*(1), 156-179.

Obama, B. (2009). Transparency and open government. *Memorandum for the heads of executive departments and agencies*.

Obama, B. (2013). Making open and machine readable the new default for Government information. *Washington DC, Executive Order*.

Ohemeng, F. L., & Ofosu-Adarkwa, K. (2014). Promoting Transparency and Strengthening Public Trust in Government through Information Communication Technologies?: A Study of Ghana's E-Governance Initiative. *International Journal of Public Administration in the Digital Age (IJPADA), 1*(2), 25-42.

Ojo, A. I. (2017). Validation of the DeLone and McLean information systems success model. *Healthcare informatics research, 23*(1), 60-66.

Ølnes, S. (2016). Beyond Bitcoin Enabling Smart Government Using Blockchain Technology. In H. J. Scholl, O. Glassey, M. Janssen, B. Klievink, I. Lindgren, P. Parycek, E. Tambouris, M. A. Wimmer, T. Janowski, & D. Sá Soares (Eds.), *Electronic Government: 15th IFIP WG 8.5 International Conference, EGOV 2016, Guimarães, Portugal, September 5-8, 2016, Proceedings* (pp. 253-264). Cham: Springer International Publishing.

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2017). Big Data technologies: A survey. *Journal of King Saud University – Computer and Information Sciences*. doi:https://doi.org/10.1016/j.jksuci.2017.06.001

Owen, B. B., Cooke, L., & Matthews, G. (2013). The development of UK government policy on citizens' access to public sector information. *Information Polity, 18*(1), 5-19.

Packer, H. S., Diochnos, D., Rovatsos, M., Gal, Y. a., & Moreau, L. (2014). Semantics and Provenance for Accountable Smart City Applications. *Semantic Web–Interoperability, Usability*.

Paik, H.-Y., Xu, X., Bandara, H. D., Lee, S. U., & Lo, S. K. (2019). Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance. *IEEE Access, 7*, 186091-186107.

Panda, B., & Leepsa, N. (2017). Agency theory: Review of theory and evidence on problems and perspectives. *Indian Journal of Corporate Governance, 10*(1), 74-95.

Pardo, T. A., Nam, T., & Burke, G. B. (2012). E-government interoperability: Interaction of policy, management, and technology dimensions. *Social Science Computer Review, 30*(1), 7-23.

Park, S., & Gil-Garcia, J. R. (2017). *Understanding Transparency and Accountability in Open Government Ecosystems: The Case of Health Data Visualizations in a State Government.* Paper presented at the Proceedings of the 18th Annual International Conference on Digital Government Research.

Parnas, D. L., & Siewiorek, D. P. (1975). Use of the concept of transparency in the design of hierarchically structured systems. *Communications of the ACM, 18*(7), 401-408.

Paroutis, S., Bennett, M., Heracleous, L. J. T. F., & Change, S. (2014). A strategic view on smart city technology: The case of IBM Smarter Cities during a recession. *89*, 262-272.

Parris, D. L., Dapko, J. L., Arnold, R. W., & Arnold, D. (2016). Exploring transparency: a new framework for responsible business management. Management Decision, 54(1), 222-247.

Pavlou, P. A., & Gefen, D. (2005). Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating role. *Information Systems Research, 16*(4), 372-399.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45-77.

Peiffer, C., & Alvarez, L. (2016). Who Will Be the "Principled-Principals"? Perceptions of Corruption and Willingness to Engage in Anticorruption Activism. *Governance, 29*(3), 351-369.

Peixoto, T. (2013). The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson's' The New Ambiguity of Open Government'.

Pelucchi, M., Psaila, G., & Toccu, M. (2017). *The challenge of using map-reduce to query open data.* Paper presented at the Proceedings of the 6th International Conference on Data Science Technologies and Applications DATA-2017, INSTICC. ScitePress, Madrid.

Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. (2015). Big data privacy in the internet of things era. *IT Professional, 17*(3), 32-39.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623-656.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*: John Wiley & Sons.

Petyuk, V. A., Gatto, L., & Payne, S. H. (2019). Reproducibility and transparency by design. *Molecular & Cellular Proteomics, 18*(8 suppl 1), S202-S204.

Phan, H. P., & Ngu, B. H. (2021). Perceived 'optimal efficiency': theorization and conceptualization for development and implementation. Heliyon, 7(1).

Philip Chen, C. L., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*, 314-347. doi:https://doi.org/10.1016/j.ins.2014.01.015

Pina, V., Torres, L., & Acerete, B. (2007). Are ICTs promoting government accountability?: A comparative analysis of e-governance developments in 19 OECD countries. *Critical Perspectives on Accounting, 18*(5), 583-602.

Pina, V., Torres, L., & Royo, S. (2007). Are ICTs improving transparency and accountability in the EU regional and local governments? An empirical study. *Public administration, 85*(2), 449-472.

Pinheiro Junior, L., Alexandra Cunha, M., Janssen, M., & Matheus, R. (2020). *Towards a framework for cloud computing use by governments: Leaders, followers and laggers.* Paper presented at the The 21st Annual International Conference on Digital Government Research.

Pinheiro, L., & Matheus, R. (2022). Designing Business Analytics Projects (BAP): A Five-Step Dashboarding Cycle. In *Handbook of Research on Foundations and Applications of Intelligent Business Analytics* (pp. 71-94): IGI Global.

Piotrowski, S. J., & Van Ryzin, G. G. (2007). Citizen attitudes toward transparency in local government. *The American Review of Public Administration, 37*(3), 306-323.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. J. J. o. a. p. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *88*(5), 879.

Porumbescu, G. (2015). Linking transparency to trust in government and voice. *The American Review of Public Administration*, 0275074015607301.

Pottier, S. W., & Sommer, D. W. (1997). Agency theory and life insurer ownership structure. *Journal of Risk and Insurance*, 529-543.

Praditya, D., Janssen, M., & Sulastri, R. (2017). Determinants of Business-to-Government Information Sharing Arrangements. *Electronic Journal of e-Government, 15*(1).

Praditya, D., Sulastri, R., Bharosa, N., & Janssen, M. (2016). *Exploring XBRL-Based Reporting System: A Conceptual Framework for System Adoption and Implementation.* Paper presented at the Conference on e-Business, e-Services and e-Society.

Puron-Cid, G., Gil-Garcia, J. R., & Luna-Reyes, L. F. (2012). *IT-enabled policy analysis: new technologies, sophisticated analysis and open data for better government decisions.* Paper presented at the Proceedings of the 13th Annual International Conference on Digital Government Research.

Rader, E., Cotter, K., & Cho, J. (2018). *Explanations as mechanisms for supporting algorithmic transparency.* Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

Rawlins, B. (2008). Give the emperor a mirror: Toward developing a stakeholder measurement of organizational transparency. *Journal of Public Relations Research, 21*(1), 71-99.

Reddick, C. G., Chatfield, A. T., & Brajawidagda, U. (2016). Open government process and government transparency in crisis communication: The case of AirAsia QZ8501 crash. *Information Polity, 21*(3), 255-271.

Relly, J. E., & Sabharwal, M. (2009). Perceptions of transparency of government policymaking: A cross-national study. *Government Information Quarterly, 26*(1), 148-157.

Resch, S., Matheus, R., & FERREIRA, M. d. F. (2012). Logística reversa: o caso dos Ecopontos do Município de São Paulo. *Revista Eletrônica Gestão e Serviços, 3*(1), 413-430.

Reynaers, A., & Grimmelikhuijsen, S. (2015). Transparency in public–private partnerships: not so bad after all? *Public Administration, 93*(3), 609-626.

Ribeiro, M. M., Vaz, J. C., & Matheus, R. (2011). *Transparency in the portals of Brazilian Federal Government: The cases of E-procurement Portal and Transparency Portal.* Paper presented at the 1st Global Conference on Transparency Research.

Richardson, G. L., Jackson, B. M., & Dickson, G. W. (1990). A principles-based enterprise architecture: Lessons from Texaco and Star Enterprise. *MIS quarterly, 14*(4), 385-403.

Ríos, A.-M., Bastida, F., & Benito, B. (2016). Budget transparency and legislative budgetary oversight: An international approach. *The American Review of Public Administration, 46*(5), 546-568.

Rizun, N., Alexopoulos, C., Saxena, S., Kleiman, F., & Matheus, R. (2023). *How do personality traits influence Open Government Data (OGD) adoption and usage? Investigating the indirect and moderating effects.* Paper presented at the Proceedings of the 24th Annual International Conference on Digital Government Research.

Rizun, N., Alexopoulos, C., Saxena, S., Kleiman, F., & Matheus, R. (2024). Do personality traits influence the user's behavioral intention to adopt and use Open Government Data (OGD)? An empirical investigation. *Telematics and Informatics, 87*, 102073.

Robbins, S., & Henschke, A. (2017). The value of transparency: Bulk data and authoritarianism. *Surveillance & Society, 15*(3/4), 582-589.

Roberts, N. (2002). Keeping public officials accountable through dialogue: Resolving the accountability paradox. *Public Administration Review, 62*(2), 658-669.

Rodrigues, D. A., & Matheus, R. (2016). *OPEN DATA: uma análise do grau de publicidade dos dados dos órgãos executivos federal e estaduais da área de trânsito.* Paper presented at the Congresso CONSAD de Gestão Pública. Brasília/DF.

Rodríguez Bolívar, M. P., Alcaide Muñoz, L., & López Hernández, A. M. (2013). Determinants of financial transparency in government. *International Public Management Journal, 16*(4), 557-602.

Rogge, N., Agasisti, T., & De Witte, K. (2017). Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. *Public Policy and Administration, 32*(4), 263-281. doi:10.1177/0952076716687355

Romi, I. M. (2013). Testing delone and mclean's model in financial institutions. *American Academic & Scholarly Research Journal, 5*(3 special issue), 121.

Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. *Health affairs, 33*(7), 1115-1122.

Rowley, J. (2011). e-Government stakeholders—Who are they and what do they want? *International Journal of Information Management, 31*(1), 53-62.

Rubinfeld, D. L., & Gal, M. S. (2017). Access barriers to big data. *Ariz. L. Rev., 59*, 339.

Salonen, J., Huhtamäki, J., & Nykänen, O. (2013). *Challenges in Heterogeneous Web Data Analytics-Case Finnish Growth Companies in Social Media.* Paper presented at the Proceedings of International Conference on Making Sense of Converging Media.

Saxena, S. (2017). "Usage by stakeholders" as the objective of "transparency-by-design" in open government data: Case study of Sri Lanka's open data initiative. *Information and Learning Science, 118*(7/8), 420-432.

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics, 36*(1), 120-132.

Scholl, H. J. (2005). Motives, strategic approach, objectives & focal areas in e-Gov-induced change. *International Journal of Electronic Government Research (IJEGR), 1*(1), 59-78.

Scholl, H. J., & Luna-Reyes, L. F. (2011). *Transparency and openness in government: a system dynamics perspective.* Paper presented at the Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance.

Schudson, M. J. A. B. S. (2020). The Shortcomings of Transparency for Democracy. 0002764220945347.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*: psychology press.

Shaharudin, A., van Loenen, B., & Janssen, M. (2023). Towards a Common Definition of Open Data Intermediaries. *Digital Government: Research Practice, 4*(2), 1-21.

Shaikh, M., & Vaast, E. (2016). Folding and unfolding: balancing openness and transparency in open source communities. *Information Systems Research*.

Sharma, C. V., & Mehta, V. (2013). Paracetamol: mechanisms and updates. *Continuing Education in Anaesthesia, Critical Care & Pain, 14*(4), 153-158.

Shirish, A., Chandra, S., & Srivastava, S. C. (2021). Switching to online learning during COVID-19: Theorizing the role of IT mindfulness and techno eustress for facilitating productivity and creativity in student learning. *International Journal of Information Management, 61*, 102394.

Shkabatur, J. (2013). Transparency with (out) accountability: Open government in the United States. *Yale Law & Policy Review, 31*(1).

Simon, H. A. (1996). *The sciences of the artificial*: MIT press.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263-286.

Snippert, T., Witteveen, W., Boes, H., & Voordijk, H. (2015). Barriers to realizing a stewardship relation between client and vendor: the Best Value approach. *Construction management and economics, 33*(7), 569-586.

Song, C., & Lee, J. (2016). Citizens' use of social media in government, perceived transparency, and trust in government. *Public Performance & Management Review, 39*(2), 430-453.

Stoyanovich, J., & Howe, B. (2018). Follow the data! Algorithmic transparency starts with data transparency. In.

Strathern, M. (2000). The Tyranny of Transparency. *British Educational Research Journal, 26*(3), 309-321.

Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information systems, 13*(1), 24.

Sunstein, C. R. (2018). Output transparency vs. input transparency. In *Troubling transparency: The history and future of freedom of information* (pp. 187-205): Columbia University Press.

Tai, K.-T. (2021). Open government research over a decade: A systematic review. *Government Information Quarterly, 38*(2), 101566.

Tambouris, E., Kalampokis, E., Janssen, M., Matheus, R., Hermans, P., & Kalvet, T. (2018). *Theory and practice of linked open statistical data.* Paper presented at the Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age.

Tan, Y. (2014). Transparency without democracy: The unexpected effects of China's environmental disclosure policy. *Governance, 27*(1), 37-62.

Themudo, N. S. (2013). Reassessing the impact of civil society: nonprofit sector, press freedom, and corruption. *Governance, 26*(1), 63-89.

Thiago, P., Victor Diogho Heuer de, C., & Ana Paula Cabral Seixas, C. (2017). The Full Knowledge of Big Data in the Integration of Inter-Organizational Information: An Approach Focused on Decision Making. *International Journal of Decision Support System Technology (IJDSST), 9*(1), 16-31. doi:10.4018/IJDSST.2017010102

Thorsby, J., Stowers, G. N., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly, 34*(1), 53-61.

TOGAF. (2009). The open group architecture framework (togaf). *The Open Group, 1*.

Tolbert, C. J., & Mossberger, K. (2006). The effects of e-government on trust and confidence in government. *Public Administration Review, 66*(3), 354-369.

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives.

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *11*(2), 5-40.

Van Loenen, B., & Grothe, M. (2014). INSPIRE empowers re-use of public sector information. *International Journal of Spatial Data Infrastructures Research, 9*, 96-106.

Van Loenen, B., Zuiderwijk, A., Vancau-Wenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., ... & Flores, C. C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM-eJournal of eDemocracy and Open Government, 13(2), 1-27.

Van Ryzin, G. G., & Lavena, C. F. (2013). The credibility of government performance reporting: An experimental test. *Public Performance & Management Review, 37*(1), 87-103.

Vaz, J., Ribeiro, M., & Matheus, R. (2011). Open government data and its impact on the concepts and practices of transparency in Brazil. *Cadernos PPGAU/UFBA*, 4562.

Vaz, J. C., Ribeiro, M. M., & Matheus, R. (2011). *New perspectives for electronic governance: the adoption of open government data in Brazil.* Paper presented at the First Global Conference on Transparency Research.

Vaz, J. C., Ribeiro, M. M., & Matheus, R. (2013). *Desafios para a Governança Eletrônica e Dados Governamentais Abertos em Governos Locais.* Paper presented at the WTRANS13-Workshop de Transparência em Sistemas.

Veale, M., Binns, R., & Ausloos, J. (2018). When data protection by design and data subject rights clash. *International Data Privacy Law, 8*(2), 105-123.

Venkatesh, V., Thong, J. Y., Chan, F. K., & Hu, P. J. (2016). Managing Citizens' Uncertainty in E-Government Services: The Mediating and Moderating Roles of Transparency and Trust. *Information Systems Research, 27*(1), 87-111.

Villeneuve, J.-P. (2014). Transparency of Transparency: The pro-active disclosure of the rules governing Access to Information as a gauge of organisational cultural transformation. The case of the Swiss transparency regime. *Government Information Quarterly, 31*(4), 556-562.

Wahyudi, A., Matheus, R., & Janssen, M. (2017). *Benefits and Challenges of a Reference Architecture for Processing Statistical Data.* Paper presented at the Conference on e-Business, e-Services and e-Society.

Wamba, S. F. (2022). Impact of artificial intelligence assimilation on firm performance: The mediating effects of organizational agility and customer agility. *International Journal of Information Management, 67*, 102544.

Wang, J., Liu, W., Kumar, S., & Chang, S.-F. (2016). Learning to hash for indexing big data—a survey. *Proceedings of the IEEE, 104*(1), 34-57.

Ward, S. J. (2014). The magical concept of transparency. *Ethics for digital journalists: Emerging best practices*, 45-58.

Weerakkody, V., Kapoor, K., Balta, M. E., Irani, Z., & Dwivedi, Y. K. (2017). Factors influencing user acceptance of public sector big open data. *Production planning & control, 28*(11-12), 891-905.

Welch, E. W., & Hinnant, C. C. (2003). *Internet use, transparency, and interactivity effects on trust in government.* Paper presented at the System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on.

Wiedenhöft, G. C., Alexopoulos, C., Saxena, S., Rizun, N., & Matheus, R. (2023). Assessing the failure of Open Government Data initiatives in Brazil. *The Electronic Journal of Information Systems in Developing Countries, 89*(6), e12286.

Wiedenhöft, G. C., Matheus, R., Saxena, S., & Alexopoulos, C. (2023). Barriers towards open government data value co-creation: An empirical investigation. *The Electronic Journal of Information Systems in Developing Countries, 89*(5), e12270.

Wielki, J. (2013). *Implementation of the big data concept in organizations-possibilities, impediments and challenges.* Paper presented at the Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*: Springer.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *73*(6), 913-934.

Worthy, B. (2010). More open but not more trusted? The effect of the Freedom of Information Act 2000 on the United Kingdom central government. *Governance, 23*(4), 561-582.

Worthy, B. (2015). The impact of open data in the UK: Complex, unpredictable, and political. *Public Administration, 93*(3), 788-805.

Worthy, B. J. A. a. S. (2014). Making transparency stick: the complex dynamics of open data.

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering, 26*(1), 97-107.

Yang, T. M., & Wu, Y. J. (2021). Looking for datasets to open: An exploration of government officials' information behaviors in open data policy implementation. Government Information Quarterly, 38(2), 101574.

Yavuz, N., & Welch, E. W. (2014). Factors affecting openness of local government websites: Examining the differences across planning, finance and police departments. *Government Information Quarterly, 31*(4), 574-583. doi:10.1016/j.giq.2014.07.004

Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.

Yoon, H., Zo, H., & Ciganek, A. P. (2011). Does XBRL adoption reduce information asymmetry? *Journal of Business Research, 64*(2), 157-163.

Zakim, D., & Schwab, M. (2015). Data collection as a barrier to personalized medicine. *Trends in pharmacological sciences, 36*(2), 68-71.

Zeleti, F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly, 33*(3), 535-551.

Zicari, R. V. (2014). Big data: Challenges and opportunities. *Big data computing, 564*.

Zuiderwijk, A., Gascó, M., Parycek, P., & Janssen, M. (2014). Special issue on transparency and open data policies: guest editors' introduction. *Journal of theoretical and applied electronic commerce research, 9*(3), I-IX.

Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special Issue on Innovation through Open Data: Guest Editors' Introduction. *Journal of theoretical and applied electronic commerce research, 9*(2), i-xiii.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity, 19*(1, 2), 17-33.

Zuiderwijk, A., Janssen, M., & Parnia, A. (2013). *The complementarity of open data infrastructures: an analysis of functionalities.* Paper presented at the Proceedings of the 14th Annual International Conference on Digital Government Research.

Zuiderwijk, A., & Matheus, R. (2017). *Collaboratively Analysing Open Research Data in Virtual Research Environments–New Visionary Use Cases.* Paper presented at the Conference for E-Democracy and Open Government.

Zyl, A. (2014). How civil society organizations close the gap between transparency and accountability. *Governance, 27*(2), 347-356.

Short CV

Ricardo Matheus is a lecturer and researcher in the field of Open government Data and Infrastructures at the Information and Communication Technology research group of the Technology, Policy and Management Faculty of Delft University of Technology.

He was a lecturer at Rotterdam School of Management of Erasmus Rotterdam University (The Netherlands) teaching Data Science, Rapid Prototyping, and Programming for Managers courses.

He led WPs in the CAP4CITY Project (www.cap4city.eu/) and WPs in the H2020 OpenGovIntelligence project (www.opengovintelligence.eu) which aimed to create transparency using open government data in six international governmental pilots.

Prior to joining TU Delft, Ricardo helped to improve Big Data processes of the IBM Operation Center in Rio de Janeiro, Brazil, as a data scientist at P3NS4 innovation group of the Rio de Janeiro City Hall during period of planning and implementation of World Cup 2014 and Olympics Games 2016. He was part of the team in charge to create the Data.rio, one of the first open data portals in Latin America in 2013, with more than 3000 datasets since its initial version. In 2013, Rio de Janeiro won the Smartest City prize in the world (1st place) at Smart City Expo World Congress, the first Latin American city to win the global award.

After his Dutch experience, Ricardo led the Digital Transformation public policies at Finance Secretariat at Salvador, Bahia, Brazil. Ricardo played as main data scientist, creating the new generation of fiscal algorithms and re-shaping Brazilian artificial intelligence tax systems. During this period, Salvador achieved the highest rank (1st place) between Brazilian capitals at FIRJAN ranking of Fiscal Management and Transparency.

In 2023, Ricardo was the General Director of Strategic Projects at Blumenau city hall (Brazil), monitoring and proposing innovative solutions for more than 200 projects (around 1 Billion of Euros).

List of Publications

The full list of publications Ricardo is author or co-author is published here: https://scholar.google.com/citations?user=kV0f5YcAAAAJ&hl=en&authuser=2

Until the date of 04/08/2024, Ricardo had a total of 1604 citations. The summary is presented in Table 34. All other scientific publications are presented in Table 35 and technical reports from EU projects or non-scientific reports in Table 36.

Table 34 – Total citations, h-index and i10-index of scientific and non-scientific publications

	Date: 04/08/2024
Citations	1604
h-index	17
i10-index	24

Source: Google Scholar (04/08/2024)

https://scholar.google.com/citations?user=sBa4qpsAAAAJ&hl=pt-BR

Table 35 – Scientific publications

#	Title / URL	Citation	Citation count	Year
Articles Published Author/Co-Author – Peer-Reviewed Journals				
1	Do personality traits influence the user's behavioral intention to adopt and use Open Government Data (OGD)? An empirical investigation. https://doi.org/10.1016/j.tele.2023.102073	(Rizun, Alexopoulos, Saxena, Kleiman, & Matheus, 2024)	0	2024
2	Are creative users more apt in reusing and adopting Open Government Data (OGD)? Gender differences. https://doi.org/10.1016/j.tsc.2024.101478	(Alexopoulos, Saxena, Rizun, Matheus, & Janssen, 2024)	0	2024
3	Role of perceived technological innovativeness (PTI) across genders vis-à-vis open government data (OGD) adoption and usage. https://doi.org/10.1108/IJILT-08-2023-0139	(Ricardo Matheus, Saxena, & Alexopoulos, 2024)	0	2024
4	Impact of information systems (IS) infusion on Open Government Data (OGD) adoption. https://doi.org/10.1108/DPRG-07-2023-0107	(Ricardo Matheus, Alexopoulos, Rizun, Loukis, & Saxena, 2024)	0	2024

5	Digital transparency and the usefulness for open government. https://doi.org/10.1016/j.ijinfomgt.2023.102690.	(Ricardo Matheus et al., 2023)	0	2023
6	Integration of IoT into e-government. https://doi.org/10.1108/FS-04-2022-0048	(Chakraborty, Bhatt, & Chakravorty, 2020)	4	2023
7	Why do Open Government Data initiatives fail in developing countries? A root cause analysis of the most prevalent barriers and problems. https://onlinelibrary.wiley.com/doi/abs/10.1002/isd2.12297	(Alexopoulos et al., 2023)	0	2023
8	Assessing the failure of Open Government Data initiatives in Brazil. https://onlinelibrary.wiley.com/doi/abs/10.1002/isd2.12286	(Wiedenhöft, Alexopoulos, Saxena, Rizun, & Matheus, 2023)	0	2023
9	Barriers towards Open Government Data Value Co-Creation: An empirical investigation. https://onlinelibrary.wiley.com/doi/abs/10.1002/isd2.12270	(Wiedenhöft, Matheus, Saxena, & Alexopoulos, 2023)	3	2023
10	Open Government Data and Urban Mobility: Comparative Analysis of Buenos Aires and Rio de Janeiro. https://doi.org/10.21118/apgs.v14i4.13111	(Garcia, Vaz, Matheus, & Jayo, 2022)	0	2022

11	Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. https://journals.sagepub.com/doi/pdf/10.1177/0894439320980118	(Janssen, Hartog, Matheus, Yi Ding, & Kuk, 2022)	81	2022
12	Forecasting customer churn: Comparing the performance of statistical methods on more than just accuracy. https://journals.open.tudelft.nl/jscms/article/view/6125	(Duchemin & Matheus, 2021)	3	2021
13	Design principles for creating digital transparency in government. https://www.sciencedirect.com/science/article/pii/S0740624X20303294	(Ricardo Matheus et al., 2021)	83	2021
14	Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. https://www.sciencedirect.com/science/article/pii/S0740624X18300303	(Ricardo Matheus, Janssen, et al., 2018)	323	2020
15	A systematic literature study to unravel transparency enabled by open government data: The window theory. https://www.tandfonline.com/doi/full/10.1080/15309576.2019.1691025	(Ricardo Matheus & Janssen, 2019)	148	2020
16	Strategies and instruments for the dissemination and promotion of open government data use in Brazil: case study of Rio de Janeiro city hall	(Ricardo Matheus,	8	2018

	https://www.redalyc.org/journal/4966/496659124011/496659124011.pdf	Ribeiro, & Vaz, 2018)		
17	Analysis of the Openness Level of Governmental Data in the Brazilian Traffic Area. http://www.periodicosibepes.org.br/index.php/reinfo/article/view/2284/pdf_1	(R Matheus, Rodrigues, Vaz, & Jayo, 2016)	0	2016
18	Administration beyond enterprises: The fundamental role of management in the government. Doi:10.6008/ESS2179-684X.2013.003.0007	(Ricardo Matheus & de Lemos, 2014)	0	2014
19	Reverse logistics: the case of 'EcoPoints' São Paulo city hall. http://www.spell.org.br/documentos/ver/31195/logistica-reversa--o-caso-dos-ecopontos-do-municipio-de-sao-paulo	(Resch, Matheus, & FERREIRA, 2012)	18	2012
20	Governance, access to information and public policy councils. https://www.revistas.usp.br/rgpp/article/view/97825	(Crantschaninov, Matheus, & Silva, 2011)	0	2011
21	Internet and citizen participation in participatory digital budgeting experiences in Brazil https://periodicos.ufba.br/index.php/ppgau/article/view/5115	(Best, Ribeiro, Matheus, & Vaz, 2010)	39	2010
22	Open government data and its impacts on transparency concepts and practices in Brazil	(J. Vaz, M. Ribeiro, & R. Matheus, 2011)	97	2010

	https://periodicos.ufba.br/index.php/ppgau/article/view/5111			
Articles Published Author/Co-Author – Scientific Conferences				
23	How do personality traits influence Open Government Data (OGD) adoption and usage? Investigating the indirect and moderating effects. https://dl.acm.org/doi/pdf/10.1145/3598469.3598521	(Rizun, Alexopoulos, Saxena, Kleiman, & Matheus, 2023)	1	2023
24	Towards a framework for cloud computing use by governments: Leaders, followers and laggers https://dl.acm.org/doi/abs/10.1145/3396956.3396989	(Pinheiro Junior, Alexandra Cunha, Janssen, & Matheus, 2020)	9	2020
25	Theory and practice of linked open statistical data. https://dl.acm.org/doi/abs/10.1145/3209281.3209341	(Tambouris et al., 2018)	0	2018
26	The role of linked open statistical data in public service co-creation. https://dl.acm.org/doi/abs/10.1145/3209415.3209446	(McBride, Matheus, Toots, Kalvet, & Krimmer, 2018)	8	2018
27	Benefits and Challenges of a Reference Architecture for Processing Statistical Data. https://link.springer.com/chapter/10.1007/978-3-319-68557-1_41	(Wahyudi, Matheus, & Janssen, 2017)	3	2017

28	Collaboratively Analysing Open Research Data in Virtual Research Environments–New Visionary Use Cases. https://repository.tudelft.nl/islandora/object/uuid%3A76a909fd-60c9-432c-81e6-9d5dae4a485c	(Zuiderwijk & Matheus, 2017)	0	2017
29	Designing and Evaluation Transparency in Open Government. https://research.tudelft.nl/en/publications/designing-and-evaluation-transparency-in-open-government	(Ricardo Matheus, 2017)	2	2017
30	Transparency-by-design as a foundation for open government. https://www.emerald.com/insight/content/doi/10.1108/TG-02-2017-0015/full/html	(Janssen, Matheus, et al., 2017)	103	2017
31	How to become a smart city? Balancing ambidexterity in smart cities. https://dl.acm.org/doi/abs/10.1145/3047273.3047386	(Ricardo Matheus & Janssen, 2017b)	8	2017
32	An evaluation framework for linked open statistical data in government. https://link.springer.com/chapter/10.1007/978-3-319-64677-0_21	(Ricardo Matheus & Janssen, 2017a)	6	2017
33	Towards an ambidextrous government: Strategies for balancing exploration and exploitation in open government. https://dl.acm.org/doi/abs/10.1145/2912160.2912192	(Ricardo Matheus & Janssen, 2016b)	15	2016

34	Open Data: Publicity level analysis of Brazilian Regional Road Traffic Executive Branches https://www.onsv.org.br/analise-de-publicidade-de-dados-na-area-de-transito/	(Rodrigues & Matheus, 2016)	2	2016
35	Exploitation and exploration strategies to create data transparency in the public sector. https://dl.acm.org/doi/abs/10.1145/2910019.2910091	(Ricardo Matheus & Janssen, 2016a)	8	2016
36	Factors impacting the development of public organizational models: Electronic government infrastructure as the foundation. https://cetic.br/media/docs/publicacoes/2/TIC_eGOV_2015_LIVRO_ELETRONICO.pdf	(Ricardo Matheus & Janssen)	1	2015
37	Transparency dimensions of big and open linked data: Transparency as Being Synonymous with Accountability and Openness. https://link.springer.com/chapter/10.1007/978-3-319-25013-7_19	(Ricardo Matheus & Janssen, 2015b)	31	2015
38	Big and open linked data (BOLD) to create smart cities and citizens: Insights from smart energy and mobility cases. https://link.springer.com/chapter/10.1007/978-3-319-22479-4_6	(Janssen et al., 2015)	83	2015
39	Open government data and the data usage for improvement of public services in the Rio de Janeiro City.	(Ricardo Matheus, Vaz, & Ribeiro, 2014b)	29	2014

	https://dl.acm.org/doi/abs/10.1145/2691195.2691240			
40	Case study: open government data in Rio de Janeiro City. https://idl-bnc-idrc.dspacedirect.org/handle/10625/55360	(Ricardo Matheus & Maia Ribeiro, 2014a)	6	2014
41	Open Data in Legislative: the case of São Paulo city council. https://idl-bnc-idrc.dspacedirect.org/handle/10625/56306	(Ricardo Matheus & Maia Ribeiro, 2014b)	3	2014
42	Open Government Data and the virtuous cycle of use of public data and information for social control, improvement of public services and economic development: Rio de Janeiro Case Study. http://wtransdev.inf.puc-rio.br/artigos/artigos_WTRANS14/paper10.pdf	(Ricardo Matheus, Vaz, & Ribeiro, 2014a)	7	2013
43	Open Data in Journalism: The Limits and Challenges of Strategies for Using and Creating a Social Value Chain encouraging transparency and social control in Latin America	(Ricardo Matheus, Angélico, & Atoji, 2014)	5	2013
44	Transparency of civil society websites: towards a model for evaluation websites transparency. https://dl.acm.org/doi/abs/10.1145/2591888.2591915	(Ricardo Matheus & Janssen, 2013)	8	2013

45	Challenges for e-governance and open government data in local governments. http://wtransdev.inf.puc-rio.br/artigos/artigos_WTRANS13/wtrans13-vaz.pdf	(Vaz, Ribeiro, & Matheus, 2013)	16	2013
46	Anti-corruption online monitoring systems in Brazil. https://dl.acm.org/doi/abs/10.1145/2463728.2463809	(Ricardo Matheus, Ribeiro, Vaz, & de Souza, 2012)	20	2012
47	New perspectives for electronic government in Brazil: the adoption of open government data in national and subnational governments of Brazil. https://dl.acm.org/doi/abs/10.1145/2463728.2463734	(Ricardo Matheus, Ribeiro, & Vaz, 2012)	48	2012
48	Interlegis program in Brazil: use of web 2.0, collaborative tools and free software in legislative power. https://dl.acm.org/doi/abs/10.1145/2072069.2072145	(Ricardo Matheus, Crantschaninov, & Nerling, 2011)	1	2011
49	Anti-Corruption online tools: Online monitoring systems against corruption in Latin America. https://dl.acm.org/doi/abs/10.1145/1693042.1693124	(Ricardo Matheus, Vaz, & Ribeiro, 2011)	9	2011
50	New perspectives for electronic governance: the adoption of open government data in Brazil. https://dl.acm.org/doi/10.1145/2463728.2463734	(J. C. Vaz, M. M. Ribeiro, & R. Matheus, 2011)	8	2011

51	Transparency in the portals of Brazilian Federal Government: The cases of E-procurement Portal and Transparency Portal. www.t.ly/rAdn	(Ribeiro, Vaz, & Matheus, 2011)	8	2011
52	Internet use for social control and participation: what are local governments doing in Latin America?. https://dl.acm.org/doi/abs/10.1145/1930321.1930389	(Ricardo Matheus, Ribeiro, Vaz, & de Souza, 2010b)	6	2010
53	Using internet to promote the transparency and fight corruption: Latin American transparency portals. https://dl.acm.org/doi/abs/10.1145/1930321.1930411	(Ricardo Matheus, Ribeiro, Vaz, & de Souza, 2010c)	18	2010
54	Case studies of digital participatory budgeting in Latin America: Models for citizen engagement. https://dl.acm.org/doi/abs/10.1145/1930321.1930328	(Ricardo Matheus, Ribeiro, Vaz, & de Souza, 2010a)	35	2010
55	Electronic Government and the right promotion of public management participation: Models of Citizenship participation in Latin America	(Ricardo Matheus, Vaz, & Ribeiro, 2010a)	0	2010
56	The Map of Transparency and Social Control of the Electronic Procurement State Legislative Assemblies (ALES): Case Study of Brazil to ALES.	(Ricardo Matheus, Vaz, & Ribeiro, 2010b)	0	2010

	http://www.contecsi.tecsi.org/index.php/contecsi/7contecsi/paper/view/2400			
57	Telemedicine in Brazilian public policy management. https://dl.acm.org/doi/abs/10.1145/1693042.1693098	(Ricardo Matheus & Ribeiro, 2009c)	6	2009
58	Public online consultation of federal ministries and federal regulatory agencies in Brazil. https://dl.acm.org/doi/10.1145/1693042.1693127	(Ricardo Matheus & Ribeiro, 2009b)	3	2009
59	Models for citizen engagement in Latin American: Case studies of public digital budgeting. https://dl.acm.org/doi/abs/10.1145/1693042.1693065	(Ricardo Matheus & Ribeiro, 2009a)	14	2009
60	Consultas públicas do Governo Federal e agências reguladoras. https://consad.org.br/wp-content/uploads/2013/02/CONSULTAS-P%C3%9ABLICAS-DO-GOVERNO-FEDERAL-E-AG%C3%8ANCIAS-REGULADORAS3.pdf	(Ricardo Matheus, 2009)	14	2009
61	A implantação da nota fiscal eletrônica no Município de São Paulo.	(Ricardo Matheus, 2007)	2	2009
Articles Published Author/Co-Author – Book Chapters				
62	Designing Business Analytics Projects (BAP): A Five-Step Dashboarding Cycle. https://www.igi-	(Pinheiro & Matheus, 2022)	0	2022

	global.com/chapter/designing-business-analytics-projects-bap/298462			
63	Brazil towards government 2.0: Strategies for adopting open government data in national and subnational governments. https://link.springer.com/chapter/10.1007/978-3-319-08081-9_8	(Ricardo Matheus, Ribeiro, & Vaz, 2015)	26	2015

Table 36 – EU-funded project technical reports and non-scientific publications

#	Title / URL	Citations	Year
1	CAP4CITY Technical Workshop: Evaluation Strategies and Metrics for Massive Open Online Courses (MOOCs) – Erasmus+ EC Project https://ieeexplore.ieee.org/abstract/document/9096670	0	2020
2	OGI Technical Report – D4.6. Pilots Evaluation results – Third round OpenGovIntelligence EC Project H2020	0	2019
3	OGI Technical Report – D5.4 Report on Dissemination Activities – Y3 – OpenGovIntelligence EC Project H2020	0	2019
4	OGI Technical Report – D5.3 Report on Dissemination Activities – Y2 – OpenGovIntelligence EC Project H2020	0	2018
5	OGI Technical Report – D5.2 Report on Dissemination Activities – OpenGovIntelligence EC Project H2020	0	2017
6	OGI Technical Report – D5.1- Dissemination and Exploitation Plan Y1 – OpenGovIntelligence EC Project H2020	0	2016

7	OGI Technical Report – D4.1- Pilots and Evaluation Plan Y1 – OpenGovIntelligence EC Project H2020	0	2016

EU-Funded Projects

The PhD author was funded by two EU projects. We thank H2020 and ERASMUS+ funding for supporting this PhD thesis. The first H2020 OpenGovIntelligence – https://cordis.europa.eu/project/id/693849. The second ERASMUS+ CAP4CITY – . https://erasmus-plus.ec.europa.eu/projects/search/details/598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP.

Part of this work was funded by the European Commission within the H2020 Program in the context of the project OpenGovIntelligence www.opengovintelligence.eu under grant agreement [No. 693849]. The authors would like to cordially thank all the experts who participated in the study.

Part of this work was funded by the European Commission within the ERASMUS+ Programme in the context of the project CAP4CITY www.cap4city.eu under grant agreement [No. 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP]. The authors would like to cordially thank all the experts who participated in the study.

Propositions

accompanying the dissertation

THE WINDOW THEORY: FACTORS AND DESIGN PRINCIPLES FOR DIGITAL TRANSPARENCY

by

Ricardo Matheus.

1. **Achieving transparency should take care of not only organizational factors, but also technical factors (this proposition pertains to Chapter 3 of this dissertation).**
2. **Existing transparency models are too narrow and ignore essential factors for use in the digital world (this proposition pertains to Chapter 3 of this dissertation).**
3. **Transparency requires embedding the creation of transparency in daily organizational structure and in the design of information systems (this proposition pertains to Chapter 3 of this dissertation).**
4. **Designers of transparency websites determine what becomes transparent and the view on transparency.**
5. **Public dashboards create only a predefined view of reality and do not provide full transparency.**
6. **The benefits of dashboards are not easily gained because design principles are difficult to bring into practice.**
7. **Transparency-by-design should be enforced via regulations.**
8. **Only neutral organizations can create transparency without bias and prejudice.**
9. **Full transparency might not be always desired.**
10. **Brazilian and Dutch planning culture are similar and different. Brazilians pretend to plan, whereas for the Dutch, planning is a national sport.**

These propositions are regarded as opposable and defendable, and have been

approved as such by the promotor Prof.dr.ir. M.F.W.H.A Janssen and Prof.dr. T. Janowski.