Government Information Quarterly 38 (2021) 101550

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Design principles for creating digital transparency in government

Ricardo Matheus [a,*,1], Marijn Janssen [a], Tomasz Janowski [b,c]

[a] Delft University of Technology, the Netherlands
[b] Gdańsk University of Technology, Poland
[c] Danube University Krems, Austria

ARTICLE INFO

Keywords:
Transparency
Digital transparency
Transparency-by-design
Open data
Open government
Design principles
Window theory

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.

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, Ríos, Gesuele, & Metallo, 2016), lack of accountability of public officials (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, Alcaraz-Quiles, & Ortiz-Rodríguez, 2016), less corruption (John C Bertot, Jaeger, & Grimes, 2010), and more openness (Frank & Oztoprak, 2015; Matheus & Janssen, 2015). 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, 2011).

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, Bertot, & Mellouli, 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.

* Corresponding author.
E-mail address: R.Matheus@tudelft.nl (R. Matheus).
[1] Jaffalaan 5, 2628 BX Delft, South Holland, The Netherlands.

https://doi.org/10.1016/j.giq.2020.101550
Received 19 November 2020; Received in revised form 21 November 2020; Accepted 21 November 2020
Available online 9 December 2020

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, Charalabidis, & Zuiderwijk, 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* (Matheus & Janssen, 2020), with many factors relevant to digital transparency and multiple windows offered to realize such transparency. According to Matheus and Janssen (2020, 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 2 presents the research approach. Section 3 identifies barriers to digital transparency, followed by design principles and how they help overcome the barriers in Section 4. Section 5 evaluates the principles using three case studies. A discussion of the principles and their use is carried out in Section 6. Finally, Section 7 provides some conclusions.

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, Bogner, Bilgeri, Fleisch, & Wortmann, 2019). The approach is outlined in Section 2.1. Section 2.2 presents the Systematic Literature Review method, which is used to derive design principles, followed by the Case Study approach in Section 2.3, which is used to evaluate the design principles in different practical scenarios.

2.1. Design science research approach

According to Chanson, Bogner, Bilgeri, Fleisch, and Wortmann (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 artefacts". Using the design science research methodology by Peffers, Tuunanen, Rothenberger, and Chatterjee (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 artefacts 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 (Matheus & Janssen, 2020).

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 Fig. 1, 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 (Matheus, Janssen, & Maheshwari, 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.

2.2. 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 A.1, and the barriers in Tables B.1 and C.1, the latter after categorizing them into political, economic, human and social, and technological areas. All three

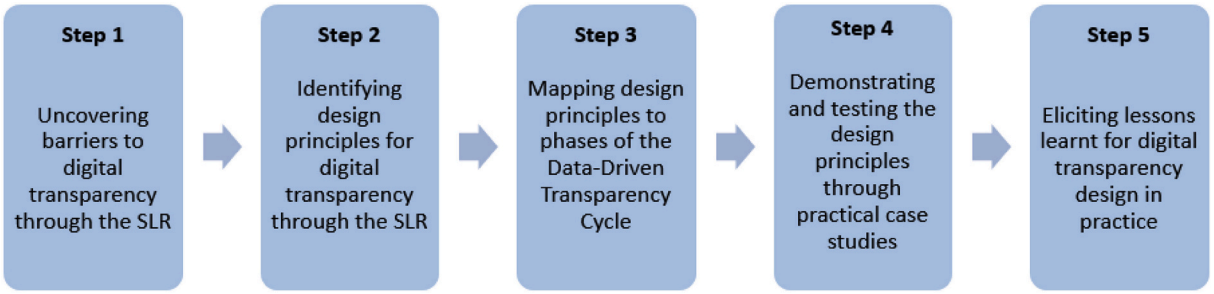

Fig. 1. Overview of the research approach.

tables are placed in Appendices A–C.

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 62 articles, 50 of which proved to be relevant to this research. In particular, the papers documenting the results of biological or medical research were excluded. The 50 remaining articles were each independently read by two researchers to identify candidates for design principles.

2.3. Evaluating design principles through case 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, policymakers, 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 Appendix D: Interview Protocol Form.

3. Barriers to digital 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 (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.

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 4.1

Table 1
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

formulates 16 design principles for digital transparency based on the Systematic Literature Review. Section 4.2 relates the 16 principles identified in Section 4.1 to the 42 barriers identified in Section 3. The resulting many-to-many mapping describes which principles help to overcome which barriers. Finally, Section 4.3 maps the design principles to different phases of the data-driven transparency cycle (Matheus et al., 2018; Matheus & Janssen, 2018), thus operationalizing the use of the principles in the engineering for data-driven transparency.

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, van Wijk, Janssen, de Winne, and Hulstijn (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 C.1 and summarized in Table 2 below.

4.2. Relating principles to barriers

The design principles for digital transparency, as described in Table 2, should help overcome the barriers to digital transparency, as described in Table 1. The matrix describing which principles address which barriers is presented in Table 3. According to this Table 3, 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 2
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

Table 3
Relationships between barriers and design principles for digital transparency.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
DQ1	x	x	x	x		x				x			x	x	x	
DQ2					x	x				x	x	x		x	x	x
DQ3	x	x		x		x	x	x	x	x	x	x	x	x	x	
DQ4		x	x		x	x				x	x	x	x			x
DQ5	x	x	x	x		x				x	x		x	x	x	
DQ6	x	x	x		x	x		x	x	x	x	x	x	x	x	
DQ7			x		x	x	x						x			
EC1		x	x	x		x	x	x	x	x	x	x	x		x	x
EC2	x	x	x							x	x		x	x	x	
EC3							x	x	x	x	x	x	x	x	x	
EC4													x			
EC5	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x
ET1		x														
ET2		x														
HR1	x	x	x	x	x					x		x	x	x	x	x
HR2	x	x	x			x				x		x	x			
HR3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HR4		x	x										x		x	
OR1		x	x	x	x	x		x		x	x	x	x	x	x	x
OR2	x	x	x			x		x	x	x	x	x	x	x	x	
OR3		x	x							x	x	x	x	x	x	
OR4		x	x										x		x	
OR5								x	x	x	x	x	x		x	
OR6	x	x	x	x		x	x	x	x	x	x	x	x		x	
OR7																x
PL1	x															
PL2	x	x											x	x	x	
PL3	x	x	x							x	x	x	x	x	x	
TE1	x			x		x	x	x	x	x	x	x	x	x		x
TE2	x		x					x	x	x	x	x	x	x	x	
TE3							x	x	x	x	x	x	x	x	x	
TE4			x			x	x						x	x		x
TE5			x							x	x	x	x	x	x	
TE6				x			x	x	x	x	x	x	x	x	x	x
TE7								x			x					
TE8		x		x		x		x	x	x	x	x	x	x	x	x
TE9	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
TE10								x	x	x	x	x	x	x	x	x
TE11	x	x	x	x		x	x	x	x	x	x	x	x	x	x	
TE12				x				x	x	x	x	x				
US1			x	x	x	x	x	x	x	x	x	x	x	x		x
US2	x	x	x									x	x			

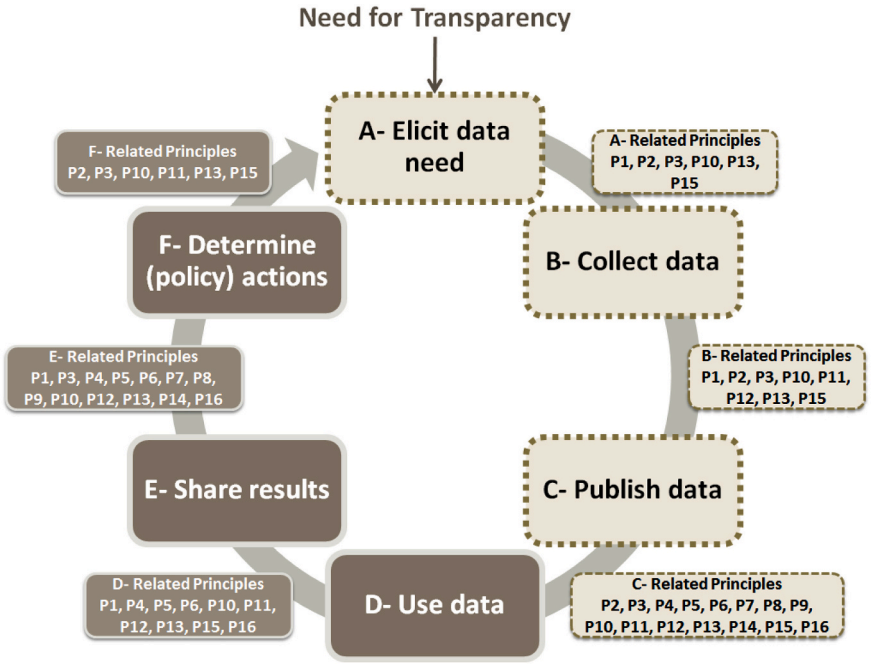

Fig. 2. Data-Driven Transparency cycle with design principles adapted from Matheus and Janssen (2018, p. 36)

4.2.1.1. 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 (*Matheus et al., *2018; Matheus & Janssen, 2018)*. The cycle is depicted in Fig. 2, adapted from fig. 8 "OGI Tools and Working Flow" in 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 Fig. 2 and elaborated in Table 4, 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.

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 5, 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.

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 6 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

Table 4
Mapping design principles to phases of the Data-Driven 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
B	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 P11 P12 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 P13 P15 P16	Privacy Data Abstraction Comprehension Data Quality Rating Data Persistency Interoperability Metadata 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		After a group of people gained meaningful insights enabled by	P2 P3	Openness Feedback

(*continued on next page*)

Table 4 (*continued*)

#	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
	Determine (policy) actions	transparency, policy action can be undertaken.	P10 P11 P13 P15	Mechanism Data Persistency Interoperability Transparency-by-Design Stewardship

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.

Fig. 3 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 Fig. 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 policymakers. 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.

Fig. 4 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,*

Table 6
The use of design principles when building applications.

Rank	Design principles		Usage	
			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%

Table 5
Overview of 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

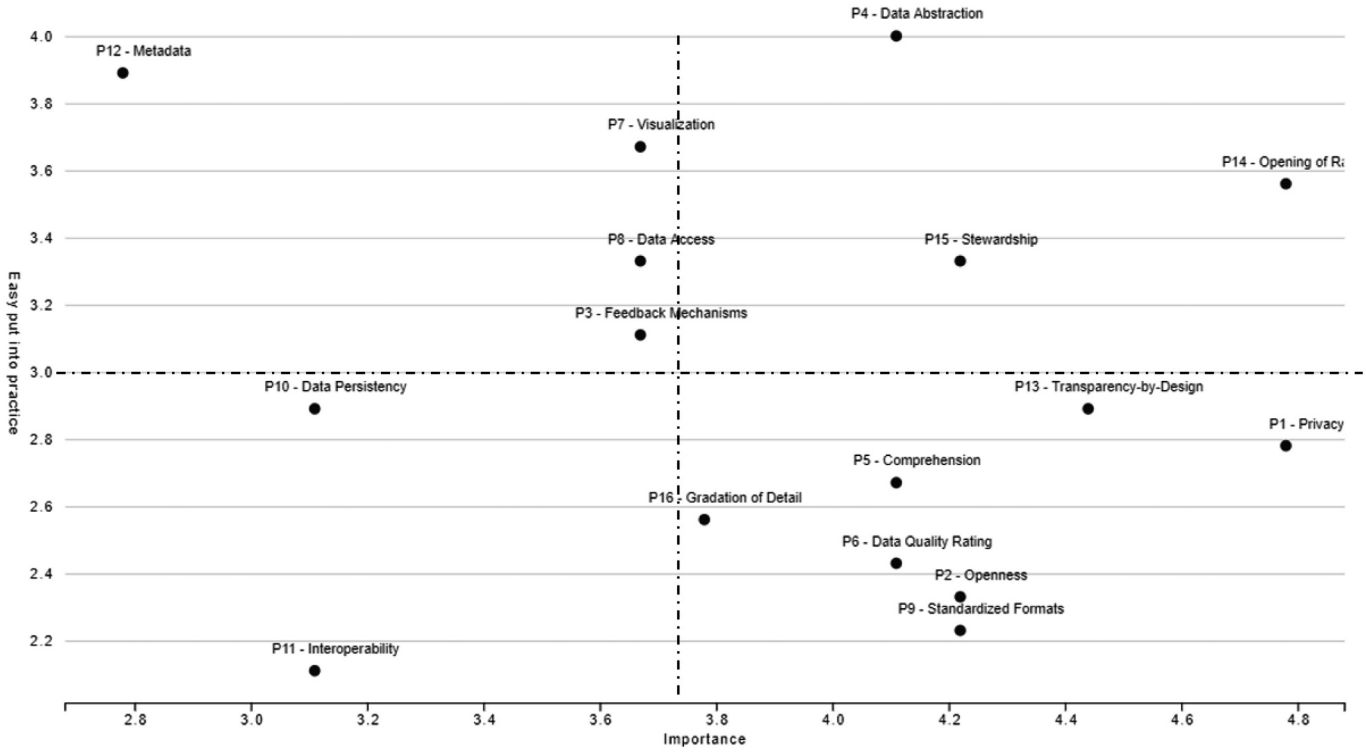

Fig. 3. The ease-of-use and the importance of design principles.

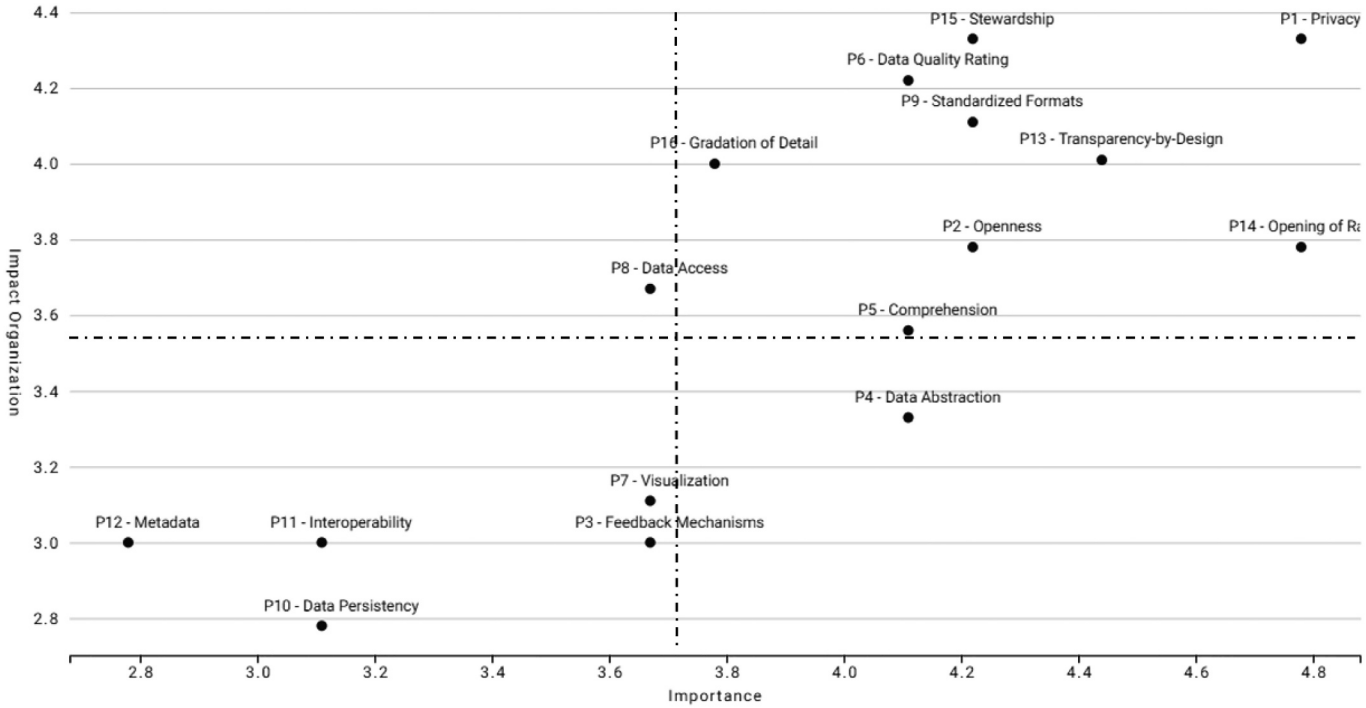

Fig. 4. The organizational impact and the importance of design principles.

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 Fig. 4 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.

6. Discussion

6.1. Do the design principles always result in digital transparency?

Disclosing data does not by itself result in digital transparency, accountability, or openness (Matheus & Janssen, 2015). 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 C.1 (Appendix C) 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.

6.2. Is full transparency possible or desired?

While full transparency is often viewed as impossible (Fung, Graham, & Weil, 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, 2011).

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.

Declaration of Competing Interest

The authors declare do not have any conflict of interest.

Acknowledgements

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.

Appendix A. List of papers containing barriers to digital transparency

Table A.1
List of papers containing barriers to digital transparency.

Paper ID	Source	Paper ID	Source
1	Sivarajah, Kamal, Irani, and Weerakkody (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, Krotov, and Bowman (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, Heuer, and Paula (2017)
9	De Laat (2017)	39	Matheus 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	M. Janssen and van den Hoven (2015)
13	Bello-Orgaz, Jung, and Camacho (2016)	43	John Carlo Bertot, Gorham, Jaeger, Sarin, and Choi (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)		

Appendix B. List of barriers to digital transparency

Table B.1
List of barriers to digital transparency.

#	Category	Code	Barrier	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.	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 in processing vast amounts of data	The vast amounts of data is a technical barrier for dealing 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

(continued on next page)

Table B.1 (*continued*)

#	Category	Code	Barrier	Description	Cite Count	Sources
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 to information sharing risks	Information sharing endangers privacy	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 do not 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 issues due to the lack of unified ontologies	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 a wide audience	A wider audience makes it 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 big data projects	11	1, 2, 3, 4, 5, 6, 18, 31, 41, 49, 50
10	Data Quality	DQ3	Data quality issues due to multiple types of data sources – unstructured vs structured datasets	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 to 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 issues due to the 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 issues due to re-identification combining datasets	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 plans	The organization has no information sharing plan or culture to help transparency and big data projects	9	1, 3, 4,11, 12, 13, 31, 42, 47
15	Organizational	OR2	Data quality issues due to 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 issues due to the 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 issues due to the 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 appropriate methods to deal with modern Big Data systems	Methods to deal with big data are still at the initial stage of development	7	1, 4, 5, 6, 29, 37, 42
19	Political and Legal	PL2	Data protection issues due to mass surveillance	Risk of big data for mass surveillance purposes	7	1, 6, 8, 17, 41, 42, 44
20	Technical	TE4	Technical issues from big data volumes creating data overload to users	A huge amount of data leading to data overload	7	7, 11, 13, 14, 15, 37, 42
21	Technical	TE5	Data quality issues due to the timeliness	Data is not accessed or published within the desired time	7	4, 30, 32, 37, 42, 43, 45
22	Organizational	OR3	Data quality issues leading to mistakes or misconducts	People make mistakes or misbehave 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 few free or low-cost analytical tools to deal with big data	5	2, 5, 6, 34, 46
24	Technical	TE6	Lack of performance due to the lack of big data system architecture plans	Organizations have no big data architecture plans, influencing on transparency-by-design	5	5, 6, 42, 43, 45
25	Technical	TE7	Performance issues due to bandwidth	There is no bandwidth available to perform big data projects	5	2, 5, 13, 14, 26
26	Technical	TE8	Security issues due to chances of leaking and hacking	Organizations are not prepared to prevent leaking or hacking	5	2, 5, 13, 14, 26
27	Economical	EC4	Lack of big data system architecture plans leading to unpredictable cost increases	Lack of or not well designed big data architectures leading to unanticipated additional costs	4	1, 6, 30, 35
28	Technical	TE9	Data quality issues due to the existence of data silos	Data silos influence big data and transparency	3	2, 42, 50
29	Data Quality	DQ7	Data quality issues due to language barriers such as the use of jargon	Language barriers such as jargons influence data quality, big data and transparency	3	3, 30, 42
30	Usage	US2	Data quality issues due to the overload of information	Information overload can lead to user mistake	3	2, 46, 49
31	Organizational	OR4	Lack of available data	No data is available	3	8, 18, 21
32	Human resources	HR4	Lack of skilled employees to lead big data projects	There are few people qualified to conduct big data projects and create transparency	3	1, 39, 42
33	Political and Legal	PL3	Legal issues due to the lack of stable regulatory frameworks	There is no stable regulatory framework for big data and transparency	3	2, 18, 33
34	Organizational	OR5	Asymmetry of information due to the lack of information sharing policies	Lack of an information-sharing policy leading to the asymmetry of information, influencing big data performance and transparency	2	2, 3
35	Technical	TE10	Lack of data accessibility	Data has a low level of accessibility	2	43, 45
36	Economical	EC5	Financial issues due to limited budgets	Organizations have limited budgets for big data and transparency	2	3, 4
37	Organizational	OR6	Lack of algorithmic openness	Algorithms used on big data are not transparent	2	1, 5
38	Organizational	OR7	Organizational issues due to the lack of awareness about big data possibilities	People are unaware of what benefits big data and transparency can bring to their organizations	2	5, 6
39	Ethical	ET1			2	8, 9

(*continued on next page*)

Table B.1 (*continued*)

#	Category	Code	Barrier	Description	Cite Count	Sources
			Prejudice due to harmful use of algorithms, e.g. discrimination of certain ethnic groups	Algorithms can have biases such as, e.g. discrimination of certain ethnic groups		
40	Ethical	ET2	Privacy issues due to human resource habits, ethics and culture	Culture influences bad habits that can lead to privacy issues	2	42, 44
41	Technical	TE10	Security issues due to the lack of log collection and forensic analysis	Organizations have no log collection to allow forensic analysis	2	1, 2
42	Technical	TE12	Technical issues due to the lack of user-friendly big data tools	Big data tools are not user friendly	2	31, 33

Appendix C. Design principles for digital transparency

Table C.1
Design principles for digital transparency.

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 (John 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 policymakers regarding the quality and use of the released data. Furthermore, data providers and policymakers 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.

(*continued on next page*)

Table C.1 (*continued*)

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.
	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 Matheus 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.

(*continued on next page*)

Table C.1 (*continued*)

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.
	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 (Matheus 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.

Appendix D. Interview Protocol Form

Introduction

You are selected as the respondent of this interview to contribute to 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 in order to achieve a level of transparency, principles should be considered in the creation of open data portals.

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 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 to push ahead and complete this line of questioning.

Section A – General information

1. What was your pilot?

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

Section B – Following Enterprise Architecture and Principles

2. I followed the Enterprise Architecture when creating applications ( ) Strongly disagree

( ) Strongly Disagree ( ) Disagree ( ) Neutral ( ) Agree ( ) Strongly Agree

3. I followed (one or more) Design Principles when creating applications ( ) Strongly disagree

( ) Strongly Disagree ( ) Disagree ( ) Neutral ( ) Agree ( ) Strongly Agree

Section C – Principle Questions.

This section contains questions about the 16 principles identified in Scientific Literature Review. You will be asked to evaluate each of the principles in the context of the OGI project. You are most welcome to recommend changes in the name and description of each principle. Below you have a figure with the list of Principles.

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

Section D – Principles Evaluation only Example Principle 1 Structure.

Principle 1 - Separating privacy and non-privacy sensitive data at the source
Name: Separating privacy and non-privacy sensitive data at the source
Statement: The essential requirement for transparency is determining the privacy level of the data. Without knowing whether the data contains privacy (including non-privacy 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 organizations should understand the impact of releasing data and find solutions for opening data that is sensitive or constrained.

[1] Do you agree with this Principle 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

[7] Do you want to explain your argument about importance? If yes, please explain below (not mandatory)

_____

[8] In practice, it is easy to implement Principle 1
( ) Strongly Disagree ( ) Disagree ( ) Neutral ( ) Agree ( ) Strongly Agree

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

_____

[10] What is your rating for Principle 1 in terms of Priority?
( ) Low Priority ( ) Somewhat Priority ( ) Neutral ( ) High Priority ( )Essential Priority

[11] Do you want to explain your argument about your rating in terms of priority? If yes, please explain below (not mandatory)

_____

[12] What is the impact of Principle 1 on the Organization?
( ) No Impact ( ) Minor Impact ( )Neutral ( ) Moderate Impact ( ) Major Impact

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

_____

References

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons, 60*(3), 285–292.

Al-Qirim, N., Tarhini, A., & Rouibah, K. (2017). *Determinants of big data adoption and success*. Computing and Systems: Paper presented at the Proceedings of the International Conference on Algorithms.

Amugongo, L. M., Nggada, S. N., & Sieck, J. (2016). Leveraging on open data to solve city challenges: A case study of Windhoek municipality. In *Paper presented at the Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on*.

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. https://doi.org/10.1111/1748-8583.12090.

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.

Bannister, F., & Connolly, R. (2011). The trouble with transparency: A critical review of openness in e-government. *Policy & Internet, 3*(1), 1–30.

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.

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. In *Paper presented at the Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review, 82*(5), 977–1008.

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.

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.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77–90. https://doi.org/10.1111/isj.12088.

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly, 31*(Supplement 1), S10–S17. https://doi.org/10.1016/j.giq.2014.01.003.

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.

Cumbley, R., & Church, P. (2013). Is "big data" creepy? *Computer Law & Security Review, 29*(5), 601–609. https://doi.org/10.1016/j.clsr.2013.07.007.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly, 27*(4), 377–383.

De Laat, P. B. (2017). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, 1–17.

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. https://doi.org/10.1007/s10796-016-9675-5.

Elragal, A. (2014). ERP and big data: The inept couple. *Procedia Technology, 16*, 242–249.

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.

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.

Frank, M., & Oztoprak, A. A. (2015). Concepts of Transparency: Open Data in UK Local Authorities. In *Paper presented at the proceedings of the international conference for E-democracy and open government 2015 (CeDEM15)*.

Fung, A. (2013). Infotopia: Unleashing the democratic power of transparency. *Politics and Society, 41*(2), 183–212.

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

George, G., Haas, M. R., & Pentland, A. (2014). *Big data and management*. In: Academy of Management Briarcliff Manor, NY.

Gil, D., & Song, I.-Y. (2016). Modeling and Management of big Data: Challenges and opportunities. *Future Generation Computer Systems, 63*, 96–99. https://doi.org/10.1016/j.future.2015.07.019.

Goëta, S., & Davies, T. (2016). *The daily shaping of state transparency: Standards, machine-readability and the configuration of open government data policies*.

González-Zapata, F., & Heeks, R. (2016). The influence of the transparency agenda on open government data in Chile. In *Paper presented at the 2016 Conference for E-Democracy and Open Government (CeDEM)*.

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.

Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism, 18*(4), 408–424.

Hardy, K., & Maurushat, A. (2017). Opening up government data for big data analysis and public benefit. *Computer Law & Security Review, 33*(1), 30–37.

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.

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), Article e1002333.

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., & 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.

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.

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.

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 Medical Informatics, 4*(4), Article e38. https://doi.org/10.2196/medinform.5359.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293–303. https://doi.org/10.1016/j.bushor.2017.01.004.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly, 32*(3), 323–332.

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*. Taylor & Francis.

Matheus, R., & Janssen, M. (2015). 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.), *Vol. 9373. Open and Big Data Management and Innovation* (pp. 236–246). Springer International Publishing.

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. (2020). A systematic literature study to unravel transparency enabled by open government data: The window theory. *Public Performance & Management Review (PPMR), 43*(3), 503–534.

Matheus, R., Janssen, M., & Maheshwari, D. (July 2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly, 37*(3), 101284.

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.

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.

Nograšek, J., & Vintar, M. (2014). E-government and organisational transformation of government: Black box revisited? *Government Information Quarterly, 31*(1), 108–118.

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.

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.*. https://doi.org/10.1016/j.jksuci.2017.06.001.

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.

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.

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.
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.
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. In *Paper presented at the proceedings of the 6th international conference on data science technologies and applications DATA-2017, INSTICC*. Madrid: ScitePress.
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.
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. https://doi.org/10.1016/j.ins.2014.01.015.
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. In *Paper presented at the conference on e-business, e-services and e-society*.
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.
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.
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. https://doi.org/10.1177/0952076716687355.
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.
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.
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.
Thiago, P., Heuer, V. D., & Paula, A. (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. https://doi.org/10.4018/IJDSST.2017010102.
TOGAF. (2009). The open group architecture framework (togaf). *The Open Group, 1*.
Ubaldi, B. (2013). *Open Government Data*.
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.
Wielki, J. (2013). Implementation of the big data concept in organizations-possibilities, impediments and challenges. In *Paper presented at the computer science and information systems (FedCSIS), 2013 federated conference on*.
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.
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.
Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.
Zakim, D., & Schwab, M. (2015). Data collection as a barrier to personalized medicine. *Trends in Pharmacological Sciences, 36*(2), 68–71.
Zicari, R. V. (2014). Big data: Challenges and opportunities. *Big data computing, 564*.

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 (The Netherlands). He was a lecturer at Rotterdam School of Management of Erasmus Rotterdam University (The Netherlands) teaching Data Science and Programming for Managers courses. He leads WPs in the CAP4CITY Project (www.cap4city.eu/) and led WPs in the H2020 OpenGovIntelligence project (www.opengovintelligence.eu) which aims to create transparency using open government data in six international governmental pilots.

Marijn Janssen is a full Professor in ICT & Governance and chair of the Information and Communication Technology research group of the Technology, Policy and Management Faculty of Delft University of Technology. His research interests are in the field of orchestration, infrastructures, and open and big data. He is Co-Editor-in-Chief of *Government Information Quarterly*, conference chair of IFIP EGOV series and is chairing mini-tracks at e-government and information systems conferences. He was nominated in 2018 and 2019 by Apolitical as one of the 100 most influential people in the Digital Government worldwide https://apolitical.co/lists/digital-government-world100. More information: www.tbm.tudelft.nl/marijnj.

Tomasz Janowski is the Head of the Department of Informatics in Management, Faculty of Economics and Management, Gdańsk University of Technology, Poland; Invited Professor at the Department for *E*-Governance and Administration, Faculty of Business and Globalization, Danube University Krems, Austria; and Co-Editor-in-Chief of Government Information Quarterly, Elsevier. Previously, he was Invited Professor at Università della Svizzera italiana, Switzerland and University of Minho, Portugal, and Head, Senior Research Fellow and Research Fellow at the United Nations University (UNU) units in Macau and Portugal. More information: https://pg.edu.pl/6260ee25a9_tomasz.janowski