Open Government Data Portal
Design Principles

Implementing Transparency, Privacy, and Information Quality by Design

Master Thesis

REVI FAJAR MARTA

Faculty of Technology, Policy, and Management

T̃U Delft

Open Government Data Portal Design Principles

Implementing Transparency, Privacy, and Information Quality Assurance by Design

Master thesis submitted to Delft University of Technology in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in Management of Technology

Faculty of Technology, Policy, and Management

By

Revi Fajar Marta

Student number: 4421639

To be defended in public on October 25, 2016

Graduation Committee

Chairperson : Prof. dr. ir. M.F.W.H.A. Janssen, ICT Section
First supervisor : drs. Jolien Ubacht, ICT Section
Second supervisor : Dr. M.L.C. de Bruijne, POLG Section
Additional supervisor : Dr. Anneke Zuiderwijk, ICT Section

This page is intentionally left blank

Acknowledgements

The process of conducting and finishing this master thesis report had been wonderful journey for me. Many difficulties were faced, but with the help of many people, I was able to go through them and finish this report and conclude my studies in the Master Management of Technology program of the Technology, Policy, and Management faculty of TU Delft. The field of open data was new to me, and exploring and understanding it was quite challenging. Thankfully, along the way I met many people who, with their knowledge and wisdom, guided and supported me through the whole process. Therefore, I would like to dedicate this chapter for the people who had supported me in finishing this master thesis report.

First of all, I would like to thank Dr. Anneke Zuiderwijk, my additional supervisor for this master thesis project. She is officially an additional supervisor, yet she was one of the people who helped me most during the initial stages of my research. Her help and guidance was invaluable to me in shaping the course of the research project. Furthermore, she also helped me during the difficult times I had by giving advices and encouragement that allowed me to continue my research through the many changes that had to be done in the course of the project.

Secondly, I want to show my gratitude to drs. Jolien Ubacht, my first supervisor in this research project. Jolien gave me many invaluable inputs and encouragement during the course of my research. One of the lessons I learned from Jolien is how to think and act systematically while paying attention to details. Her inputs are what shaped the final version of this master thesis report.

Next, I would like to thank Prof. Dr. Marijn Janssen, the chairperson of my graduation committee, for his invaluable advice when I had doubt during the course of the research project, and for his feedbacks for my master thesis report. I also thank Dr. Mark de Bruijne, the second supervisor of this research project, for his feedbacks and his inputs which allowed me to see different perspectives regarding the results of the research.

Furthermore, I would like to express my gratitude for the interviewees for the supports and the inputs which allowed me to finish this research project as well as my friends from the MoT program who always gave me encouragement.

Finally, I would like to thank my family, especially my parents for their supports and prayers, as well as my wife and children, who supported me unconditionally through everything and giving me a reason to go forward despite all the difficulties I had.

Revi Fajar Marta

Delft, October 2016

Abstract

Transparency, privacy, and information quality assurance are profound issues that often emerge due to the publication or use of open data. Realizing the three values are important, and the design of open government data portals should be able to accommodate the values. However, how the open data portals should be designed in order to accommodate the three values is still unclear. Existing principles of open government data focus on the data itself, rather than how open government data portals should be designed with regard to the values of transparency, privacy, and information quality. Therefore, there is a need for explicit design principles for incorporating these values into the design of open government data portals. This research does qualitative research on open government data portal design principles. The approach taken involves reviewing literatures to find out the dimensions of transparency, privacy, and information quality which are relevant to open government data portal design. Next, design requirements from open government data portals are obtained through interviewing five academic researchers from TU Delft while focusing on the dimensions of transparency, privacy, and information quality. Then, design observation is done on five selected open government data portals, focusing on the implementation of design requirements acquired in the previous phase of the research. Lastly, principles are formulated from the findings obtained from each phase of this research. The result of this research project is a set of design principles which serves as a guideline on how to design open government data portals that incorporate transparency and privacy while ensuring information quality.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

The movement on making governments more open has been adopted by many countries (Bertot, Jaeger, & Grimes, 2010). Local, regional, and national governments from all over the world have started to participate by making their data open to public. Data published by governments for public use are called open government data. These data are published through data portals owned by the governments. For example, the UK government publishes its data through its data portal data.gov.uk. The site contains thousands of datasets from various sectors such as transportation, environment, health, economy, etc. which are published by institutions under the UK government.

Governments publish data for certain reasons. Public data can be used in public policy development, and they have an important role as well in creating innovative products and services which can be useful for the citizens (Janssen, 2011). These innovative products and services ultimately can be used in generating social and economic values for the citizens (Dawes & Helbig, 2010). For example, public data can be used to create an application which displays real-time traffic information or a location finder service for many public infrastructures such as schools, public toilets, etc. (Janssen, 2011).

The other main reason for governments to publish their data is to create transparent governments. Data published by governments in open government data portals, such as data from transportation, health, and education sectors, can give citizens a perspective of what the government is doing. Furthermore, scholars believe that opening government data for use and reuse will lead to increased government transparency (e.g. Bertot, et al., 2010; Luna-Reyes, et. Al., 2014; Janssen, et al., 2012). Governments may seek to enhance transparency as it is beneficial for the governments themselves and the society. Janssen & van den Hoven (2015) argue that transparency is an important component for establishing a democratic society. Government transparency allows the public to have an oversight of their governments and allows citizens to keep the governments accountable for their actions or their lack of actions (Janssen & van den Hoven, 2015).

1.1. Research problem and research objective

Opening government data may cause undesired effects. In publishing government data, breach of privacy may become a concern (Conradie & Choenni, 2014; Zuiderwijk & Janssen, 2014). Data owned by governments may contain information of individuals. When this information is published through open government data, the privacy of those individuals may be violated. Although governments may have treated the data before publishing them on open government data portals to ensure that no personal data are published, there may be certain situation where the use of government data can lead to privacy issues. For example, combining different datasets from open government data portals may increase the granularity of the data and lead to privacy issues. Therefore, in publishing government data, it is important that the privacy of individuals are taken into account.

Another issue concerning open government data is that the act of publishing government data does not automatically translate to increased transparency (Janssen, Charalabidis, & Zuiderwijk, 2012). With more data published in a government's data portal, finding the right information from the abundant datasets may become difficult, and instead of increasing transparency, the open government data may result in less transparency (Janssen & van den Hoven, 2015). A research by Lourenço (2015) showed that most open data portals do not actually lead to increased transparency.

One of the causes for this is the lack of information quality assurance mechanisms in publishing the data (Lourenço, 2015). This may result in the open government data having low quality, e.g. the information provided may not be complete or it may not be relevant and thus has low value and usability. Information quality also affects public's trust in governments. Publishing low quality data may tarnish the reputation of the agency who publishes the data as well as harm public's trust on the agency (Lee & Kwak, 2011). Therefore, it is imperative that governments publish high quality data.

Based on the discussion above, it can be concluded that in the effort of increasing government transparency through open government data, the preservation of individual's privacy as well as high information quality must be ensured. The open government data portals, in particular, should be able to support the value of transparency, privacy and information quality assurance from the design level. In other words, open government data portals should be designed and developed while taking into account the value of transparency, indvidual's privacy, and quality of information.

However, there are no clear design principles yet on how to include the value of transparency, privacy, and information quality into the design of open government data portals. In 2015, the Open Government Partnership, the international organization focused on promoting open government initiative, launched the International Open Data Charter, which contains six principles in ensuring that data released by governments is open and thus leads to transparency and government accountability (Principles, 2016). Open government advocates have also formulated eight principles in publishing open government data, which also promotes transparency (The 8 Principles of Open Government Data, 2016). However, all the mentioned principles are focused on the data itself and do not describe the way open government data portals should be designed in order to incorporate the value of transparency, privacy, and information quality assurance.

The objective of this research is to formulate principles for designing open government data portals which accommodate transparency, privacy, and information quality assurance. This research will deliver a set of design principles which can be used by governments and portal developers in designing open government data portals which takes into account the issue of transparency, privacy, and information quality assurance.

1.2. Knowledge gap and problem statement

From the previous section, a gap in the body of knowledge regarding open government data can be identified. This gap refers to the lack of clearly defined design principles for open government data portals which emphasize the value of transparency, privacy, and information quality assurance. The lack of explicit design principles for open government data portals can cause difficulties in designing (or redesigning) open government data portals as clear and well-defined principles are required by system designers and system architects in guiding the design and redesign of information system architectures (Janssen & van den Hoven, 2015; Richardson, Jackson, & Dickson, 1990).

This gap can be overcome by defining principles for designing open government data portals. Explicit design principles can help system designers and architects to design and redesign information systems with high degree of consistency (Richardson, et al., 1990). By having clear and well-defined principles for guiding the design and redesign of government data portals, the value of transparency, privacy, and information quality assurance can be implemented consistently throughout the government data portals. Therefore, there is a need for explicit principles in guiding the design of

open government data portals so that the three values (i.e. transparency, privacy, and information quality assurance) can be accommodated by the open government data portals.

1.3. Practical and scientific relevance

This research has both practical relevance and contribution to science. From a practical perspective, this research will contribute to the effort of designing open government data portals that are able to support the value of transparency, privacy, and information quality assurance by providing design principles which acts as guidelines in designing open government data portals. The guidelines provided by the open government data portal design principles can be used by both policy makers or governments and open government data portal designers in order to ensure that the portals that they develop can accommodate the values of transparency, privacy, as well as information quality assurance.

Open government data can provide benefits to both the governments who publish their data and their citizens. Identified benefits from open government data include the political and social benefits such as government accountability and creation of public's trust in government, economic benefits such as economic growth and contribution towards the improvement of processes, product, or services, and operational and technical benefits such as improvement of public policies (Janssen, et al., 2012). In a grander scope, this research will contribute to the efforts of governments to gain these benefits.

From a scientific perspective, this research will contribute to the current scientific studies on open government data. Currently, there is a need for both fundamental and applied open data research, especially related to the conceptualization of transparency and privacy in open data (Janssen & van den Hoven, 2015). A knowledge gap regarding how open government data portals should be designed wih regard to transparency, privacy, and information quality assurance currently exists. This research aims to fill this knowledge gap and contribute to open data research in general by delivering the design principles for open government data portals.

1.4. Research questions

In the process of achieving the research objective, the formulation of research questions are required. Answering research questions will lead to the knowledge required in reaching the objective of a research (Verschuren & Doorewaard, 2010).

This research will focus on the following main research question:

RQ: What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance?

In order to answer the main research question above, three sub research questions must first be answered. The answers to these sub-questions contribute to the knowledge needed in answering the main research question.

The three sub-questions are:

SQ1: What are the dimensions of transparency, privacy, and information quality that are relevant to open government data portal?

This research question focuses on exploring the definition of transparency, privacy, and information quality as well as their dimensions. Literature review will be used in exploring and answering the research question.

SQ2: What design requirements of open government data portals are relevant to transparency, privacy, and information quality?

Open government data portal design requirements define what features and functionality an open government data portal should provide. This knowledge can be obtained through interviewing experts in open data research.

SQ3: How do existing open government data portals implement the design requirements of open government data portals related to transparency, privacy, and information quality?

Open government data portal design observation provides insight into the existing design of the portals. This insight may help in formulating the design principles of open government data portals. The observation will evaluate whether the design requirements of open government data portals obtained from sub-question 3 are implemented in existing open government data portals.

Research instruments are required in order to be able to answer the research questions above. In a summary, the research instruments used for answering each sub-question can be observed in Table 1.

Table 1. Research questions and research instruments

Research questions	Research instruments
SQ1: *What are the dimensions of transparency, privacy, and information quality that are relevant to open data portal design*?	Literature review
SQ2: *What design requirements of open government data portals are relevant to transparency, privacy, and information quality?*	Interview
SQ3: *How do existing open government data portals implement the design requirements of open government data portals related to transparency, privacy, and information quality?*	Observation of open government data portals
RQ: *What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance?*	Synthesis of findings from interview and design observations

1.5. Research approach and methodology

This research is an exploratory research aimed at formulating principles for designing open government data portals which incorporates transparency, privacy, and information quality assurance. The principles are extracted from solutions or experiences in dealing with transparency, privacy, and information quality problems in open government data. As it is exploratory and open ended in nature, this research can be categorized as a qualitative research.
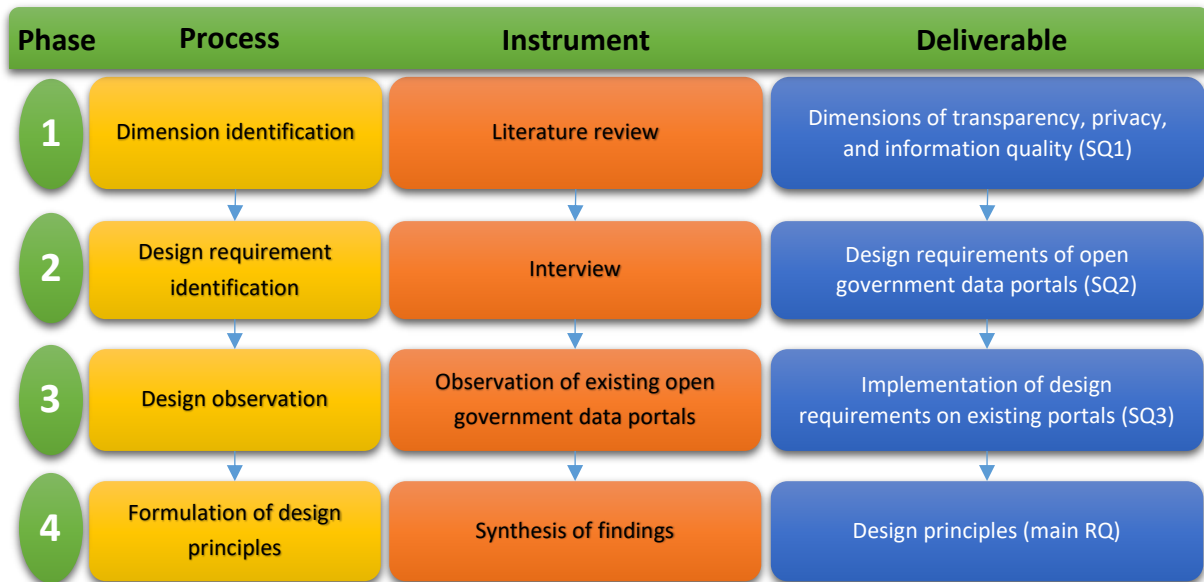
Phase	Process	Instrument	Deliverable
1	Dimension identification	Literature review	Dimensions of transparency, privacy, and information quality (SQ1)
2	Design requirement identification	Interview	Design requirements of open government data portals (SQ2)
3	Design observation	Observation of existing open government data portals	Implementation of design requirements on existing portals (SQ3)
4	Formulation of design principles	Synthesis of findings	Design principles (main RQ)

Figure 1. Research approach

This research adapts the methods used by several scholars in order to arrive to the design principles for open government data portals. Bharosa & Janssen (2015) used interviews in order to formulate design principles. Dimensions are used in order to be able to focus on the problems presented to the interviewees. A dimension can be defined as an aspect or attribute which belongs to a concept. For example, the dimensions of information quality may include accuracy and timeliness. Lourenço (2015) performs open government data portal observation using design requirements as the focus of the observation. The design requirements are developed from characteristics or dimensions of transparency in literatures. Meanwhile, Fedorowicz, et al. (2014) used a multiple case study research design to arrive at design observations. Design observations describe the current condition of an architecture and provide insight into existing architectural or social conditions (Fedorowicz, et al., 2014). Furthermore, Fedorowicz, et al. (2014) also argue that design principles can arise from design observations.

In this research, design principles are derived from design observations as well as expert interviews (Figure 1). The expert interviews provide design requirements that will be the focus of the observations on existing open government data portals. The interview will also be used to obtain insight from experts on open data on how open government data portals should be designed with regard to transparency, privacy, and information quality. Literature review provides the dimensions of transparency, privacy, and information quality which are presented to the interviewees in order to obtain relevant design requirements.

There are four phases involved in this research:

1. **Phase 1: dimension identification**

    The first phase of this research is to identify the dimensions of transparency, privacy, and information quality. The identification of these dimensions is important to what elements of transparency, privacy, and information quality are relevant to open government data portal design. The identification of dimensions of transparency, privacy, and information quality also allows the identification of design requirements, which is done in the next phase, to be focused on the aspects of transparency, privacy, and information quality that are relevant to open government data portal design. The identification of dimensions is done by exploring existing

literature regarding transparency, privacy, and information quality. Exploration on existing literature on open government data and design principles is also done to provide definition on the concepts discussed in this research as well as gain understanding on how design principles should be defined. The deliverable of the first phase is a set of dimensions of transparency, privacy, and information quality which are relevant to open government data portal design.

2. **Phase 2: design requirement identification**

    The second phase of this research takes the output from the first phase and process it using different instrument. The dimensions of transparency, privacy, and information quality obtained from the previous phase are used as a starting point to explore the design requirements of open government data portals. Design requirements define what functionalities an open government data portal should or should not have. Having a set of design requirements can help in determining how an open government data portal should be designed.

    Interviews are used to gain insight from researchers of open data regarding the design requirements of open government data portals. Interview is selected as the method to gather data for this phase as it allows focus on the goal of the phase, which is identifying design requirements, and it can provide insight through the explanation from the interviewees (Yin, 2003). The second phase of this research will deliver design requirements of open government data portals and answer the second sub-question of the research. The findings from the interviews, including the design requirements, will be used as inputs for the formulation of the design principles.

3. **Phase 3: design observation**

    The third phase of this research uses the design requirements obtained from the previous phase. This phase involves observing how existing open government data portals implement the design requirements of open government data portals. By confronting the perspective of the interviewees with the reality on the field, the observation of existing open government data portals can provide additional insight by reinforcing the findings from the interviews or providing new findings. In order to observe existing open government data portals, a selection of portals first should be done. The deliverable of this phase is a set of findings from the observation of existing open government data portals. The result from this phase will contribute as an input for the formulation of design principles.

4. **Phase 4: Formulation of design principles**

    The formulation of design principles involves synthesizing the findings from the previous phase of this research. The end result and deliverable of this phase is a set of design principles of open government data portal which accommodates transparency, privacy, and information quality.

Further explanation regarding the methodology used in this research is presented below.

1.5.1. Literature review

Literature review can be defined as "the selection of available documents (both published and unpublished) on the topic, which contain information, ideas, data, and evidence written from a particular standpoint to fulfill certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed" (Hart, 1998, p. 13). Following Hart's (1998) definition, the literature review

used in this research aims to gain understanding about the topic of the research and provide theoretical background for the research, as well as to find out the views of scholars with regard to the topic being discussed. With regard to the second aim, in particular, the literature review is used to explore the views of scholars regarding the dimensions of transparency, privacy, and information quality and evaluate which dimensions are relevant to open government data portal design.

Due to the exploratory nature of the research, it is important to first have an understanding about the concepts discussed in this research, including open government data, open government data portals, design principles, as well as transparency, privacy, and information quality. Therefore, literatures related to open government data, open government data portals, design principles, as well as transparency, privacy, and information quality are examined and analyzed. The literature review is also used to explore the dimensions of transparency, privacy, and information quality that are relevant to open government data portals and this will provide the answer to sub-question 1 of this research.

In order to be effective, a literature review should follow a systematic process. Levy and Ellis (2006) argued that a literature review should follow three stages in order to be effective: 1) Input, 2) Processing, and 3) Output.

1. Input

   In searching for relevant literatures, this research will use literature databases such as Scopus, Google Scholar, and TU Delft Repository as the main sources of literatures. The keywords used in searching the literature databases include "open government data," "open government data transparency," "open government data privacy," "open government data information quality," and "design principles." In addition to searching for literatures directly in the literature databases, a backward and forward search is also conducted on the found literatures. The backward search consists of searching for relevant articles from the references found in the literatures, as well as searching for previous publications by the author. The forward search includes searching for additional articles that cites the original articles as well as articles that are published by the author after the publication of the original articles.

2. Processing

   This stage is about the processing of articles that have been found in the input process. Relevant articles are read and important materials from the articles are extracted and put in a note. Similar concepts from different articles are clustered together. Therefore, it will help in understanding the concepts and compare different findings from different articles. Different articles read are also critically assessed and confronted with each other in order to derive a conclusion regarding the articles.

3. Output

   In this stage, conclusions from the findings in the previous stage are drawn. The conclusions are put together into paragraphs by first defining the logical structure of the findings. The result of this stage is the theoretical background section of the research document.

1.5.2. Interview

The purpose of the interview is to find out what open government data portal design requirements are correlated to the dimensions of transparency, privacy, and information quality. A set of dimensions of transparency, privacy, and information quality is presented to the interviewee and the

interviewee will be asked questions regarding the design requirements of open government data portals relevant to the dimensions.

The complete interview protocol used in this research, including interview structure and questions, is provided in the Appendix section of this report.

1.5.3. Open government data portal observation

The observation of open government data portals is conducted in order to gain insight on how existing open government data portals implement the design requirements obtained from the interview. The open government data portals selected for the design observation are taken from the cases provided by the 2016 Open Government Course MOOC participants. Instead of an in-depth observation into one portal, this research conducts investigation into many open government data portals. Conducting studies on multiple units of analysis can have stronger generalizability than a single case study, and the overall study can be regarded as more robust (Herriott & Firestone, 1983; Yin, 2003).

In order to be able to select a number of portals, a list of open government data portals for selection has to be obtained first. The 2016 Massive Open Online Course (MOOC) on Open Government held on edX platform discusses cases related to open government data. The course, run by TU Delft, invited its participants to discuss cases regarding open government. The participants of this course came from various countries, and they were each required to submit an open government cases from their own country. One of the focuses of the cases is the sharing of data and information by governments, which was relevant for this research. For example, a participant from Belgium submitted a case regarding data and information sharing by the Belgium government through its data portal. The participant provided a short description of the case, including the objective of the data sharing by the Belgium government, stakeholders involved, technology used, geographical coverage of the case, and a link to the open government data portal discussed in the case. Therefore, the cases submitted by the MOOC participants can be used to obtain a list of open government data portals from various government levels around the world.

1.5.4. Data analysis

Data analysis is done with the help of ATLAS.ti. ATLAS.ti is a software tool commonly used in qualitative research and qualitative data analysis. The software helps in analyzing qualitative data and provides assistance in coding. As this research is a qualitative research and gathers qualitative data through interview, the use of ATLAS.ti helps in conducting the analysis.

The features of ATLAS.Ti include the ability to code from text, image, video, and audio files. The tool also provides the ability to group codes into code families to help in the categorization of codes, write notes and memos about an idea obtained from the materials, as well as building network views to visualize the correlation between codes and gain insight from it. Therefore, primarily, ATLAS.Ti will be used in the coding process to make the process easier and to be able to organize the codes more easily. The process of data analysis can also be done without the help of ATLAS.Ti and it should end in the same result as when it is done with the help of ATLAS.Ti. However, the use of ATLAS.Ti can greatly help in speeding up the process and make it easier.

1.6. Research design

This research consists of several activities which are required in order to reach the objective of the research which are derived from the phases involved in the research approach. Some other research

activities are required for answering the research questions (e.g. literature review). Furthermore, most of the activities are sequential, which means that an activity must be finished first before it can continue to the next activities. Some activities, however, are parallel and can be conducted at the same time with little effect to each other. The logical flow of the combination of the research activities makes up the design of this research. In a diagram, the research design can be observed in Figure 2.
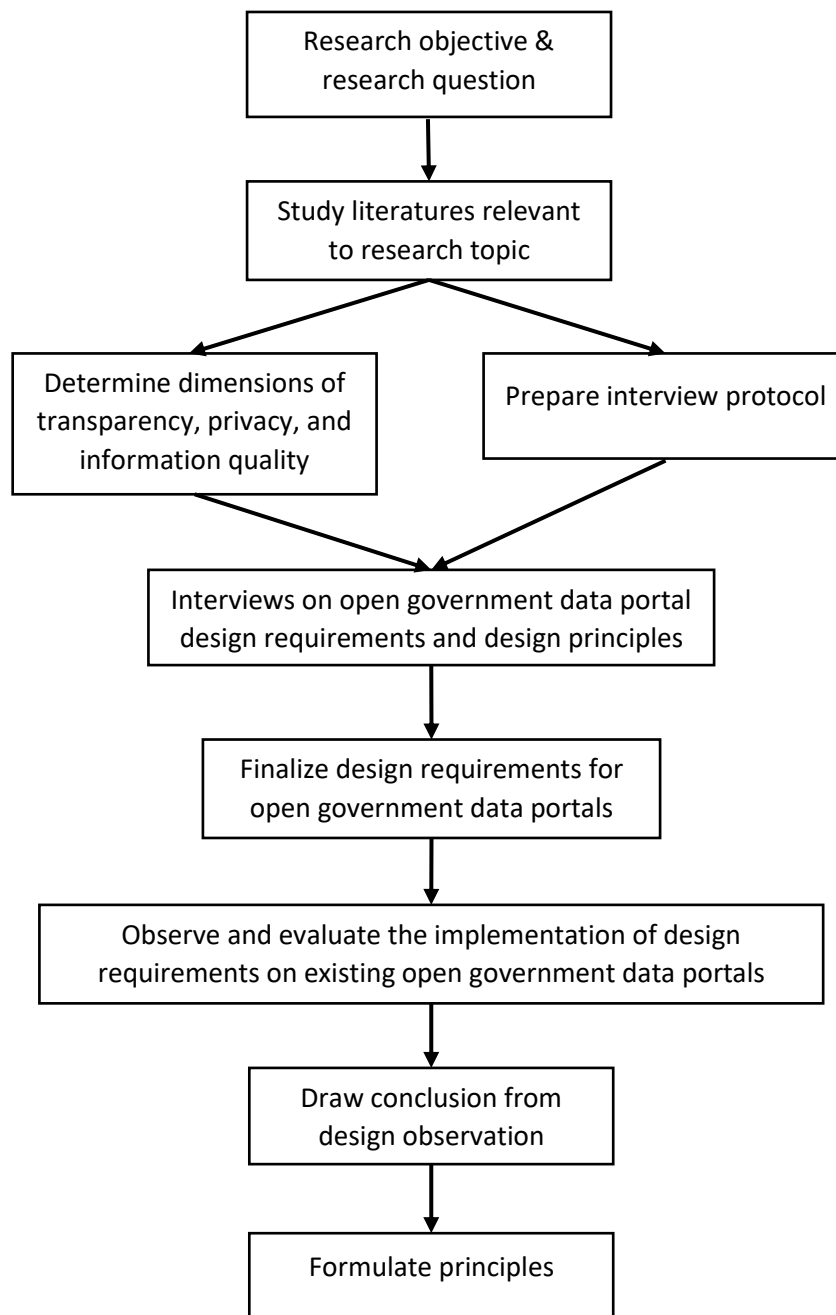
Figure 2. Research design

Chapter 2
Theoretical Background

The theoretical background section will discuss the theory regarding open government data, design principles, as well as transparency, privacy, and information quality. It is the first phase of the research and serves as a foundation to understand the problem being dealt within this research. Furthermore, this chapter will also answer the first sub-question of this research, which is:

"What are the dimensions of transparency, privacy, and information quality that are relevant to open government data portal?"

The deliverable of this chapter will be a set of dimensions of transparency, privacy, and information quality which are relevant to open government data portal. The dimensions will be used as the starting point for identifying open government data portal design requirements which will be discussed in the next chapter.

In summary, the process, instrument, and deliverable of the second phase of the research can be observed in the following figure.

Phase	Process	Instrument	Deliverable
1	Dimension identification	Literature review	Dimensions of transparency, privacy, and information quality (SQ1)
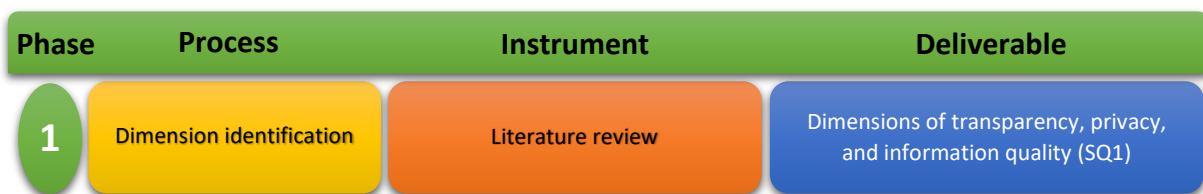
Figure 3. Phase 1 of the research

2.1. Open government data

Before one can understand what open government data is, one should have an understanding on the term "data" (Zuderwijk, 2015). Data can be defined as "symbols that represent properties of objects, events, and their environments" (Ackoff, 1989, p. 3). One of the most common models which helps in understanding data is the Data, Information, Knowledge, Wisdom (DIKW) model. According to the model, which is usually represented by a pyramid, data is at the lowest level. Information is gained from the interpretation of data, and knowledge is gained from information. Wisdom lies on the top level of the pyramid. This model suggests that data can be used to generate information, knowledge, and understanding or wisdom (Ackoff, 1989; Zuderwijk, 2015).

There are several existing definitions regarding open data and open government data. Lindman, Kinnari, and Rossi (2014, p. 740) define open data as "data, which is legally accessible through the Internet in a machine-readable format." The Open Knowledge Foundation (2016) describes that data is open "if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness." Geiger and von Lucke (2012, p. 269) define open government data as "all stored data of the public sector which could be made accessible by government in a public interest without any restrictions for usage and distribution." Meanwhile, Zuiderwijk (2015, p. 43) define open government data as the following:

"Open Government Data are structured, machine-readable and machine-actionable data that governments and publicly-funded research organizations actively publish on the internet for public reuse and that can be accessed without restrictions and used without payment."

This research will use the definition of open government data from Zuiderwijk (2015) since it has the most comprehensive definition of open government data. Furthermore, the definition offered by Zuiderwijk (2015) is built on the common elements of other open government data definitions. These commmon elements include: 1) open government data are collected and published by governments and publicly-funded research organizations, 2) open government data are published on the internet for reuse by public, 3) open government data are provided without restrictions and are provided free of charge, and 4) open government data are structured and machine-readable (Zuderwijk, 2015).

2.2. Open government data portals

Open government data are published on the internet for public reuse through open government data portals. Examples of open government data portals include data.gov.uk which belongs to the UK government and data.gov which belongs to the US government. These sites host datasets from various sectors which are created by various institutions under the governments. The Merriam-Webster dictionary defines a portal as "a site serving as a guide or point of entry to the World Wide Web and usually including a search engine or a collection of links to other sites arranged especially by topic" (Merriam-Webster, 2016). According to Lourenço (2015), open data portals act as centralized points of access to government data. Meanwhile, Krishnamurthy & Awazu (2016, p. 669) describe data.gov, a data portal owned by the United States government, as "an online portal to share data from various public agencies, which can be downloaded and manipulated for public use and re-use". From these definitions, it can be concluded that an open government data portal is a website which serves as a centralized point of access to open government data, which can be downloaded and manipulated for public use and reuse (Lourenço, 2015; Krishnamurthy & Awazu, 2016).

2.3. Design principles

Principles are often used in designing a system or an architecture. There is no common definition for principles, and scholars have different opinions regarding the definition of principles (Bharosa & Janssen, 2015). Albert Cherns (1976), who was one of the first scholar to use principles in designing socio-technical system (Bharosa & Janssen, 2015), describes the use of principles as a "checklist, not a blueprint" (p. 785). Housel, Sawy, and Donovan (1986, p. 396) define principles as "generic prescriptions for the design and implementation of information systems." Gilb (1997, p. 1) defines principles as "rules of thumb that guide the choices and actions of engineers." Richardson, Jackson, and Dickson (1990, p. 389), define principles in designing an information technology architecture as "an organization's basic philosophies that guide the development of the architecture." The Open Group defines principle as "general rules and guidelines, intended to be enduring and seldom amended, that inform and support the way in which an organization sets about fulfilling its mission" (TOGAF, 2011). Meanwhile, Bharosa and Janssen (2015, p. 472) define principles as "normative, reusable and directive guidelines, formulated towards taking action by the information system architects."

Although scholars define principles differently, there are some similarities in the definition of principles that they offer and how they use principles. All definitions show that principles are used as

a guideline in designing an artefact. Principles are similar to rules and laws in that they give directions on how to act. However, unlike rules and laws, principles are normative and prescriptive in nature and not as rigid as rules or laws (Bharosa & Janssen, 2015; Gilb, 1997). Due to its characteristics, principles are useful in designing an information system as they guide and give direction without being restrictive. Therefore, based on the characteristics and the existing definitions of design principles, this research defines open government data portal design principles as general rules and guidelines in designing and developing open government data portals.

2.4. Transparency, privacy, and information quality

This subsection discusses the concepts of transparency, privacy, and information quality which are used as the focal point in formulating the open government data design principles. Each of the concepts has dimensions, which can be defined as the aspects or properties which belong to the concepts. For example, one of the aspects of information quality is completeness, which is the "degree to which a given data collection includes data describing the corresponding set of real-world objects" (Batini, Cappiello, Francalanci, & Maurino, 2009). Thus, completeness is a dimension of information quality.

2.4.1. Transparency

The concept of transparency cannot be separated from open government data. Achieving a transparent government is one of the main reasons for governments to publish their data (Dawes & Helbig, 2010; Linders & Wilson, 2011; Lourenço, 2015). Furthermore, transparency is often associated with government accountability. Janssen and van den Hoven (2015) argue that transparency acts as a balance and control mechanism that is needed for a democratic society to function properly and that it allows governments to be held accountable for their actions.

There is no common definition of transparency, and scholars define the term "transparency" differently. Piotrowski and Van Ryzin (2007) define governmental transparency as "the ability to find out what is going on inside a public sector organization through avenues such as open meetings, access to records, the proactive posting of information on Web sites, whistle-blower protections, and even illegally leaked information" (p. 308). Meijer, Curtin, and Hillebrandt (2012) argue that transparency is one of the two elements of openness (of a government), the other being participation. According to Meijer, et al. (2012), transparency refers to the ability of "observing government decision-making processes" (p.13) while participation refers to "the opportunity to participate in these decision-making processes" (p. 13). However, other scholars may not conform to this definition of openness, and use the term "openness" and "transparency" differently. For example, Bauhr and Grimes (2012) and Kim (2009) put openness as an element of transparency, instead of the other way around. Despite the differences in definition, there is a common conception that in a transparent government, citizens should be able to know what the government is doing.

This research has an objective to derive design principles which emphasizes transparency for open government data portals. For this research, the most relevant method in achieving transparency is through the release of open government data to public. Therefore, based on the relevance of open government data in achieving transparency and the conception that transparency is about knowing what the government is doing, this research defines government transparency as "the ability to find out what is going on inside a government by observing data published by the government through open government data portals".

Scholars have different views regarding the dimensions of transparency. Bauhr & Grimes (2012) mentions three dimensions of transparency which can be used in measuring the transparency level of a government: government openness (access to information), whistleblower protection, and publicity. Meanwhile, in measuring transparency, Hollyer, Rosendorff, & Vreeland (2014) focuses on the specific dimension of governments' collection and dissemination of aggregate data. Kim (2009), on the other hand, in regard to governance by local government, describes that transparency has three dimensions: openness, participation, and integrity.

Bauhr & Grimes (2012), Kim, and Matheus & Janssen (2015) all mention openness as one of the dimensions of transparency. The openness of a government in an open data context is related to the publication of open government data. This is reflected in the definition used by Bauhr and Grimes (2012) who mention the publication of government data as part of openness of a government. In this case, the dimension of transparency mentioned by Hollyer, et al. (2014), which is the collection and dissemination of aggregate data by governments, also has a similar context with the dimension of openness mentioned by Bauhr & Grimes (2012), Kim (2009), and Matheus & Janssen (2015). The dimension of participation mentioned by Kim (2009) is related to the public involvement in the decision making process. Meanwhile, the accountability dimensions mentioned by Matheus & Janssen (2015) is about how governments can be held accountable for their actions or inactions.

Table 2. Dimensions of transparency according to scholars

Author	Transparency dimensions
Bauhr & Grimes (2012)	• Government openness (access to information) • Whistleblower protection • Publicity
Hollyer, Rosendorff, & Vreeland (2014)	• Governments' collection and dissemination of aggregate data
Kim (2009)	• Openness • Participation • Integrity
Matheus & Janssen (2015)	• Accountability • Openness

However, some of the dimensions of transparency mentioned by the scholars above have little relevance to open government data context. Whistleblower protection (Bauhr & Grimes, 2012) is more about the internal affair context of a government. Publicity (Bauhr & Grimes, 2012) is related to whether the media is able to publish negative reports or news regarding a government. Integrity (Kim, 2009) is related to the quality of public officials in managing themselves so that their performance is not influenced by the obligations to external individuals or organizations. These dimensions, while relevant to the broader context of transparency, have little to do with open government data.

2.4.2. Privacy

Privacy is often discussed in open data literature. However, there is no common and clear definition of the concept of privacy itself. Some literatures relate privacy with keeping personal identities from being disclosed (Conradie & Choenni, 2014; Choenni, et al., 2010; Janssen, et al., 2012). Janssen and van den Hoven (2015, p. 363) writes that "privacy is our right to live our lives without any external involvement." Based on those definitions and in relation to the release of government data to public, this research defines privacy as "the state of being free from identity disclosure due to the release or use of open government data." Several authors identify the dimensions of privacy in the form of privacy concerns. These dimensions of privacy concerns, although not directly being the elements of privacy, can be used to identify the aspects of open government data that may affect people's privacy.

Table 3. Dimensions of privacy concerns according to scholars

Author	Privacy concern dimensions
Smith, Milberg, & Burke (1996)	- Collection - Unauthorized secondary use (internal) - Unauthorized secondary use (external) - Improper access - Errors - Reduced judgment - Combining data
Borgesius, Eechoud, & Gray (2015)	- Chilling effects - Lack of control over personal information - Social sorting and discrimination
Martínez-Ballesté, Pérez-Martínez, & Solanas (2013)	- Identity privacy - Query privacy - Location privacy - Footprint privacy - Owner privacy

Smith, Milberg, & Burke (1996) identify the dimensions of concerns for information privacy as: collection, unauthorized secondary use (internal), unauthorized secondary use (external), improper access, errors, reduced judgment, and combining data. These dimensions of privacy concern are related to how data is being handled using an information system. Among these dimensions, combining data is the most relevant to open government data, as it is related to the use of published data. However, the other concerns mentioned by the author are not as relevant with open government data. The concern of collection and errors are more related to the collection phase of open government data, instead of the publication phase. As such, it has little relevance with the publication of open government data on open government data portals. The unauthorized secondary use of data (both internal and external), like combining data, is also related to the use of published data. However, in open data context, the purpose of data use should be left to the users of data (The 8 Principles of Open Government Data, 2016). The same case is also applicable to the

concern of improper access. According to the 8 principles of open government data, data should be available to the widest range of users (The 8 Principles of Open Government Data, 2016). Therefore, the concern of unauthorized use of data and improper access both have little relevance with open government data. Reduced judgment is related to the concern about excessive automated decision making processes and also has little relevance with open government data.

Borgesius, Eechoud, & Gray (2015) identifies three dimensions of privacy concerns, which are chilling effects, lack of control over personal information, and social sorting and discrimination. The lack of control over personal information is related to the concern that people will lose control over their personal information when it is released as part of open government data. Social sorting and discrimination is related to the concern that people can receive different treatment according to the information that other people have about them. These concerns are related to the use of published data, and as such, is relevant with open government data context. Chilling effects, however, is more related to the collection of data as it is about the reluctance that people have in interacting with public sector agencies due to the fear that their information will be stored (Borgesius, et al., 2015). Therefore, it has little relevance with the publication of open government data on open government data portals.

The dimensions of privacy concern identified by Smith et al. (1996) and Borgesius, et al. (2015) are related to the publication of data. However, there also privacy concerns that can arise due to the use of an information system. In the context of this research, privacy concerns can arise from the use of open government data portals. Martínez-Ballesté et al. (2013) identified five dimensions of privacy concerns that are related to the use of an information system: identity privacy, query privacy, location privacy, footprint privacy, and owner privacy. Although they are originally used in identifying privacy concerns in smart city context, some of the dimensions are relevant with the context of open government data portal as well. In the context of the use of open government data portals, the identity privacy concern is related to the concern that a user of an open government portal's identity can be disclosed and his or her activities on the open government data portal correlated with his or her identity. Query privacy concern is related to the possibility that the queries made by users of an open government data portal allows the users to be profiled based on the queries. Location privacy concern is related to the possibility that the location of an open government data portal's location can be disclosed. The other two dimensions of privacy concern mentioned by Martínez-Ballesté et al. (2013), footprint privacy and owner privacy, however has little relevance in the context of the use of open government data portals. This is because the two dimensions are specific to the context of smart city implementation.

2.4.3. Information quality

In the effort of achieving government transparency by publishing government data, information quality plays a significant role. Behkamal, et al. (2014) argue that the success of open data initiatives depend strongly on quality of released datasets. Information quality assurance is one of the necessary factors in conceptualizing transparency and privacy (Janssen & van den Hoven, 2015). Information quality is also an important factor in building public trust on the government. Low quality of published data can harm the reputation of government agencies who publish the data and decrease public's trust in the agencies (Lee & Kwak, 2011). Furthermore, the reuse of open government data requires that the users of the data trust that the datasets released has sufficient

quality (O'Hara, 2012). If the datasets are of low quality, data users will be reluctant in using them and thus the datasets will become less useful.

Information quality has many dimensions. Batini, et al. (2009) mention that among the these dimensions, accuracy, completeness, consistency, and timeliness are the most basic dimensions of information quality and often become the focus of other scholars. Strong, Lee, and Wang (1997) grouped information quality dimensions into four categories: 1) intrinsic, 2) contextual, 3) representational, and 4) accessibility. Intrinsic information quality means that the information has quality in its own right. Contextual information quality refers to the fact that the context of use of information determines the information quality required and that information must be relevant, timely, complete, and appropriate in order to add value. Representational information quality means that the information must be understandable, easy to manipulate, and concise and consistent in its representation. Accessibility information quality refers to how accessible the information is (Lee, Strong, Kahn, & Wang, 2002).

Table 4. Academics' view of information quality (Lee, et al., 2002)

	Intrinsic IQ	Contextual IQ	Representational IQ	Accessibility IQ
Wang & Strong (1996)	Accuracy, believability, reputation, objectivity	Value-added, relevance, completeness, timeliness, appropriate amount	Understandability, interpretability, concise representation, consistent representation	Accessibility, ease of operations, security
Zmud (1978)	Accurate, factual	Quantity, reliable/timely	Arrangement, readable, reasonable	
Jarke & Vassiliou (1997)	Believability, accuracy, credibility, consistency, completeness	Relevance, usage, timeliness, source currency, data warehouse currency, non-volatility	Interpretability, syntax, version control, semantics, aliases, origin	Accessibility, system availability, transaction availability, privileges
Delone & McLean (1992)	Accuracy, precision, reliability, freedom from bias	Importance, relevance, usefulness, informativeness, content, sufficiency, completeness, currency, timeliness	Understandability, readability, clarity, format, appearance, conciseness, uniqueness, comparability	Usableness, quantitativeness, convenience of access
Goodhue (1995)	Accuracy, reliability	Currency, level of detail	Compatility, meaning, presentation, lack of confusion	Accessibility, assistance, ease of use (of h/w, s/w), locatability
Ballou & Pazer (1985)	Accuracy, consistency	Completeness, timeliness		
Wand & Wang (1996)	Correctness, unambiguous	Completeness	Meaningfulness	

Furthermore, Lee, et al. (2002) compiled the academics' view of information quality. Information quality dimensions mentioned by scholars are gathered and categorized based on the four categories of information quality dimensions. Lee, et al.'s (2002) academics' view of information quality can be observed in Table 4.

Although Lee, et al.'s (2002) academic views of information quality contains an extensive list of information quality dimensions identified by many scholars, only the accessibility information quality dimensions, which is related to how accessible the information is, have the highest relevance to the context of open government data portal. This is because, as it had been mentioned in section 2.2, an open government data portal serves as the central point of access for open government data. Meanwhile, intrinsic, contextual, and representational information quality dimensions are more relevant to the content (open government data) being uploaded into the portal itself, instead of the access to the content.

2.4.4. Relevant dimensions of transparency, privacy, and information quality

Based on the literatures regarding transparency, privacy, and information quality which has been discussed in section 2.4.1, 2.4.2, and 2.4.3, a list of dimensions of transparency, privacy, and information quality which are relevant to the context of open government data portal can be identified (Table 5). Some dimensions which are mentioned by different scholars are combined together. For example, the openness of government are combined together into the dimension of publication of open government data. Dimensions which have little relevance to the context of open government data portals are not included in this list.

Table 5. Dimensions of transparency, privacy, and information quality that are relevant with open government data portal

Value	Dimension	Definition
Transparency	*Publication of open government data* (Bauhr & Grimes, 2012; Hollyer, Rosendorff, & Vreeland, 2014; Kim, 2009; Matheus & Janssen, 2015)	"The extent to which governments publish information electronically, as well as their clarification of reasons for restricting information when the wider public interest clearly demands it" (adapted from Bauhr & Grimes, 2012, p. 7, and Kim, 2009, p. 11)
	Participation (Kim, 2009)	"The practice of involving members of the public in the agenda-setting, decision-making, and policy-forming activities of organizations/institutions for policy development" (Rowe & Frewer, 2005, p. 253)
	Accountability (Matheus & Janssen, 2015)	"Answerability for the government's actions or inactions and the responsibility for their consequences" (Matheus & Janssen, 2015, p. 238)
Privacy concern (due to the publication of open government data)	*Combining data* (Smith, Milberg, & Burke, 1996)	"Concern that personal data in disparate databases may be combined into larger databases, thus creating a 'mosaic effect'" (Smith, et al., 1996, p. 172)

Value	Dimension	Definition
	Lack of control over personal information (Borgesius, Eechoud, & Gray, 2015)	"The lack of control over one's personal information if that information is released as open data" (Borgesius, et al., 2015, p. 2089)
	Social sorting and discrimination (Borgesius et al., 2015)	"Classification of people and populations according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access, and so on" (Lyon, 2003, p. 30)
Privacy concern (due to the use of open government data portals)	*Identity privacy* (Martínez-Ballesté, Pérez-Martínez, & Solanas, 2013)	The possibility of disclosure of a data portal user's identity, thus allowing his/her activities in the data portal to be correlated to his/her identity (Martínez-Ballesté, et al., 2013)
	Query privacy (Martínez-Ballesté, et al., 2013)	The possibility that the queries made by users to services are saved and thus allowing users to be profiled and information about their habits can be obtained (Martínez-Ballesté, et al., 2013)
	Location privacy (Martínez-Ballesté, et al., 2013)	The possibility that a user's physical location can be disclosed (Martínez-Ballesté, et al., 2013)
Information quality	*Accuracy* (Batini, Cappiello, Francalanci, & Maurino, 2009)	"The extent to which data are correct, reliable, and certified free of error" (Wang & Strong, 1996, p. 31)
	Completeness (Batini, et al., 2009)	"The degree to which a given data collection includes data describing the corresponding set of real-world objects" (Batini, Cappiello, Francalanci, & Maurino, 2009, p. 16:7)
	Consistency (Batini, et al., 2009)	The degree to which the semantic rules defined over a set of data items are not violated by the data (Batini, et al., 2009)
	Timeliness (Batini, et al., 2009)	"The extent to which the age of data is appropriate for the task at hand" (Wang & Strong, 1996, p. 32)
	Accessibility (Wang & Strong, 1996; Jarke & Vassiliou, 1997)	"The extent to which data are available or easily and quickly retrievable" (Wang & Strong, 1996, p. 32)
	Ease of operations (Wang & Strong, 1996)	"The extent to which data are easily managed and manipulated (i.e. updated, moved, aggregated, reproduced, customized)" (Wang & Strong, 1996, p. 32)
	Assistance (Goodhue, 1995)	"The extent to which help is provided in accessing and understanding the data" (Goodhue, 1995)

Value	Dimension	Definition
	Ease of use (of h/w, of s/w) (Goodhue, 1995)	The extent to which the open government data portal is easy to learn and use (Goodhue, 1995)

The dimensions of privacy concern are separated into two categories: privacy concerns due to the publication of open government data and privacy concerns due to the use of open government data portals. The information quality dimensions are mostly comprised of the accessibility dimensions compiled by Lee, et al. (2002). However, although the intrinsic, contextual, and representational information quality dimensions has less relevance, it is possible that they may also affect the design of an open government data portal. For this reason, as well as to keep the research to be focused on the highest relevant dimensions, the basic set of information quality dimensions mentioned by Batini, et al. (2009) are also included in Table 5.

Table 5 answers the first sub-question of this research, which is the following:

> *What are the dimensions of transparency, privacy, and information quality that are relevant to open government data portal?*

The dimensions listed in Table 5 will be used in extracting the design requirements of open government data portals which will be discussed in the next chapter of this report.

2.4.5. Connection between transparency, privacy, and information quality

Although transparency, privacy, and information quality are separate concepts, they are closely connected to each other and affect each other. For example, the concept of privacy is closely connected to transparency. Violation of privacy has become one of the concerns in the effort to increase government transparency through the publication of government data (Conradie & Choenni, 2014; Zuiderwijk & Janssen, 2014). Janssen and van den Hoven (2015) argue that the value of transparency and privacy compete with each other and that sometimes trade-offs between transparency and privacy should be made when sharing information. Unclear trade-offs between the two values (transparency and privacy) can become a barrier in publishing government data (Janssen, et al., 2012). Therefore, in order to achieve the goal of government transparency through open government data, privacy concerns are something that has to be dealt with.

Information quality can also affect transparency. Based on the definition of transparency, transparency is achieved if people know what their government is doing. In the context of open government data, the open government data contains the information regarding the actions of a government. However the ability of people to obtain information from open government data is affected by its information quality. The accessibility information quality dimensions compiled by Lee, et al. (2002), for example, affect how accessible these data are. When people are not able to access open government data, it is possible that transparency, which is one of the goal of publishing open government data, may not be achieved.

Chapter 3
Design Requirements of Open Government Data Portals and Additional Findings from Interviews

This chapter focuses on the interview conducted during this research, which is part of the second phase of the research. The chapter will present the selected interviewee and their backgrounds, the interview process, the data analysis process, and findings from the interview. Findings from the interview will be presented as the design requirements of open government data portals and answer the second sub-question of this research, which is the following:

"What design requirements of open government data portals are relevant to transparency, privacy, and information quality?"

The findings from this chapter will contribute to the objective of this research by providing input for the formulation of the design principles. At the end of this chapter, the design requirements of open government data portals extracted from the interview as well as additional findings that may contribute to the formulation of design principles will be delivered.

In summary, the process, instrument, and deliverable of the second phase of the research can be observed in the following figure.

Phase	Process	Instrument	Deliverable
2	Design requirement identification	Interview	Design requirements of open government data portals (SQ2)

Figure 4. Phase 2 of the research

3.1. Interview purpose

The interview aims to explore and identify the design requirements of open government data portals. The design requirements of open government data portals obtained from the interview will be used as an input for the formulation of design principles, as well as used in the design observation by observing how existing open government data portals implement the design requirements. In order to conduct the interview systematically, an interview protocol is developed and used. The interview protocol includes how to select interviewees, what to prepare before the interview, the structure of the interview, and the questions asked during the interview. The complete interview protocol can be found in the Appendix B of this report.

3.2. Interviewee selection criteria

The interviewees are selected based on the following criteria:

1. Works as an academic researcher in TU Delft
2. Focuses on open data research
3. Has experience in using an open government data portal
4. Willing to participate in the interview

Open government data portal design requirements which are obtained through the interviews are derived from the dimensions of transparency, privacy, and information quality. Therefore, in order

to derive the open government data portal design requirements, understanding of the concepts of transparency, privacy, and information quality, which are some of the key concepts in open data research, is imperative. Academic researchers who focus on open data research are chosen as the interviewees since they fit into this requirement. Moreover, since the interview results will be used to observe cases from diverse countries and government levels, a selection of interviewees' with experience on different open government data portals each is also desired in order to have a more diverse view on design requirements of open government data portals.

TU Delft has many experts in open data research and it is one of the highest contributing institutions in open data research (Scholl, 2014; Hossain, Dwivedi, & Rana, 2016). Due to the expertise of the researchers of TU Delft as well as practical reasons, one of the criteria used in selecting interviewees is that they work as an academic researcher in TU Delft. Some academic researchers may also have experience in open government data portal development or design. In this case, the experience of the academic researchers may provide a valuable input in reaching the objective of the interview. Therefore, if possible, the selection of interviewees will focus on academic researchers who have prior experience in designing or developing an open government data portals. However, this is not a strict criterion, and as such, will not be included in the interviewee selection criteria.

The number of interviews involved in this research is not fixed. Although the interview process follows the initial list of potential interviewees, the interview process may be stopped or a number of interviews may be added depending whether saturation has been achieved or not.

3.2.1. Selected interviewee

Using the interviewee selection criteria, five interviewees are selected. The interviewees all currently work as an academic researcher at the TPM Faculty of TU Delft, with different focuses of research. However, their research fields or their previous works are all related to open data. Some interviewees also have practical experience as open data users and one of the interviewees has an experience in designing and developing an open government data portal in his home country.

1. Interviewee 1

   Interviewee 1 is an academic researcher from Indonesia, and is currently working in the field of big and open linked data. His main focus is on exploring how organizations can create value from big and open linked data. He has been working as an academic researcher in TU Delft for more than one and a half year. His current work also involves observing open data portals such as Dublin open data portal and Rotterdam smart city open data portal. Therefore, he has some experience as an observer of open data portals. During his previous work at a telecommunication company, he often used data that are open in nature, such as statistical data, in order to get insight and create new products for the company.

2. Interviewee 2

   Interviewee 2 is an academic researcher from Brazil. He is currently focusing on the field of transparency in big and open linked data. He has about ten years of experience working on the field of open data. Aside from working as an academic researcher, he also works as a data scientist, collecting, treating, analyzing data, and creating a dashboard which helps in decision making process. Interviewee 2 has experience in open government data portal design. He created the open government data portal for the government of Rio de Janeiro in Brazil.

3. Interviewee 3

Interviewee 3 is an academic researcher from Belarus. She has worked for more than three years in the field of open data. Her current academic research focuses on the topic of information sharing, specifically data sharing between different sectors and the concept of data collaboratives. Her previous work was focused on the topic of challenges to participation and was related to open data initiatives. Her previous work was also focused on the use or demand side of open data. This included the usability, impact, and outcome of open data initiatives as well as organizational aspects of information sharing.

4. Interviewee 4

Interviewee 4 is an academic researcher from China. His current research primarily focuses on privacy and trust enhancement for data sharing. In relation to open data, his work focuses on the security of open data, enhancing the security and privacy of data before they are being published. His research also focuses on open research data. It has a similar concept with open government data, although the type of data being handled is different. However, due to the similarity in concept, it can be said that he has some experience in open data portal design.

5. Interviewee 5

Interviewee 5 is an academic researcher from Indonesia. He has been working as an academic researcher in the field of open data for three months, with the focus on the evaluation of open government data impacts. However, he has experience in using open government data since 2014 when the presidential election in Indonesia took place. Besides working as an academic researcher, he also works for the government of Indonesia, specifically the Supreme Audit Board of the Republic of Indonesia.

After the interviewees are selected using the interviewee selection criteria, schedules for interview were arranged with the interviewees. The summary of the selected interviewee as well as their interview date can be observed in Table 6.

Table 6. Summary of selected interviewees

Interviewee	Field of research	Date of interview
Interviewee 1	How organizations create value from big and open linked data	4 August 2016
Interviewee 2	Transparency in big and open linked data	11 August 2016
Interviewee 3	Information sharing and open data initiatives	12 August 2016
Interviewee 4	Privacy and trust enhancement for data sharing	12 August 2016
Interviewee 5	Evaluation of the impacts of open data	16 August 2016

3.3. Interview process

The interview takes around a one hour time and was conducted in a semi-structured manner. The interview questions mainly revolved around the transparency, privacy, and information quality dimensions which are explained in section 2.4.4. The main focus of the interview is to find out what open government data portal design requirements are correlated to the dimensions of transparency, privacy, and information quality. However, during the process of the interview, knowledge related to the design principles of open government data is also explored.

The interviews were divided into five sections: introduction of the interviewer, introduction of the interviewee, transparency section, privacy section, and information quality section. At the beginning of the transparency, privacy, and information quality sections, the definition of the corresponding

concept (i.e. transparency, privacy, or information quality) are explained to the interviewee so that the interviewee has the same perspective on the concept as the interviewer. The interviews are recorded using a recording device under the permission of the interviewee.

The following are the topics of questions and discussion asked during the interviews.

Introduction of the interviewer

Before the interview starts, the interviewer gives an introduction about the interview. The following subjects are explained by the interviewer:

1. Name of the interviewer
2. Brief introduction to the interview topic and objective
3. Confidentiality agreement
4. Asking interviewee's permission to record interview

Section 1: Introduction of the interviewee

1. What is your name?
2. From which country are you?
3. What research field are you primarily associated with?
4. How long have you been working in the field of open data?
5. How do you associate yourself with the field of open data?
6. What experience do you have in open government data portal design?

Section 2: Transparency

This section focuses on and iterate each of the dimensions of transparency. For each of the dimensions, the definition of the dimension is explained and the interviewees are asked the following questions:

1. Identification of transparency dimensions on open government data portals
   a. Are you familiar with the concept of <name of dimension>?
   b. Which features and functionality of an open government data portal are related to <name of dimension>?
2. Open government data portal design principles
   a. What design requirements should an open government data portal comply with in regard to <name of dimension>
   b. What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 2, the following question is asked to the interviewee:

3. Do you have any more comments to add regarding the transparency dimensions?

Section 3: Privacy

This section focuses on and iterate each of the dimensions of privacy concerns. For each of the dimensions, the definition of the dimension is explained and the interviewees are asked the following questions:

1. Identification of privacy concern dimensions on open government data portals
   a. Are you familiar with the concept of <name of dimension>?

       b.   Which features and functionality of an open government data portal is related to <name of dimension>?

  2.   Open government data portal design principles

       c.   What design requirements should an open government data portal comply with in regard to <name of dimension>

       d.   What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 3, the following question is asked to the interviewee:

  3.   Do you have any more comments to add regarding privacy dimensions?

Section 4: Information quality

This section focuses on and iterate each of the dimensions of information quality. For each of the dimensions, the definition of the dimension is explained and the interviewees are asked the following questions:

  1.   Identification of information quality dimensions on open government data portals

       a.   Are you familiar with the concept of <name of dimension>?

       b.   Which features and functionality of an open government data portal is related to <name of dimension>?

  2.   Open government data portal design principles

       c.   What design requirements should an open government data portal comply with in regard to <name of dimension>

       d.   What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 4, the following question is asked to the interviewee:

  3.   Do you have any more comments to add regarding information quality dimensions?

3.4.   Data analysis process

After the interviews are conducted, the recording from the interviews are transcribed and analyzed. The analysis involves labeling the interview transcript with codes as well as finding the relations between codes. The coding process is done with the help of the software tool for qualitative data analysis ATLAS.Ti. Two categories of codes were used in the coding process. The first are the codes that have been prepared before the coding process begins. These codes are derived from the dimensions of transparency, privacy, and information quality which are the focus of this research. The following are the codes from this first category: publication of open government data, participation, accountability, combining data, lack of control over personal information, social sorting and discrimination, identity privacy, query privacy, location privacy, accuracy, completeness, consistency, timeliness, accessibility, ease of operations, assistance, and ease of use. The purpose of having these codes is so that it would be easier to identify which part of the interview is related with the dimensions of transparency, privacy, and information quality since the interview is conducted in a semi structured manner and it is possible to jump to different kind of relevant topics without strictly following the structure of the interview.

The second category of codes consists of open codes that are found from the interview transcripts. These codes were not prepared beforehand. Instead, they were formulated from the statements that the interviewee made during the interview. The process of looking for this kind of codes is called open coding. After the open coding process is done, the relationships between the codes

found during the open coding are explored. This is done to gain insight into how the statements of the interviewees are related to each other.

3.5. Findings from the interviews

This subsection will discuss the findings from the interviews. The findings are categorized into two: findings that contributes to design requirements for open government data portals, and findings that are not directly related to design requirements, but contributes to design principles. The design requirements for open government data portals found through the interviews will be used in design observation which will be discussed in the next chapter of this report.

3.5.1. Design requirements

As what have been discussed in section 1.5 (p. 6), design requirements define what functionalities an open government data portal should or should not have, and having a set of design requirements can help in determining how an open government data portal should be designed. Among the questions asked during the interviews are what functionalities of open government data portals are related to the dimensions of transparency, privacy, and information quality, as well as the design requirements of open data portals related to the functionalities. From the interviews, a set of 12 design requirements for open government data portals are identified. The design requirements are explained below.

Table 7. Design requirements for open government data portals

#	Design requirement	Relevant dimension
[1]	Dataset request a) Provide form for requesting datasets b) Provide tracking of dataset request	Publication of open government data, accountability
[2]	User feedback a) Allow users to comment on datasets b) Allow users to rate datasets c) Allow users to upload a better version of a dataset	Participation, consistency
[3]	Provide space for interaction between users	Participation, accountability
[4]	Help/assistance a) Provide FAQ b) Provide channel for users to ask question c) Provide tutorials d) Provide examples of dataset use and apps that have been built	Publication of open government data, accountability, accuracy, completeness, consistency, timeliness, ease of operations, assistance
[5]	Tagging of datasets	Publication of open government data, accessibility, ease of use
[6]	No registration for download	Identity privacy
[7]	Search function a) Filter for search results b) Search suggestion	Publication of open government data, query privacy, timeliness, accessibility, ease of use
[8]	Provide API for dataset access	Accessibility, ease of operations
[9]	Use open, machine processable data format a) Allow users to choose format to download	Accessibility, ease of operations

#	Design requirement	Relevant dimension

[10]	Provide metadata
	  a) Use standardized metadata
	  b) Publish data owner
	  c) Publish semantic rules
	  d) Publish history of changes
	  e) Publish update frequency/schedule
		Publication of open government data, accountability, accuracy, completeness, consistency, timeliness, ease of operations, assistance

[11]	Provide dataset visualization a) Allow comparison of different datasets	Timeliness, ease of operations, assistance

[12]	User friendly interface
	  a) Categorize datasets into topics
	  b) Use universal icons
	  c) Highlight important datasets
	  d) Social media integration
	  e) Use consistent design between pages
		Publication of open government data, accessibility, ease of use

1. Dataset request

All interviewees agree that open government data portals should provide a function for users to request datasets. This requirement is related to the dimension of "publication of open government data" and "accountability". Sometimes, the government may not be able to publish all government data. When a user cannot find the datasets that they need on the open government data portals, he or she may want to request the datasets from the relevant government agencies that are responsible for the data. The obligation of governments in fulfilling the request for datasets can be limited by regulations. Interviewee 2 commented:

"*The first thing is to observe what the law points out, the context (city, country), the 15 principles of open government data, and the 5 stars of linked data. From these, we can conclude what the government obligation is with regard to the publication of open government data*" (interviewee 2).

Different governments may have different regulation concerning what kind of data a government must publish and whether they should publish data when the public requests it or not, which can affect how dataset requests are being fulfilled. However, all interviewees agree that the function to request datasets is important for open government data portals. Interviewee 3 responded when asked about the importance of the functionality to request datasets:

"*Of course it depends on which country context you're looking into if there is, for example, an obligation that the government has to actually provide it. Then it would make more sense than in the other context. But requesting dataset is a really important feature*" (interviewee 3).

When users request for datasets that are related to the accountability of a government, such as the report of tax money use by the government, it becomes important for the users to be able to know how their request is being handled by the related government agency. With regard to this concern, open government data portals should provide the ability to track the handling of dataset requests, such as whether or not the request has been received by the responsible government agency, whether or not the request can be fulfilled, and when the dataset can/will be published.

2. User feedback

When governments publish datasets, users may have different opinions regarding the datasets. One user may feel that the datasets are lacking in quality. For instance, the data may be too out-of-date that some users feel that the dataset is no longer relevant, or users may feel that the dataset contains too little information to be useful. Open government data portals should give users the functionality to provide feedback regarding the datasets. This functionality can be implemented in various ways. For example, it can be implemented as the feature to rate datasets. One interviewee made the following response with regard to the functionality to rate datasets:

"Another feature that can be useful is the ability of users to rate the quality of information themselves outside of the metadata so that other people can see how many 'stars' a dataset has" (interviewee 3).

Another interviewee gave the following response with regard to dataset rating:

"It is relevant with the dimensions of information quality. If, for example, a user found that a data is no longer timely, they can rate the data with low number to show other users their opinion regarding the timeliness of the data. The data provider may also benefit from the feedback by knowing which aspect to improve from the datasets" (interviewee 1).

Besides rating datasets, another feedback can be implemented as the ability for users to comment on datasets. Commenting allows users to express their opinions with arguments, and thus giving more input for the data provider regarding the datasets. With regard to the dimension of participation, one interviewee responded with the following comment:

"From user's perspective, I think the participation feature is the ability of users to make comments about the datasets because they need to judge whether the datasets are relevant or valid" (interviewee 4).

A more advanced feedback that a user can give is by providing an enriched or corrected dataset. For example, a dataset may contain some inaccuracies or inconsistencies. A user can provide feedback by uploading his or her version of a corrected dataset in the feedback section. Regarding this functionality, an interviewee gave the following response:

"…provide space for users to upload qualitative review of datasets or upload a new corrected or enriched version of datasets through combination of datasets. The feature of uploading datasets is important as it completes the cycle of data publication: publishing, enriching, and cleansing datasets" (interviewee 2).

User feedback can benefits government in many ways. First, users may gain an insight from datasets that they use, and this insight may help in government's decision making process. For example, a user can find an insight about a potential problem from traffic and transportation data. Feedback from the user regarding the potential problem can help government in designing better traffic policies. Next, it can help data providers in detecting errors such as inconsistencies or inaccuracies in the datasets by receiving reports from users. For users of open government data portals, the benefit that they can receive by the implementation of feedback function is, among others, revealing the average perceived quality of the datasets.

3. **Provide space for interaction between users**

Users of open government data can have different background and abilities in using the open government data portals and manipulating the data. For example, one user can be a data scientist while another user can be a software developer. Discussion between different kinds of users can lead to a new insight or development of new services based on the available datasets. To accommodate the interaction between users, an open data portal should provide a space where users can have discussion between themselves. Interviewee 3 gave the following comment with regard to user space in open government data portals:

"If you look at open data portals, participation of users can be reflected through such features as user to user communication tools, for example integration with social media which would allow discussion of datasets, exchange of knowledge on how to find certain datasets, or how to interpret datasets, what tools to use, for example" (interviewee 3).

In discussing datasets, interaction between users can take the form of a showcase of what insight users have found from the datasets or what kind of tools can be used to manipulate datasets. Furthermore, interaction between users can give insight for data providers. For example, data providers may find what kind of datasets are popular among users and what kind of datasets are less popular. This insight can lead to a certain decision by the data provider, such as which datasets should be focused on in the future.

4. **Help/assistance**

Not every users may be familiar with the functionalities of an open government data portals. First time users may find difficulties in searching for datasets or using the functions provided by the open government data portals. In another case, they may find it difficult to work with the datasets that the government provides on the open government data portal since they do not know which tools to use to manipulate the data. In order to allow users to get familiar with the functionalities of an open government data portals or to work with the datasets, assistance should be provided for the users. Assistance can be provided in various ways, such as providing frequently asked questions and tutorials in the form of videos, presentations, or free online courses. One interviewee provided the following response when asked about the functionalities that an open government data portal should provide in regard to assistance:

"Examples of usage on the metadata, real data usage such as apps that have been built based on certain datasets, free online courses on how to use the data or the portal, frequently asked questions, the metadata, providing channel for people to ask questions about datasets" (interviewee 2).

Users may still have questions regarding datasets, such as asking about the meaning of a field in a datasets. To help users in resolving their problems, the open data portal should provide a channel for users to ask questions. This channel can be implemented as a helpdesk which is in charge of answering users' questions. Another interviewee gave the following response with regard to assistance:

"It is really important for data providers to provide some kind of channels to get back to the data providers to ask questions and get help when something does not work. It is also possible to provide frequently asked questions, tutorials, and videos that show you the basics on operations and how to do things" (interviewee 3).

Examples of real usage of data from the open government data portals as well as apps that have been built based on the data should also be provided on the open government data portals. Such examples can help users in learning the use of open government data in real life and finding inspiration on how to use the datasets from the open government data portals.

5. **Tagging of datasets**

Datasets often have keywords that they are mainly associated with. For example, a dataset about a list of high schools in Delft can be tagged with "high-school" and "Delft". This indicates that both keywords, "high-school" and "Delft", are relevant with the dataset. Tagging can help users in finding the correct dataset when they are browsing datasets or searching a dataset by entering the relevant keywords. Interviewee 1 provided the following response when asked about tagging function in open government data portals:

"It will ease the users in looking for the datasets that they are interested in. This may help the government in achieving their goal of creating transparency" (interviewee 1).

Interviewee 5 suggested that tagging functionalities are related with how accessible the datasets are:

"Categorization of datasets and tagging is also related to the accessibilities of datasets" (interviewee 5).

Meanwhile, interviewee 2 gave the following response when asked which features of an open government data portal is related to ease of access:

"Tags, search functionalities, universal icons, colors, design of interface, position of headers and search box" (interviewee 2).

6. **No registration for download**

Registration is a common feature in many websites. Usually, during registration to use a website, users may have to provide their email address and username. However, for accessing open government data, registration may not be a feature that users want. The requirement for registration before downloading datasets in itself contradicts the "non-discriminatory" principle of the 8 principles of open government data, which states that data should be "available to anyone, with no requirement of registration" (The 8 Principles of Open Government Data, 2016). Interviewee 2 responded with the following comment when asked about which functionalities of open government data portal are related to identity privacy:

"Login feature, and I think it should not be implemented as it is discriminatory and contradicts the 15 principles of open government data" (interviewee 2).

Interviewee 4, however, suggested that a registration feature may be necessary in order to monitor the access of datasets by users. Interviewee 4 gave the following comment when asked about his opinion on identity privacy:

"From my personal experience, in order to manage the operation of the portal, I need to record who has downloaded these datasets. In this way, I want users to be registered in order to use the portal, because if something wrong happens I would need to check the history log. Therefore, users' personal identity can be exposed to the open data portal. However, the logs are not exposed to public" (interviewee 4).

However, with regard to query privacy and query recording, the same interviewee responded that users should be able to choose their privacy preferences, through the following comment:

"At least provide an opt-out option for users, so that they can have more freedom about their privacy preferences" (interviewee 4).

Another interviewee emphasized that the choice to implement registration on open government data portals depends on the context of the country and the purpose of the function itself, through the following comment:

"…it really depends on the context: what kind of country, what kind of level, what the purpose is, and what other functionalities does the portal have" (interviewee 3).

In short, by default, open government data portals should not require users to do registration before they download datasets as it contradicts the 8 principles of open government data. Not requiring users to register also increases the user-friendliness of the open government data portals since users can explore the portals and download datasets more freely. However, depending on the context of the country or purpose, a government may choose to implement the need to register before downloading datasets.

7. **Search function**

In order to reach the goal of transparency through the publication of open government data, citizens need to be able to access datasets published by the government. When too many datasets have been uploaded to the open government data portal, sometimes the fastest way to find a dataset is by searching for the datasets that the open government data portal provides. Therefore, for mature open government data portals, a search function may become a crucial functionality for users and accessibility to datasets that users are interested in in some way relies to the search function. As an important functionality of an open government data portal, search function also affects the ease of use of the open government data portal.

All interviewees agree that a search function implemented in an open government data portal has to provide several features. Firstly, besides simply searching by keywords or title of datasets, it needs to at least provide the ability to filter search results based on domain/topic, the government agency that published the datasets, and the time period when the datasets are uploaded. By allowing users to filter search result based on the time period when datasets are uploaded, users can make sure that the search can result in timely datasets. With regard to the search function, interviewee 3 gave the following comment:

"In open data portals, you can see that the datasets are divided into topics, and it is possible to do a search using the keywords or tags which are used in the portal. It would be also useful if the users are also prompted with similar keywords that they can use to search for similar datasets. Another mechanism is to have a good filter which would allow you to refine your search, for example what kind of area, what period of time, or what unit of government you are interested in" (interviewee 3).

Meanwhile, on the same topic, interviewee 4 responded with the following comment:

"Users should be able to search by different categories, different domains, and other criteria defined by users. Furthermore users should be able to filter criteria for search results, such as time range, organization, size, formats, etc. A searching suggestion like what Google use can also be useful for users" (interviewee 4).

Secondly, the search function should help users in picking the right search term by providing suggestions of search terms. This suggestion can be implemented as autosuggestion functionality, like what has been implemented in many popular search engines. It can also be implemented as relevant search terms when the user has finished entering one search term. For example, after a user search for "tax report" in an open government data portal, the open government data portal also suggest the search term "tax report 2016" along with the search results for the previous search term.

Search function can be related to query privacy of open government data portal users. However, most interviewees agree that query privacy is not a problem for open government data portals. Most interviewees also suggested that queries from users should be tracked so that it can be used to improve the services provided by the open government data portals. For example, by knowing which datasets or which topic is searched more extensively by users, data provider can focus on providing more datasets that the users are interested in.

8. **Provide Application Programming Interface (API) for dataset access**

For transparency purpose, providing datasets as downloadable documents in open data format may be enough. However, for other goals of publication of open government data such as creating innovative products or services (Janssen, 2011), access to datasets require a faster method. API can increase accessibility of open government data by allowing the access of datasets through a programming interface, which is useful for dataset access by external applications. However, not everyone may be familiar with using an API, and it may be suitable for people who have the required competence, as what one of the interviewee said in his interview response:

"APIs are for technical people. If data are only provided through API, then it will be discriminatory for users who does not understand how to use API" (interviewee 2).

For people who can use API for creating new service, the use of API can increase the ease of operations of datasets. However, the decision to use provide API in an open government data portal should be based on who the target users of the open government data portal are. For example, mature open government data portals which provide a large number of datasets may reap benefit from providing API to its users. However, for a smaller open government data portal which focuses more on transparency goal, providing API in its early growth stage may not be necessary. Interviewee 4 gave the following response when asked about API:

"If the portal focuses on transparency, then you would have to make it as accessible as possible for regular people. But if it is a specialized portal which focuses on providing a way for businesses and entrepreneurs to get the data, then APIs are more efficient. You have to know your user group and the level of skills and knowledge that they have and estimate their abilities to work with data" (interviewee 4).

Therefore, providing APIs may be beneficial in general. Due to this, it is listed as one of the design requirements of open government data portals. However, depending on the context of open government data portal, not providing API can also be acceptable.

9. **Use open, machine processable data format**

Open government data can be provided in different formats. However, according to the 8 principles of open government data, open government data should be provided in a non-proprietary format and should be machine processable (The 8 Principles of Open Government

Data, 2016). This means that proprietary data format such as XLS (Microsoft Excel data format) or open format that is not machine processable such as PDF (Portable Document Format) does not fall into the category suggested by the 8 principles of open government data.

Most interviewees agree that providing data in machine processable format can lead to higher ease of operations and open government data portals should be designed to provide users with formats that are ready for operations. For example, interviewee 1 commented:

"How can you do an operation on PDF files? The data needs to be in a format that is ready for operations by data users" (interviewee 1).

As data are provided by government agencies and uploaded into open government data portals by different agencies, the data format used in open government data portals need to be standardized. Interviewee 4 gave the following response with regard to data format:

"There are some widely accepted formats, such as PDF, CSV, or XML. The format in an open data portal needs to be standardized" (interviewee 4).

Furthermore, as users may use different tools to manipulate datasets, they should be able to download datasets in their preferred format. Interviewee 2 gave the following comment with regard to data format:

"People should be able to download data in the formats that they prefer" (interviewee 2).

10. Provide metadata

The transparency goal of publication of open government data can succeed when users of open government data portal understand what the data is about. In the effort to help users understand the datasets better, the open government data portal should provide metadata of the published datasets. Metadata contains data about data, and the contents of metadata, when all relevant information about a dataset is provided, may help users in judging the accuracy, completeness, consistency, and timeliness of the datasets.

Since datasets are provided by different government agencies, there needs to be a standardization on what kind information the data provider puts into the metadata. According to the interviewees, the metadata should at least include: data owner, semantic rules, history of changes, and update frequency/schedule. Publishing the data owner who is responsible for the datasets can help in the effort to increase government's accountability. Interviewee 4 gave the following response when asked about the dimension of accountability:

"I think there is at least a specific area to clearly define the responsibility of who and what should be liable for the datasets" (interviewee 4).

Publishing semantic rules can help users check the consistency of data in a dataset. Meanwhile, publishing history of changes and update frequency/schedule can help users in determining the timeliness of a dataset. Interviewee 3 gave the following comment with regard to metadata:

"It is useful to have some sort of history, where you can track back what has happened to the datasets that you are interested in, how, when, and by whom it was updated, if any of the users cite the dataset and what they have been doing with the datasets" (interviewee 3).

11. **Provide dataset visualization**

Visualization is the feature to view data on open government data portals. Data from a dataset can be viewed in different formats, including graphs when the data supports it. In the simplest form, it can be implemented as the ability to view the content of a dataset directly on the open government data portal. Most interviewees agree that visualization is an essential functionality of an open government data portal. Among the benefits of the function of visualization which were mentioned by the interviewees are providing convenience for the users, and allow people whose system has low processing power to view dataset directly on the open government data portal. Interviewee 2 gave the following response when asked about functionalities of open government data portals that are related to accessibility:

"People can decide on which format they want the datasets, and the visualizations of the datasets. Sometimes people do not have the processing power to handle the data on their own systems" (interviewee 2).

For visualization function, one of the requirements is that users should be able to view and compare different datasets. Comparing data from different time periods can accommodate the timeliness dimension of information quality. Interviewee 2 responded with the following answer when asked about functionalities of an open government data portal that are related to timeliness:

"It will be also useful to create a visualization of datasets from different time. Then, users can, for example, compare datasets of inflation from 2010 and 2016" (interviewee 2).

12. **User friendly interface**

User interface is an important part of open government data portal that affect the success of the publication of open government data. It also affects the accessibility of datasets and determine how easy the open government data portal is to use. The requirements for user interfaces mentioned by the interviewees include the categorization of datasets into different topics, use of universal icons, highlight of important datasets, social media integration, and the use of a consistent, standardized design between different pages.

Categorization of datasets into different topics can help users to find relevant datasets that they need faster. The use of universal icons can help users identify what topics are available and make sure that they click on the right link on the interface of the open government data portals. Interviewee 2 gave the following response with regard to icons:

"…for people that prefer browsing open data portals, the open data portals should have universal icons that represent different topics or domains of datasets and lead to different categories of datasets" (interviewee 2).

Highlights on important datasets may help users in accessing certain information faster. Interviewee 5 gave the following example with regard to important datasets:

"I think transparency in Indonesian perspective still emphasize on financial data as it is related to the accountability of the government. Every year, the process of financial reporting and budgeting follows a certain timeline, starting from the preparation to the final reporting. So, the open data portal should be able to highlight or promote the relevant financial datasets based on the ongoing process of the reporting" (interviewee 5).

Social media integration is not a core functionality of an open government data portal. This means that without the functionality, an open government data portal may still achieve its goals. However, integration with social media may encourage more user discussion on datasets. Interviewee 3 provided the following comment with regard to integration of social media on open government data portals:

"If you look at open data portals, participation of users can be reflected through such features as user to user communication tools, for example integration with social media which would allow discussion of datasets…" (interviewee 3).

Standardized design between pages can help users in getting used to the interface of an open government data portal, and thus use the portal more efficiently. The interviewees mentioned several criteria for the design of open government user interface, including the choice of suitable colors and good placement of items.

3.5.2. Additional findings from the interviews

Besides the findings that are directly related to the design requirements above, there are more findings from the interviews which are not directly related to design requirements of open government data portals but could contribute to the design principles of open government data portals. The additional findings are the following.

1. **Privacy is an issue that has to be dealt with before the publication of open government data.**

    Although privacy issues can become a problem when published data contain privacy-sensitive contents, the measures to prevent privacy issues needs to be taken before the publication of open government data. If we take a look at the design requirements for open government data portals which are discussed in the previous subsection, no design requirements listed are related to combining data, lack of control over personal information, and social sorting and discrimination, which are the privacy concerns due to the publication of open government data. Most interviewees agree that the measures to prevent privacy issue due to the publication of open government data is not related to any functionality of an open government data portal. Rather, proper steps to ensure privacy during preparation of a dataset is what should be done by government agencies. What the open government data portal can do in regard to this issue is more to providing explanation on the steps that government agencies has taken to ensure that no privacy issue will arise from the publication of open government data.

    Another privacy concern that is not related to the design requirements is location privacy. The closest data that websites can use to track user's location is IP. However, even if his or her IP is known, a user's precise location cannot be easily found. Some interviewees agree that location privacy is not an issue in open government data portal. Moreover, by knowing the general location of a user, a more customized service can be developed for the user. Rather than IP, interviewee 4 suggested that GPS location that can be contained in a picture uploaded by a user may be a more relevant issue for location privacy, as GPS data embedded in some pictures can be used to track the precise location where the pictures were taken.

2. **Functionalities implemented in an open government data portal should match the context of regulation, target users, and other situation surrounding the portal.**

    Some requirements for the open government data portals are requirements that are more global and can be effective in many contexts. For example, a user friendly interface may be something that every portal should implement, although the implementation may be different

from portal to portal. However, some functionalities on an open government data portal are functionalities that may be changed depending on the target users of the open government data portal. For instance, the effective implementation of functionalities such as API depend on the purpose of the portal and the target users of the portal. If the portal is still in its early stage, with a limited number of datasets uploaded and the main purpose is for transparency purpose, providing API to access datasets in the portal may not be necessary. Many other functionalities, such as registration, and dataset request, depend on the regulation that determines the obligation and the rights of the government as well the citizens. The purpose of the implementation of certain functionalities also affect how the functionalities should be implemented. With regard to the negative consequence of not implementing registration, interviewee 4 gave the following comment:

"As data provider, there is a need to know who have been using the data. If the registration of using the data portal is not implemented in the process, I think it is not fair for the government" (interviewee 4).

3. **Potential users should be involved in the development process of an open government data portal.**

Users are one of the most essential stakeholders of an open government data portal. Most interviewees agree that potential users should be involved in the development of an open government data portal. Taking into account the inputs from users during the development of an open government data portal can result in a portal design that matches what the users really need. Interviewee 3 gave the following comment with regard to user involvement in open government data portal development:

"It's really important to consider the target community in thinking about the overall design requirements of a portal and what kind of uses are you focusing on. These different group of users may have different requirements. Involving the users in the design process of a portal may allow the design of the portal to be in accordance with the needs of the actual users" (interviewee 3).

The involvement of users can be done in various ways, such as consulting them about the design requirements that need to be implemented in an open government data portal. Another way to involve users is by taking their inputs during the prototyping phase of an open government data portal. Interviewee 4 gave the following comment with regard to prototyping of open government data portals:

"First, you need to interview the potential users and get some idea about their perception about a certain open government data portal. Next you need to give them something that they can play with. Then, they can provide more useful information" (interviewee 4).

3.6. Conclusion of chapter 3

The interviews result in two types of findings. The first type of findings are focused on the design requirements of an open government data portal. From the interviews, 12 open government data portal design requirements are acquired. They are: dataset request [1], user feedback [2], provide space for interaction between users [3], help/assistance [4], tagging of datasets [5], no registration for download [6], search function [7], provide API for dataset access [8], use open data format [9], provide metadata [10], provide dataset visualization [11], and user friendly interface [12]. These open government data portal design requirements are the answer to the 2nd sub-question of this research project, which is:

"What design requirements of open government data portals are relevant to transparency, privacy, and information quality?"

The second type of findings from the interviews are findings that are not directly related to design requirements, but are still relevant to the objective of this research. These findings are: privacy is an issue that has to be dealt with before the publication of open government data [1], functionalities implemented in an open government data portal should match the context of regulation, target users, and other situation surrounding the portal [2], and potential users should be involved in the development process of an open government data portal [3]. The first type of findings will be used in the design observation of open data portals, while the second type of findings will be used in formulating the open government data portal design principles which are the objective of this research project (see section 5.1).

Chapter 4
Open Government Data Portal Design Observation

This chapter focuses on the design observation of open government data portals, which is the thirds phase of the research. An overview of the selected open government data portals will be provided and findings from the design observation of the portals will be presented in this chapter. Findings from the design observation will answer the third sub-question of this research, which is:

"How do existing open government data portals implement the design requirements of open government data portals related to transparency, privacy, and information quality?"

The deliverable for this chapter is a set of findings from the result of the design observation. The findings from this chapter will contribute as an input to the formulation process of design principles.

In summary, the process, instrument, and deliverable of the third phase of the research can be observed in the following figure.
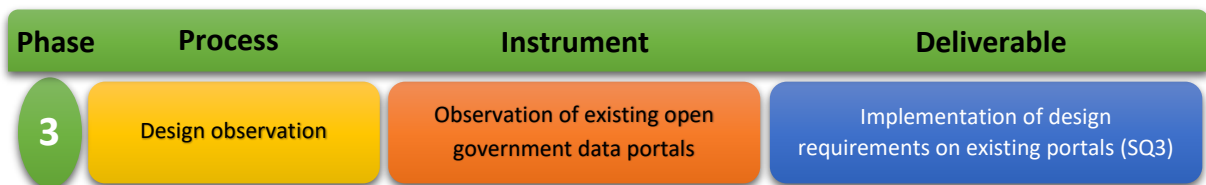

Phase	Process	Instrument	Deliverable
3	Design observation	Observation of existing open government data portals	Implementation of design requirements on existing portals (SQ3)

Figure 5. Phase 3 of the research

4.1. Design observation

The design observation of open government data portals is a continuation of the previous phase of the research. In the previous phase, design requirements of open government data portals have been identified through expert interviews. However, the interview result mostly provides the perspective from the user's side, as the interviewees discusses what functionalities a portal should or should not have based on the needs of a user. This means that the perspective of the data provider, or the supplier's side, is not explored. Furthermore, the interview result consists of the opinions of open data researcher about how an open government portal should be designed. However, how applicable these design requirements are in the real world cannot be derived from the interviews.

The design observation attempts to bridge the design requirements identified from the user's side with the actual implementation of the design requirements, which may reflect the perspective of the data provider on the design requirements, by finding out how existing open government data portals implement the design requirements identified from the interviews. At the same time, the design observation will also observe the applicability of the design requirements by connecting the context of the existing open government data portals and the implementation of the design requirements. Ultimately, the design observation contributes to the formulation of the design principles by providing findings that may reinforce previous findings, or provide new findings that has not been obtained in the previous phase of the research.

In conducting the design observation, five existing open government data portals are selected and observed using the selection criteria and design observation protocol explained in the next sections.

The portals are selected from the cases submitted by the participants of the 2016 MOOC on Open Government (see section 1.5.3). The MOOC is chosen as the starting point to select the open government data portals to be observed as it provides information regarding various open government data portals which can be used as a list for the portal selection purpose.

4.2. Open government data portal selection

171 cases were retrieved from the MOOC, posted by 190 participants. The number of cases are fewer than the number of participants as some participants submitted the same cases to the MOOC. Cases that are the same were grouped together and counted as a single case. The administrator of the MOOC provided five submission categories for participants to submit their cases: data and information sharing, service development, citizen participation and social media, legislation, and other. Participants were required to submit their cases to the submission category relevant to their cases. However, in practice, many participants submitted their cases in the wrong category or outside of the categories provided by the MOOC administrator. Therefore, although the most relevant category for this research is the category of data and information sharing, this research does not only include cases submitted in the category. Instead, all cases submitted were included as potential cases to be used in this research project.

4.2.1. Open government data portal selection criteria

Although many cases were retrieved, not all of them were relevant to the topic of this research project. Furthermore, to be able to get more in-depth analysis into the cases, the timeframe of this research allowed the use of only several cases. Therefore, a selection of cases is necessary to ensure that only the most relevant cases are used in this research. In selecting the cases, it is necessary to do a classification of the cases in order to get an overview of what each cases are about and its relevancy to this research. Then, the cases are filtered using the initial case selection criteria. The initial selection criteria are the criteria of the cases that are relevant to this research. There are five initial selection criteria used in the case selection of this research:

1. Cases are from MOOC

    This research project will only use cases from the course of Open Government of the 2016 MOOC. The course is run by TU Delft, and the cases used in this research are taken from assignment 1 of the Open Government course. The number of cases submitted by the participants of the MOOC should be sufficient to provide a list of open government data portals to be used in this research.

2. Related to data sharing

    Only cases related to data sharing are selected since they are relevant for this research and can be used as materials for analysis. There are other topics from the cases taken from the MOOC, such as the legislation of open government or the use of social media in open government. For example, a participant from the USA submitted a case regarding the Open Meeting & Public Records Act in California, and a participant from Indonesia submitted a case regarding the use of social media in enhancing public policy formulation. Although they are related to open government, they are not relevant to the objective of this research, and as such, will not be selected to be used in the design observation.

3. Provides sufficient information

Participants of the MOOC were required to provide information regarding the case that they submitted, including the objective of the case, the actors involved in the case, general description of the case, geographical coverage of the case, and references regarding the case. To be able to use the case submissions in desk research and case analysis, sufficient information regarding the cases are required. Therefore, only cases that fulfill this criteria are selected for the design observation.

4. Relevant to the concept of transparency, privacy, and information quality

Cases presented by the participants of the MOOC should be relevant to the concept of transparency, privacy, and information quality because the concepts are used as the foundation for case analysis in reaching the objective of this research. In filtering the case using this criteria, an examination of the description of the case submitted by the participant will be conducted to determine whether the case has relevance to privacy, transparency, and information quality.
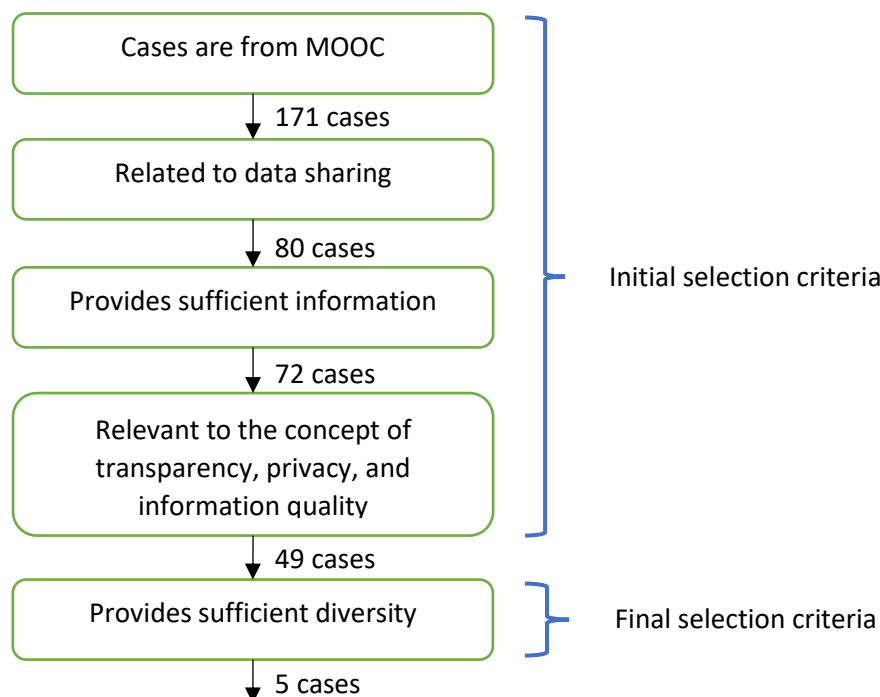
Figure 6. Case selection criteria

The selection of cases for this research are done in stages, as shown in Figure 6. Firstly, cases are selected using the first selection criterion, which selects cases from the MOOC as the initial potential cases for case studies. This results in 171 cases from different topics related to open government. Next, the 171 cases are filtered using the second case selection criterion. At this stage of case selection, only cases related to data sharing are selected. The resulting number of cases are then filtered using the next case selection criterion. This filtering goes on until it reaches the final selections criterion. The final result of the initial case selection are cases which are relevant to this research, namely cases from the MOOC which are relevant to the concept of transparency, privacy, and information quality in data sharing by governments.

The initial case selection results in 49 cases. This number of cases is still too many for this research. Ideally, this research should select around five cases which are the most relevant to the research and

provide sufficient diversity to ensure high relevance of the resulting design principle. The number five is chosen so that the selected cases may provide sufficient material for analysis. Therefore, final case selection is conducted by selecting five most relevant cases from the 49 cases acquired from the initial selection criteria. A small number of cases also allows a more in-depth analysis on the cases can be conducted. However, this number is only an approximation as the number of cases can be adjusted in order to fit the final case selection criteria.

The final selection criteria used in selecting the cases is that the selected cases should provide sufficient diversity. The participants of the MOOC are required to submit cases from their own countries. Therefore, geographical coverage of the cases will be diverse. The difference in geographical relevance may also causes differences in the cultural aspects that affect the cases. Selected cases should have sufficient diversity to ensure that the design principles extracted from the cases are relevant to as wide geographical and cultural coverage as possible. For that reason, in the final selection of cases, these criteria of diversity are sought:

1. Geographical diversity
   Selected cases should be diverse geographically to ensure that the countries where the cases from are culturally diverse. Since culture may affect how governments design and implement open data portals, a diverse selection of cultures are sought to ensure that the design principles which are the objective of this research can be relevant for a broader spectrum of culture.
2. Diversity of the level of government
   The cases retrieved from the MOOC discuss open government at different levels of government (local, regional, national, global). The selected cases should have a diversity in the level of government so that the design principles extracted from the cases are relevant for different levels of governments.
3. Diversity of the concept of transparency, privacy, and information quality
   Finally, not all cases are focused on all of the concepts of transparency, privacy, and information quality. Some cases only touch the aspect of transparency while other cases discuss the problem of privacy or information quality. Since the design principles should emphasize on all of the three concepts, it is imperative that the selected cases should cover the concepts of transparency, privacy, and information quality.

Even though a broader relevance of the resulting design principles is sought, it may be possible that there will be principles that only apply to a certain culture, level of government, or concept. This kind of design principle will still be taken into account during the design principle formulation and therefore, in this research project, we aim for an open mind to such possibility.

After a sufficient number of open government data portals are selected using the final selection criteria, they are then used in the open government data portal design observation phase.

4.2.2. Result of initial selection of open government data portals

The initial selection of the open government data portals from the cases submitted by the participants of the Open Government Course of the 2016 MOOC using the initial case selection criteria (see section 1.5.3) results in 49 open government data cases. These cases are from different levels of governments, including cities, provinces/regions, and countries. There is also a case that discusses open data implementation at global worldwide level. Table 8 shows the number of cases for each different level of governments.

Table 8. Number of cases by level of government

| | City | Region | | | | Country | Worldwide | Total |
		County	Province	Autonomous community	State			
Number of cases	13	1	1	1	1	29	1	**49**

Meanwhile, from a geographical perspective, the cases selected using the initial case selection criteria come from 29 different countries and 1 global (worldwide) case (see Table 9). The countries can be categorized using the continents where they are located in. There are 7 cases from Asia from 7 different countries, 5 cases from Africa from 5 different countries, 19 cases from Europe from 10 different countries, 5 cases from North America from 2 different countries, 12 cases from South America from 5 different countries, and 1 global case.

Table 9. Number of cases by geographical region

	Asia	Africa	Europe	North America	South America	Worldwide	Total
Number of cases	7	5	19	5	12	1	49
Number of countries	7	5	10	2	5	1	30

4.2.3. Selected open government data portals

From the 49 cases submitted by the participants of the MOOC, 5 cases are selected using the final selection criteria to be used in this research. The final selection criteria is used to allow the selected cases to have sufficient diversity and therefore lead to higher relevance of the design principles produced. As the cases come from countries from all of the continents in the world, a high geographical diversity would be achieved if each of the five cases represents one different continent. However, the development level of the countries where the cases come from must also be taken into account so that the selected cases represent both developed countries as well as developing countries. This is because for the 49 cases, most countries from some regions, such as Europe, are developed countries. Meanwhile, countries from other regions, such as Asia, can be a mix of developed and developing countries. It is preferable for the selected cases to represent both developed countries and developing countries so that higher diversity can be achieved.

The 49 cases also discuss different levels of government, from local (cities), regional (provinces/states), to national (countries). To allow higher diversity, the selected cases should represent different levels of governments. A higher diversity in the level of governments would allow the resulting design principles to be relevant to different levels of governments.

Using the final selection criteria, five cases from the MOOC are selected:

1. **Fingal County Open Data Portal (Ireland)**
   This case discusses the implementation of open government data sharing by the County of Fingal, Ireland. Fingal launched an open data portal, http://data.fingal.ie/, in 2010 which hosts Fingal related data published by public sector organizations in Fingal.

2. **Portal Site of Official Statistics of Japan**

   The portal site of official statistics of Japan provides statistical information collected from various ministries in Japan. These data can be shown as graphs on the open data portal http://www.e-stat.go.jp/. Graphs can also be exported and downloaded in various formats, such as CSV or XLS.

3. **Kenya Open Data Portal**

   Kenya open data portal, https://opendata.go.ke/, provides open government data from various sectors in Kenya. The datasets in the portal are published by Kenyan government and can be viewed directly on the portal or downloaded in various formats, including CSV, XLS, and XML.

4. **US open government**

   This case discusses the United States open data portal, https://data.gov, which hosts public sector data published by various government agencies. The open data portal, which was launched in 2009, provides visualizations of the datasets as well as allows datasets to be downloaded in ready-to-use formats (CSV, XLS, etc.).

5. **Open data Rio de Janeiro (Brazil)**

   The city of Rio de Janeiro, Brazil, launched its own data portal to host open government data published by the municipal government of Rio de Janeiro. Various information can be accessed on the portal, such as public transportation data and healthcare facilities. Data can be downloaded from the portal, http://data.rio, in various formats, including CSV and XML.

Table 10. Overview of selected cases

#	Case title	URL	Continent	Country	Government level	Country development level
1	Fingal County Open Data Portal (Ireland)	http://data.fingal.ie/	Europe	Ireland	County (regional)	Developed country
2	Portal Site of Official Statistics of Japan	http://www.e-stat.go.jp/	Asia	Japan	Country (national)	Developed country
3	Kenya Open Data Portal	https://opendata.go.ke/	Africa	Kenya	Country (national)	Developing country
4	US open government	https://data.gov	North America	USA	Country (national)	Developed country
5	Open data Rio de Janeiro (Brazil)	http://data.rio	South America	Brazil	City (local)	Developing country

The five cases (can be observed in Table 10) should provide high diversity, as many of the diversity categories are represented by the cases. First of all, each case comes from different region in the world since one case is selected for each of the world continents. Next, the cases also represent both developed and developing countries. Three cases come from developed countries (Ireland, Japan, USA) while two cases are from developing countries (Kenya, Brazil). Lastly, three of the cases discuss open government data portal implementation on a national level, one case discusses open

government data portal implementation on the regional level, and the last case discusses open government data portal implementation on a local level. Therefore, local, regional, and national government levels are also represented by the selected cases.

4.3. Design observation protocol

The purpose of the open government data design observation is to find out how existing open government data portals implement the design requirements obtained from the interviews. The result of the observation can provide additional insights that will contribute in the formulation of the design principles of open government data portals. In order to be able to do a design observation of open government data portal systematically, a design observation protocol is required. The design observation protocol defines what steps need to be taken during the design observation process. Using the defined steps, the design observation results from different open government data portal can be combined and analyzed together, or compared when necessary.

The protocol for design observation of open government data portal is as follows:

1. The observation starts with the first open government data portal (Fingal County), and after the whole observation for the first open government data portal finishes, the observation continues to the next portal, and so on until all five open government data portals are observed. When required, further observation on an open government data portal after the sequence of observation ends will be conducted directly on the open government data portal without following the previous sequence.
2. The observation involves checking whether the design requirements discussed in section 3.5.1 are implemented in the open government data portal, and how they are implemented.
3. The observation for each open government data portal will start at the main page of the open government data portal and continue to other pages under the same main domain of the portal.
4. All relevant links provided in the main page will be opened to do the observation on the opened pages. The relevant links provided in the opened page will be opened to do observation, and so on until no more finding from design observation is obtained.
5. Observation which requires downloading or viewing datasets from the open government data portal will proceed using the first found relevant dataset from the portal.

4.4. Result of design observation

All five open government data portals are observed in the month of September, 2016. During the design observation, the result of the observation for each open government data portal is recorded. The result of the design observation can be observed in Table 11.

Table 11. Result of design observation

Design requirement		Portal	Observation result
[1]	Dataset request	FGL	Form to request dataset is provided (email required), previous requests are displayed, but tracking of request cannot be observed
		JPN	No dataset request function is available
		KEN	Form to request dataset is provided (login required), previous requests are displayed as well as their status (open, approved, rejected)
		US	Form to request dataset is provided, previous requests are displayed as well as their status (open, closed)
		RIO	No dataset request function is available
[2]	User feedback	FGL	Datasets can be rated. No functionality to comment on datasets or upload a better version of a dataset
		JPN	No user feedback functionalities are provided with regard to datasets

Design requirement		Portal	Observation result
		KEN	Datasets can be rated. No functionality to comment on datasets or upload a better version of a dataset
		US	Datasets cannot be rated by users. Users can report problem for datasets through web form or email. No functionality to comment on datasets or upload a better version of a dataset
		RIO	No user feedback functionalities are provided with regard to datasets
[3]	Provide space for interaction between users	FGL	Interaction between users only available under the page of a dataset request or a showcase of apps built by users
		JPN	No interaction space is provided for users
		KEN	Interaction between users only available under the page of a dataset request
		US	Interaction between users are managed on external site, but link is provided on the portal
		RIO	No interaction space is provided for users
[4]	Help/assistance	FGL	A simple FAQ is available, no channel for users to ask questions, no tutorial. However, a simple guide on which application can be used to open datasets is provided. A page is provided to showcase the use of datasets by users
		JPN	Extensive help and tutorial page is provided (in Japanese language), users can ask questions using a web form (email required)
		KEN	FAQ is provided, users can ask questions by sending email, postage mail, or through phone call. Tutorial is not provided, example of dataset use and apps built is not provided.
		US	FAQ is provided, users can ask questions through web form or email. Tutorial is not provided, examples of dataset use and apps built are provided.
		RIO	A simple FAQ is available. No channel for users to ask questions. Manual is provided on external website, but link to manual is provided on the portal. Examples of apps built are provided.
[5]	Tagging of datasets	FGL	Datasets are tagged with keywords
		JPN	No tagging visible
		KEN	Datasets are tagged with keywords
		US	Datasets are tagged with keywords
		RIO	Datasets are tagged with keywords
[6]	No registration for download	All	No registration required to download datasets
[7]	Search function	FGL	Search function is provided, but no filtering options are available. No search suggestion available
		JPN	Search function is provided. Filtering based on topic, government agency, and time period is possible. No search suggestion available
		KEN	Search function is provided. Filtering based on topic is possible. No search suggestion available
		US	Search function is provided. Filtering based on topic, government agency, and time period is possible. No search suggestion available
		RIO	Search function is provided. Filtering based on topic, government agency, and time period is possible. No search suggestion available
[8]	Provide API for dataset access	FGL	No API is provided for users
		JPN	API is provided (registration is required to use API)
		KEN	API is provided
		US	API is provided
		RIO	API is provided
[9]	Use open, machine processable data format	FGL	Datasets are provided in open, machine processable data formats (CSV, XML, RDF, KML). Choice of data format is limited on certain datasets.
		JPN	Most datasets are provided in proprietary format (Excel), but some dataset can be downloaded as CSV or XML after being viewed using the visualization function on the portal.
		KEN	Datasets can be downloaded in various open, machine processable formats (CSV, JSON, RDF, RSS, XML)
		US	Datasets can be downloaded in various open, machine processable formats (CSV, JSON, RDF, XML, etc.)
		RIO	Datasets can be downloaded in various open, machine processable formats (CSV, JSON, XML, etc.)

Design requirement	Portal	Observation result
[10] Provide metadata	FGL	Metadata is provided and standardized across datasets. Data owner, release date, semantic rules, and update frequency is provided. Information provided regarding changes to dataset is limited to last revision date.
	JPN	Metadata is provided and standardized across datasets. Data owner and release date are provided. Semantic rules are provided in Japanese language only.
	KEN	Metadata is provided and standardized across datasets. Data owner information and release date is provided, but semantic rules and update frequency is not.
	US	Metadata is provided and standardized across datasets. Data owner, semantic rules, release date, and update frequency is provided. No information provided regarding changes to dataset.
	RIO	Metadata is provided and standardized across datasets. Data owner and release date is provided. Semantic rules and update frequency are not provided. History of changes to datasets is provided.
[11] Provide dataset visualization	FGL	No visualization provided
	JPN	Visualization available in table and graph format, but only for one dataset
	KEN	Viewing dataset content directly on portal is possible, visualization is provided in various graph format.
	US	Visualization is provided through 3rd party website for selected datasets (registration to the 3rd party website may be required)
	RIO	Visualization is provided but not working
[12] User friendly interface	FGL	Datasets are categorized into topics, but no icons are used to represent topics. There are featured datasets, social media integration is provided for commenting (interaction between users), and sharing the link of opened pages on social media.
	JPN	Datasets are categorized into topics, but can only be viewed based on topics using the filter feature in the search function. No highlights on datasets, no universal icons, no social media integration, standardized design of pages
	KEN	Datasets are categorized into topics, icons are used to represent different topics, there are featured datasets, integration with twitter on portal for tweet viewing purpose, standardized design for the pages under the main domain
	US	Datasets are categorized into topics, icons are used to represent different topics, there are featured datasets, integration with google+, twitter, and Facebook on portal for sharing on social media purpose, standardized design for the pages under the main domain
	RIO	Datasets are categorized into topics, icons are used to represent different topics, there are featured datasets, integration with google+, twitter, and Facebook on portal for sharing on social media purpose, standardized design for the pages under the main domain

The summary of the result of observation for each of the design requirement is as follows.

1. **Dataset request**

   Among the five open government data portals observed, only the Japan statistics portal and Rio open data portal do not provide the functionality for the users to request datasets. The other three portals provide users with the ability to request datasets from the data provider. Kenya open data portal and the US open data portal further provide the functionality to track the status of a dataset request, whether the request is still open or has been resolved. Meanwhile, Fingal County open data portal does not provide the tracking functionality. Kenya open data portal, although provide the functionality for users to request dataset, require that users register and login first in order to request a dataset. Login using one of the user's social media accounts (Facebook, twitter, or google+) is also possible. Requesting a dataset on the Fingal County open data portal requires users to provide their email address.

2. **User feedback**

   Fingal County and Kenya open data portal allows users to rate the datasets published on the portals. This functionality is not found in US open data portal. However, US open data portal

provides a web form and email address to report a problem with datasets. Japan statistics and Rio open data portal does not provide any functionality for feedback on datasets. None of the five open government data portals provide the functionality to put comment on datasets or upload a corrected/enriched version of datasets.

3. **Provide space for interaction between users**

Fingal County and Kenya open data portal provides limited interaction space for users. Meanwhile, although US open data portal does not provide an interaction space for users directly on the portal, it provides a space for user interaction on external websites where users can discuss or ask questions about datasets or features of the portal. Japan statistics and Rio open data portal does not provide any functionalities for interaction between users.

4. **Help/assistance**

Fingal County, Kenya, US, and Rio open data portal provide FAQ-like information. Japan statistics, Kenya, and US open data portal all provide channels for users to ask questions. Japan statistics provide extensive help and tutorial page in Japanese language. The Japan statistics portal itself is available in Japanese or English language. Rio open data portal provide manual in external website. US and Rio open data portals both provide a showcase of apps built by users based on the datasets published on the portals. Fingal County open data portal, on the other hand, provides a showcase of dataset use by its users, which is also available on US open data portal.

5. **Tagging of datasets**

All observed open government data portals except the Japan statistics portal provide tagging functionalities for their datasets.

6. **No registration for download**

None of the five observed open government data portals requires users to register to download datasets directly from the portals.

7. **Search function**

All five observed open government data portals provide search functionalities, although different in capabilities. Fingal County open data portal does not provide filtering functionality, unlike the other four portals. Kenya open data portal only allows filtering based on topic and tag. Meanwhile, Japan statistics, US, and Rio open data portals allow filtering based on topic, government agency, and time period. No search suggestion functionality is provided in any of the five open government data portals.

8. **Provide API for dataset access**

All observed open government data portals except Fingal County open data portal provide an API, although their capabilities may be different. For example, in its API information page, it is written that the US open data portal provides an API only to access metadata of datasets but not the content of the datasets itself.

9. **Use open, machine processable data format**

All observed open government data portals except Japan statistics open data portal provide datasets in various open, machine processable data formats. Data format choice is limited on

certain datasets on the Fingal County open data portal. Meanwhile, Japan statistics portal provide most of its datasets in the proprietary Excel format.

10. Provide metadata

Metadata is provided by all observed open government data portals. Japan statistics, Kenya, and Rio open data portal provide information on data owner and release date. Meanwhile, Fingal county and US open data portals provide information about data owner, release date, semantic rules, and update frequency. History of changes to datasets is available on Rio open data portal. However, on Fingal County open data portal, history of changes is limited to the last revision date of datasets.

11. Provide dataset visualization

Japan statistics open data portal allows visualization for certain datasets in table and graph format. Kenya open data portal allows visualization for all of its datasets in table and graph format. Meanwhile, US open data portal allows visualization for certain datasets through 3rd party websites, which may require registration on the 3rd party website. No visualization functionality is available on Fingal County open data portal. A visualization functionality is supposed to be available on Rio open data portal, but during the observation, the functionality was not functioning.

12. User friendly interface

Generally, all observed open data portals have a standardized design across different pages, which can be seen from the consistent placement of the main menus on different web pages on the portals. Datasets are also categorized on different topics on the portals. However, the Japan statistics open data portal does not allow users to browse datasets under the different topics directly. Instead, users have to use the search function to list the datasets and use the filtering function to see datasets under a certain category. No icons are used to represent the different topics on Japan statistics and Fingal County open data portals. Fingal County, Kenya, US, and Rio open data portal provides highlight on certain datasets by showing them as "featured datasets" on the portals.

4.5.   Findings from design observation

To be able to identify how the existing design of open government data portals reflects previous findings, understanding of the context of each of the observed open government data portal is necessary. During the observation of the five open government data portals, additional information regarding the portals is also acquired from the open government data portal. This information provide context in understanding why certain design element may or may not be implemented on the open government data portals. Besides information taken from the open government data portals, information from other sources such as articles or conference proceedings may also be used to provide context to the observed portals. The additional information can be observed in table below.

After the design observation on all five open government data portals is finished, the result of the observation is analyzed in order to find insight. The result of observation is compared with the design requirements acquired in the previous phase of the research (see section 3.5.1 for detailed description of design requirements). Based on the comparison, there are generally three kinds of findings obtained from the design observation: 1) findings that conform to previous findings of this

research, 2) findings that do not conform to previous findings of this research, and 3) new findings that are not described in previous findings. Previous findings here are the findings form the interview, including the design requirements and other findings which may contribute to open government data portal design principles (see section 0).

Table 12. Information regarding observed open government data portals

Information about portal	Fingal County	Japan Statistcs	Kenya	US	Rio
Launch year	2010	2008*	2011	2009**	2014
Current number of datasets	273	78000+	904	18000+	1000+
Data scope	County	Country	Country	Country	City
Platform used	-	-	Socrata	CKAN + Wordpress	CKAN

* Source: Akatani (2015)
** Source: Wikipedia (data.gov - Wikipedia, 2016)

1. **Findings that conform to previous findings of this research**

The design observation shows that there are several design implementations on the open government data portal which match the findings from the previous phase of this research. First, there are functionalities that are part of the design requirements and are implemented on most of the observed portals. For example, dataset request, help/assistance functionalities, dataset tagging, no registration requirement to download datasets, search function, API, open and machine-processable data format, metadata, and standardized design of user interface are implemented by most of the open government data portals observed. This shows that these functionalities are indeed required by the open government data portals. Therefore, this finding confirms some of the conclusion from the previous chapter regarding the design requirements of open government data portals.

However, there are several exceptions regarding the implementation of the design requirements. For example, unlike four other open government data portals, Fingal County open data portal does not provide API for its users. A possible explanation can be obtained by looking at the context of Fingal County open data portal. API is designed for accessing datasets quickly through applications. The more datasets exist in the portal, the more valuable the existence of an API would be. However, the number of the datasets that Fingal County open data portal provide, by the time the portal was observed, was 273. This is despite the fact that it has been launched for six years. Seeing the low number of datasets it provide, it may be true for Fingal County open data portal that providing an API for dataset access is not yet necessary. In this case, it confirms the previous finding that functionalities implemented in an open government data portal should match the context of the portal.

Another example of how context affects the design of an open government data portal can be seen on the Japan statistics open data portal. Among the five observed portals, the Japan statistics open data portal was launched the first in 2008. The portal uses its own platform, instead of using an open source platform such as CKAN (which is used by US and Rio open data portal), or other platforms that are specially designed for open data portals such as Socrata (which is used by Kenya open data portal). Furthermore, the Japan statistics has been focusing on the use of API and making more of its datasets available through the API (Akatani, 2015). Considering the history of the open data portal and its focus on API, it may be understandable

how the functionalities provided by Japan statistics open data portal are often different from the other four portals. For example, the Japan statistics portal does not provide topic browsing capabilities similar to the other portals as it uses its search function as the way to browse datasets. Furthermore, the interface design used in the Japan statistics open data portal is different from the other observed portals, such as the design of the tab menu.

2. **Findings that do not conform to previous findings of this research**

The observation result also shows how certain design requirements are not implemented by most open government data portals. This means that it contradicts the previous findings where the interviewees said the design requirements are essential to be implemented in open government data portals. The design requirements not implemented by most open government data portals include user feedback, space for interaction between users, and visualization. The dataset feedback functionalities implemented in most observed portals are limited only to rating of the datasets. The ability to rate a dataset is not even available in US open data portal which is one of the most mature portals among the five observed portals. The US open data portal, however, provides the possibility for users to provide feedback through a web form or by email.

Space for interaction between users is not provided in most of the observed open data portals, with the exception of US open data portal. Even if there is a possibility for users to interact with between each other on the Fingal County and Kenya open data portal, the implementation is limited only to commenting on other users' request on datasets, and there is no possibility for users to discuss existing datasets. US open data portal, on the other hand, provides the space for interaction on an external website. The possible benefit of this implementation is that users can focus more on different topics of discussion and offload the internet traffic burden to the external website.

Visualization function is available on Japan statistics and Kenya open data portal. However, the visualization function on Japan statistics open data portal is limited to certain datasets, unlike the Kenya open data portal. The visualization function on Kenya open data portal is provided by the Socrata platform it uses, and it works on most datasets provided on the Kenya open data portal. The US open data portal, on the other hand, uses third party services to give visualization to datasets. This feature works only on certain datasets, and users will be redirected to external website to view the visualization.

The limited implementation of user feedback, user space, and visualization function in most observed open data portals shows that it contradicts the findings from the interviews, where these functionalities are considered important in the effort to accommodate transparency, privacy, and information quality assurance through the design of open government data portals. However, this does not mean that the functionalities are left out by most of the observed open data portals because the design requirements are not important. Unfortunately, the result of the design observation and known context regarding the open data portals are not enough to provide an explanation, and further exploration on the context of the open data portals may be required to know the reason why these features are left out.

3. **New findings that are not described in previous findings**

The third finding from the design observation is that there are functionalities that are implemented by the observed portals and are related to the dimensions of transparency, privacy, and information quality, yet not explored during the interviews. These functionalities

are language and the requirement for registration to request a dataset. Language is an essential element in presenting the content of an open government data portal to its users. In some cases, target users of an open government data portal may have different needs in term of language. This may require the portal to provide a user interface with the capability to switch language. The Japan statistics portal may fall into this category of a portal, as it allows the user to switch the portal interface to either Japanese language or English language, although some contents such as the help pages are strictly provided in Japanese language. In comparison, other observed open government data portals only provide a single language for their user interface, which is English for the Fingal County, Kenya, and US open data portals, and Portuguese for Rio open data portal. Providing the user interface with the language that users are familiar with can increase the ability for users to access datasets and quickly retrieve datasets that they need. Therefore, this functionality is related to the accessibility dimension and the dimension of publication of open government data.

Another functionality that may have a relation with the dimensions of transparency and privacy is the requirement to register in order to request datasets. This functionality is implemented by Kenya open data portal. As what have been discussed in the previous chapter, the requirement to register before downloading datasets is considered discriminatory and thus should not be implemented by open government data portals. Requesting a dataset is one of the efforts that can be done when a user requires access to a dataset that is not available on the open government data portal. Therefore, the end purpose is the same as downloading a dataset, which is to get one's hands on a required dataset. In this sense, the functionality should be provided to the users without discrimination, and thus the design requirement of "no registration for download" may need to be extended to "no registration to download or request dataset".

The requirement of registration can also be found in the visualization function of the US open government data portal. However, this is due to a restriction on the third party website used to generate the visualization, and not a restriction found natively in the portal itself. Furthermore, the requirement to register to view a visualization of dataset does not reduce the ability of a user to download the dataset itself. Hence, this requirement is irrelevant to the requirement to allow users to download datasets without having to register first.

4.6. Conclusion of chapter 4

The design observation phase of this research involves observing five selected open government data portals: Fingal County open data portal, Japan statistics open data portal, Kenya open data portal, US open data portal, and Rio open data portal. Each portal is selected to represent different variation of open government data portal and is aimed to add diversity to the whole selection of portals. The observation process itself is done by observing whether the design requirements of open government data portals which are acquired in the previous phase of this research are implemented by the five selected open government data portals.

The result of the design observations shows a high variation in the implementation of the design requirements by the five open government data portals. There are design requirements that are implemented by the majority of the portals, including dataset request, help/assistance functionalities, dataset tagging, no registration requirement to download datasets, search function, API, open and machine-processable data format, metadata, as well as the characteristics of user friendly interface, and there are design requirements that are not implemented by the majority of the open government data portals, including user feedback, space for interaction between users,

and visualization. However, even though a design requirement is implemented by the majority of the observed portals, the implementation of the functionality related to the design requirement may be different from portal to portal. The result of the design observation answers the third sub-question of this research, which is:

"*How do existing open government data portals implement the design requirements of open government data portals related to transparency, privacy, and information quality?*"

In general, there are three kinds of findings that can be generated from observing the implementation of the design requirements by the open government data portals. The first kind of findings is the findings that conform to previous findings of this research. These include: 1) design requirements that are implemented relatively uniformly by the five observed portal, including dataset request, help/assistance functionalities, dataset tagging, no registration requirement to download datasets, search function, API, open and machine-processable data format, metadata, as well as the characteristics of user friendly interface, and 2) the implementation of certain design requirements on an open government data portal, such as API, should match the context of the open government data portal.

The second kind of findings is the findings that do not conform to previous findings of this research. These include the design requirements that are not implemented by most of the observed open government data portals, including user feedback, space for interaction between users, and visualization. Although these design requirements are not implemented by most of the open government data portals, it does not mean that the design requirements are not important. It may be the case that the context of the open government data portals play a role in the missing functionalities. However, the result from the design observation does not provide sufficient information to formulate a conclusion about this.

The last kind of findings are new findings that are not described in previous findings. These include the design requirement of: 1) providing the language that the users need for the user interface, and 2) no requirement of registration should be implemented for requesting dataset. These requirements are not explored during the interview process. However, they are related to the dimensions transparency, privacy, and information quality.

Chapter 5
Open Government Data Portal Design Principles

This chapter focuses on the fourth and the final phase of the research approach, which is the formulation of open government data portal design principles, and answer the main research question of this research, which is the following:

"*What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance?*"

The design principles are formulated from the synthesis of the findings from the literature review, the expert interviews as well as the design observation. Three sets of input are used in synthesizing the open government data portal design principles: 1) design requirements of open government data portals, 2) additional findings from the interview, and 3) findings from the design observation. The deliverable for this chapter is a set of design principles for open government data portals which focuses on accommodating transparency, privacy, and information quality.

In summary, the process, instrument, and deliverable of the final phase of the research can be observed in the following figure.

Phase	Process	Instrument	Deliverable
4	Formulation of design principles	Synthesis of findings	Design principles (main RQ)

Figure 7. Final phase of the research

5.1. Design principle formulation process

The formulation process of the open government data portal design principles is the process of synthesizing the findings from each phase of the research. In the design principle formulation process, each phase of the research contributes to different area of the design principles. During the literature review phase, dimensions of transparency, privacy, and information quality which are related to open government data portal design are identified. Since the objective of this research is to formulate open government data portal design principles which accommodate the value of transparency, privacy, and information quality, the dimensions of transparency, privacy, and information quality identified during the literature review phase will provide a frame for the design principles. Using the frame built from the dimensions, it will be easy to identify which value (i.e. transparency, privacy, or information quality) and dimension each of the design principles contribute to.

The phase of design requirement and design observation, which are conducted through the means of interview and observation of existing open government data portal, will provide the input to formulate the design principles. There are three sets of input obtained through the interview and design observation: 1) a set of design requirements for open government data portal, 2) additional findings from interview, and 3) findings from the design observation of existing open government data portal. Synthesis of the three sets of input results in the design principles of open government data portals. The three sets of inputs are explained below:

1. **A set of design requirements for open government data portal**

    The design requirements of open government data portal contributes to the design principles of open government data portals by providing input on which functionalities are required by an open government data portal. There are twelve design requirements of open government data portals. They are: dataset request [1], user feedback [2], provide space for interaction between users [3], help/assistance [4], tagging of datasets [5], no registration for download [6], search function [7], provide API for dataset access [8], use open data format [9], provide metadata [10], provide dataset visualization [11], and user friendly interface [12].

2. **Additional findings from the interview**

    Additional findings from the interview provide input to the formulation of design principles on the factors that affect how open government data portals should be designed outside of the functionalities that need to be implemented on an open government data portal. These factors include the understanding of context regarding the open government data portal and the development process of the open government data portal.

3. **Findings from the design observation of existing open government data portals**

    The findings from the design observation provide the input for the design principles by providing insight into how the design requirements of open government data portals are implemented on existing open government data portals. Knowing which design requirements are implemented by the majority of the observed portals, as well as taking into account the findings from the interviews, allows the identification which implementations of design requirements are sensitive to context and which are not. Meanwhile, new findings from the design observation contributes to enriching the design requirements. Furthermore, the result of observation can provide examples on how each of the design requirements can be implemented on the open government data portals. These examples, although not part of the design principles, can provide more practical examples on how to translate the design principles taken from the design requirements into functionalities of open government data portals.

5.2. Design principles

Based on the three sets of inputs used in the formulation process, three types of design principles can be identified: function, context, and process. Function-related design principles are taken from the design requirements of open government data portals. These design principles explain which functionalities an open government data portal should have. Meanwhile, context-related and process-related design principles are taken from additional findings from the interview. Context-related design principles explain how an open government data portal should be designed with regard to the context surrounding it. Process-related design principles explain how the open government data portal should be developed. In total, there are 15 design principles of open government data portals which accommodates transparency, privacy, and information quality (Table 13).

Table 13. Design principles of open government data portals

	Design principle	Benefits	Example of implementation	Transparency, privacy, and information quality dimension affected
[1]	Provide the mechanism for users to request datasets and to track whether their request has been approved or rejected	• Allows data providers to align dataset supply with users' needs	• Request through web form • Request through email	Transparency: • Publication of open government data • Accountability
[2]	Provide the mechanism for users to give feedback to datasets	• Detect inconsistencies of published datasets • Allow users to participate in determining dataset quality	• Rate datasets • Report problem of datasets through web form or email	Transparency: • Participation Information quality: • Consistency
[3]	Stimulate knowledge exchange between users by providing space for interaction between users	• Encourage more dataset use by allowing discussion between users	• Provide user forum internally on the portal • Provide user forum on external website	Transparency: • Participation Information quality: • Accountability
[4]	Assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage	• Encourage more dataset use by users of the portal	• Provide examples of dataset usage in a dedicated page • Provide web form or email address for users to ask questions • Provide manual on how to use the website	Transparency: • Publication of open government data • Accountability Information quality: • Accuracy • Completeness • Consistency • Timeliness • Ease of operations • Assistance
[5]	Tag datasets with relevant keywords to allow faster searching of datasets	• Allow users to identify the topic of a dataset easily • Allow users to find the datasets that they need more quickly	• Allow filtering of datasets using tags • Use tags as part of search variable	Transparency: • Publication of open government data Information quality: • Accessibility • Ease of use
[6]	Allow users to download and request datasets without having to register first	• Encourage more dataset use • Attract more users to use the open government data portal	• Provide direct link to datasets on dataset page	Privacy: • Identity privacy
[7]	Assist users in searching datasets by providing filtering and suggestion feature on the search function	• Allow users to search the datasets that they need faster	• Filter based on topic, government agency, time period • Implement autosuggestion in searching form	Transparency: • Publication of open government data Privacy: • Query privacy Information quality • Timeliness • Accessibility • Ease of use
[8]	Provide API when users require to access datasets through external applications	• Allow apps to access datasets from the API function • Encourage more services built based on datasets	• Use open data portal platform which supports API such as CKAN or Socrata	Information quality: • Accessibility • Ease of operations

	Design principle	Benefits	Example of implementation	Transparency, privacy, and information quality dimension affected
[9]	Provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer	• Avoid a lock-in to a proprietary application • Encourage more dataset use	• Provide data in different formats such as CSV, XML, or JSON	Information quality: • Accessibility • Ease of operations
[10]	Provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets	• Help users in understanding what is contained within a dataset • Encourage more dataset use	• Provide link to semantic rules on metadata • Provide history of changes through a dedicated tab on the dataset page	Transparency: • Publication of open government data • Accountability Information quality: • Accuracy • Completeness • Consistency • Timeliness • Ease of operations • Assistance
[11]	Provide the function to visualize datasets to allow quick viewing of datasets	• Allow users to view datasets more quickly	• View content of datasets on portal before downloading • Allow the creation of graph from datasets directly on the portal	Transparency: • Accountability Information quality • Timeliness • Ease of operations • Assistance
[12]	Use user friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need	• Provide a good user experience for the users • Encourage more access to the open government data portal • Allow users to find datasets more quickly	• Highlight important datasets as "featured datasets" • Consistent placement of menu • Allow users to switch between different languages (e.g. Japanese and English)	Transparency: • Publication of open government data Information quality: • Accessibility • Ease of use
[13]	Deal with privacy issues before dataset is uploaded to the portal	• Reduce the possibility of privacy problems after datasets are published	• Build standardized framework for government agencies on how to deal with and prepare privacy sensitive data	Privacy: • Combining data • Lack of control over personal information • Social sorting and discrimination
[14]	Take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities	• Align the functionalities of open government data portals with the needs of the users as well as regulations	• Implement API only when the target users need it • Adjust the implementation of dataset request function to match regulation	All

	Design principle	Benefits	Example of implementation	Transparency, privacy, and information quality dimension affected
[15]	Involve users in early and later development of the portal to match portal design with users' need	• Align the functionalities of open government data portals with the needs of the users	• Focus group discussion to extract user requirements • Gain user input in prototyping phase • Involve users in continuous development of the portal	All

5.2.1. Function-related design principles

Function-related design principles are mainly derived from the twelve design requirements obtained from the interview as well as findings from the design observation. They provide guidelines on which functionalities should be implemented in an open government data portal from the perspective of transparency, privacy, and information quality. There are twelve design principles related to the function of an open government data portal:

1. **Provide the mechanism for users to request datasets and to track whether their request has been approved or rejected**

   This design principle originate from the design requirement of dataset request. Tracking function is also included as part of the principle as it is useful for users in knowing whether their request has been approved or rejected. This design principle accommodates the value of transparency, specifically the dimensions of publication of open government data and accountability.

   The benefit of providing users with the mechanism to request datasets is that it helps data providers in finding out the datasets that users need but not available yet on the open government data portal. Therefore, it allows data providers to align the supply of datasets with the needs of the users.

2. **Provide the mechanism for users to give feedback to datasets**

   This design principle originates from the design requirement of user feedback. It is considered an important requirement by the interviewees. However, the design observation shows that implementation of user feedback feature is limited on the majority of the portals. Furthermore, the portals which implement certain functionalities related to user feedback show variation in which functionality is being implemented. This design principle accommodates the participation dimension of transparency and the consistency dimension of information quality.

   The benefits of feedback functionalities include providing data providers with the ability to detect inconsistencies in the published datasets, and allows the participation of users in determining the quality of published datasets.

3. **Stimulate knowledge exchange between users by providing space for interaction between users**

   The design requirement of providing space for user interaction aims to stimulate knowledge exchange between users as well as more discussions regarding datasets uploaded to the portal. Design observation shows limited implementation of this requirement in the observed portals.

However, as open data portal becomes more mature, this requirement may become more relevant for the data provider. For example, the US open data portal, although it provides user interaction space in an external website, allows the possibility of users to exchange knowledge regarding the datasets which may increase use of data by the users. This design principle accommodates the participation and accountability dimensions of transparency.

The benefits of providing space of interaction for users include allowing users to exchange knowledge regarding datasets, and stimulating more data use from the knowledge exchange. A user may have something they want to discuss with other users regarding how to use datasets or they may even have questions regarding datasets. Allowing users to have these discussions can encourage more dataset use.

4. **Assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage**

This design principles originate from the help/assistance design requirement. The purpose of providing help/assistance is to allow users to understand datasets better, as well as to allow new users to learn on how to use the open government data portal faster. This may stimulate more data use by the users of the portal. FAQ, channel for users to ask questions, tutorials, and examples of data usage are among the common functionalities of open government data portals related to assistance that were found during the design observation. This design principle accommodates the publication of open government data and accountability dimensions of transparency, and the accuracy, completeness, consistency, timeliness, ease of operations, and assistance dimensions of information quality.

The benefit of providing help and assistance in understanding datasets as well as using the open government data portals is that it can lead to more data use by the users of the portal.

5. **Tag datasets with relevant keywords to allow faster searching of datasets**

This design principle is derived from the design requirement of tagging of datasets. Design observation result shows that dataset tagging is implemented in the majority of the observed portals. The purpose in tagging datasets is to allow users to search dataset faster as users can use the keywords to search for datasets as well as filter search results. This design principle accommodates the publication of government data dimension of transparency as well as the accessibility and ease of use dimensions of information quality.

The benefit of tagging datasets is that it allow users to identify the topic of a dataset easily and allow them to find the datasets that they need more quickly.

6. **Allow users to download and request datasets without having to register first**

This design requirement is derived from the design requirement of allowing users to download datasets without having to register first. Findings from the design observation further extends this requirement to allowing users to download datasets without having to register as well. This design principle accommodates the dimension of identity privacy.

The benefit of allowing users to download and request datasets without having to register first is that it may allow more data use and attract more users to use the open data portals. This is because users may see registration as a hindrance in accessing datasets, and not all users will bother to go through the process of registration to download or request a dataset that they need.

7. **Assist users in searching datasets by providing filtering and suggestion feature on the search function**

   Assistance in searching for datasets can allow users to search for dataset faster. The design observation result shows that assistance in searching dataset commonly comes in the form of filtering functionality. However, interview result also shows that suggestion of keywords to search can also help users in searching faster. This design principle accommodates the publication of government data dimension of transparency, query privacy dimension of privacy, and the timeliness, accessibility, and ease of use dimensions of information quality.

   The benefit of filtering and suggestion feature is that it allows users to search the datasets that they need faster. Browsing through a large collection of datasets can be a hassle, and filtering function allows users to view only datasets that match the criteria that they define. Thus, they can find the datasets that they need more quickly.

8. **Provide API when users require to access datasets through external applications**

   This design principle is derived from the design principle of "provide API for dataset access." Design observation result shows that the majority of the observed portals provide API for exploring datasets in the portal. In some cases, the API is provided by the platform used to develop the portal (CKAN and Socrata). However, it is also possible to use custom platform and implement API, as what the Japan statistics portal does. Implementation of API can be context dependent and depends on the need of the users. This design principle accommodates the accessibility and ease of operations dimensions of information quality.

   The benefit of API functionality is that it allows users to access datasets through a programming interface, which allows apps to access the datasets by making use of the API function and provide more flexibility for the apps to provide different services based on the datasets.

9. **Provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer**

   This design principle is derived from the design requirement of using open and machine-processable data format. Design observation result shows that the majority of the observed open government data portals provide their datasets in open and machine processable data formats. The exception is Japan statistics open data portal, which provides its datasets mostly in proprietary data format. Providing datasets in open format ensures that all users can open the datasets without relying on proprietary applications. This design principle accommodates the accessibility and ease of operations dimensions of information quality.

   The benefit of providing datasets in open and machine-processable formats is that it may avoid a lock-in to a proprietary application. Furthermore, many free applications are usually available to access files in an open format. Therefore, it may also increase the possibility of data use compared with using proprietary formats.

10. **Provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets**

    This design principle is taken from the design requirement of providing metadata. Metadata provides context for users to understand the content of a dataset. Among useful context regarding datasets that is viewed as important by the interviewees are data owner, semantic rules, update frequency, and history of changes to datasets. Design observation result shows

that all observed open government data portals provide metadata along with each dataset published. This design principle accommodates the publication of open government data and accountability dimensions of transparency as well as the accuracy, completeness, consistency, timeliness, ease of operations, and assistance dimensions of information quality.

The benefit of providing metadata is that it helps users in understanding what is contained within a dataset. Thus, it may also help increase the possibility of data use by the users.

11. Provide the function to visualize datasets to allow quick viewing of datasets

This design principle is derived from the design requirement of provide dataset visualization. The purpose of having visualization is to help users in viewing the datasets without having to download and process the datasets first. Design observation result shows that only some portals implement the function of visualization, and the functionality is often available for only selected datasets. However, interview result shows that this functionality can be useful in allowing users to view datasets quickly. This design principle accommodate the timeliness, ease of operations, and assistance dimensions of information quality.

The benefit of providing visualization is that it allows users to view datasets more quickly. This can help users in doing operations on the datasets more quickly by accessing them directly on the portal without having to download them first.

12. Use user friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need

This design principle is derived from the design requirement of user friendly interface and the findings from design observation. The purpose of having a user friendly interface is to allow easy navigation which helps users in finding what they are looking for in an open government data portal. Interview result shows that user friendly interface can be characterized by the categorization of datasets into different topics, use of universal icons to allow finding the topics that the users are interested in easier, highlighting of important datasets, use of consistent user interface design between pages, and providing the user interface in the language that the users need. This design principle accommodates the publication of open government data dimension of transparency, and the accessibility and ease of use dimensions of information quality.

The benefit of user friendly interface is that it can provide a good user experience for the users. Bad user experience can lead to the loss of user interest in accessing the portal, while good user interface may attract users to come back to access the portal another time. Furthermore, a user friendly interface can help users in accessing the datasets that they need more easily.

5.2.2. Context-related design principles

Context-related design principles are the design principles that are related to the context surrounding an open government data portal. They are derived from additional findings of the interviews as well as the findings from the design observation. Context-related design principle provide guidelines on the actions that should be taken in designing an open government data portal which are related to the context surrounding the open government data portal, such as regulation and target users of the portal. Context-related design principles are explained below.

1. **Deal with privacy issues before dataset is uploaded to the portal**

   This design principle is derived from the additional findings of the interviews. Most interviewees agree that the measure to deal with privacy issues related to datasets is supposed to be taken before the datasets are published on open government data portals. This means that the burden to protect citizen's privacy due to the publication of open government data should not be placed upon the design of an open government data portals. This design principle accommodates the combining data, lack of control over personal information, and social sorting and discrimination dimensions of privacy.

   The benefit of dealing with privacy issues prior to publishing it on an open government data portal is that it can avoid the possibility of privacy problems due to the publication of open government data.

2. **Take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities**

   This design principle is derived from the additional findings of the interviews and reinforced by the findings from the design observation. The decision to implement certain functionalities in an open government data portal should take into account the regulation, target users, as well as the purpose of the functionality itself. Taking context into account can make sure that the functionalities that are implemented will be the ones that are needed by the users and the data provider. This design principle is related to transparency, privacy, and information quality.

   The benefit of taking into account the context in implementing a portal's functionalities is that it ensures that the functionalities implemented conform to regulation as well as aligned with users' needs.

5.2.3. Process-related design principle

Process-related design principle is the design principle that is related to the development process of an open government data portal. It is derived from the additional findings of the interviews. Process-related design principles provide guidelines on what kind of action should be taken during the development and design process of an open government data portal. From the 15 design principles, one is related to development process and is explained below.

1. **Involve users in early and later development of the portal to match portal design with users' need**

   This design principle is taken from the additional findings from the interviews. User involvement in every stage of an open government data portal development can ensure that requirements of potential users can be accommodated by the open government data portal. The design observation result shows that after the portal is launched, users can still be involved in the continuous improvement of the portal. A good example of this is the ability for users to be involved in the improvement of the US open government data portal through Github. This design principle is related to transparency, privacy, as well as information quality.

   The benefit of involving users in the development of an open government data portal is that it allows the alignment between the functionalities implemented on the portal and the needs of the users.

5.3. Context sensitivity of design principles

Design principles provide general rules and guidelines, but it is not strict in its implementation. This means that portal developer may follow the design principles in developing an open government data portal, but there can be some deviation depending on the situation related to the portal development. One of the design principles suggested that context should be taken into account in implementing functionalities into an open government data portal. If we take a look into the findings from the design observation, there are implementations of design requirements which have high variations or do not conform to the expectation of the design requirements. Furthermore, some variations in the implementation of design requirements such as API are related to the context of the open government data portals.

Meanwhile, there are also design requirements that are implemented quite uniformly by the observed open government data portals. This can be an indication that some design requirements are sensitive to context, and there are design requirements that are less affected by context. As function-related design principles are mainly derived from the designed requirements, the context sensitivity of the design requirements may be applied also to the design principles. This subsection will examine which design principles are sensitive to context, and which are not as sensitive.

Table 14. Context sensitivity of design principles

Category		Design principles
High context sensitivity	[2]	Provide the mechanism for users to give feedback to datasets
	[3]	Stimulate knowledge exchange between users by providing space for interaction between users
	[11]	Provide the function to visualize datasets to allow quick viewing of datasets
Moderate context sensitivity	[1]	Provide the mechanism for users to request datasets and to track whether their request has been approved or rejected
	[5]	Tag datasets with relevant keywords to allow faster searching of datasets
	[8]	Provide API when users require to access datasets through external applications
	[9]	Provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer
Low context sensitivity	[4]	Assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage
	[6]	Allow users to download and request datasets without having to register first
	[7]	Assist users in searching datasets by providing filtering and suggestion feature on the search function
	[10]	Provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets
	[12]	Use user friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need
	[13]	Deal with privacy issues before dataset is uploaded to the portal
	[14]	Take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities
	[15]	Involve users in early and later development of the portal to match portal design with users' need

In general, the design principles can be divided into three categories based on how sensitive it is to context of the open government data portals (Table 14). Further explanation on which design principles fall into which category is as follows.

1. **Design principles with high context sensitivity**

   The implementation of design principles with high sensitivity have high dependency on the context of the open government data portal. This means that how portals follow the guidelines from the design principles can be different depending the context of the portals. Design principles [2], [3], and [11] are part of the design principles that have high sensitivity to context. This is shown by the high variance in the design observation result for the design requirements from which the design principles are derived.

2. **Design principles with moderate context sensitivity**

   The second category of design principles is design principles which have moderate context sensitivity. This means that the implementation of the design principles may have variation from portal to portal depending on the context, but the variance is lower from the first category (high context sensitivity) of design principles. Design principles [1], [5], [8], and [9] falls into this category. The indication of the context sensitivity of the design principles is reflected from the design observation result of the design requirement from which the design principles are derived from. The design observation result for design requirement from which design principles [1], [5], [8], and [9] shows that there is a slight variance in the implementation of the requirements.

3. **Design principles with low context sensitivity**

   The last category of design principles are design principles with low sensitivity to context. This means that the design principles may be implemented in most situations and are less dependent to the context of the open government data portal. Design principles [4], [6], [7], [10], and [12] fits into this category as the design observation result for the design requirements from which the design principles are derived shows little variance. Furthermore, design principles [13], and [14], and [15] can also fit into this category as they can apply in most situations.

5.4. Implementation of the design principles

The design principles provide guidelines for governments and portal developers in designing an open government data portal that can accommodate transparency, privacy, and information quality. Since the nature of principles are not strict, the implementation of design principles can vary from case to case. However, there may be principles that should be prioritized in its implementation. To identify which principles should be prioritized, we should reflect back to the goal of the open government data itself, which in the case of this research topic is to increase government transparency.

If we examine the list of design principles, there are several principles that contributes and can be related to transparency. They can be identified by asking this question for each of the principles: "if the principle is not implemented, will government transparency be difficult to achieve?" If the answer is difficult to achieve, then it may be an indication that the principle should be prioritized. From the function-related principles formulated before, principle [1], [2], [6], [7], [9], and [12] may fall into the category of principles that has to be prioritized. The reason is because these principles correlates with the ability of users to find and access datasets. When users are unable to access datasets, then the goal of increasing government transparency through the publication of open

government data may not be reached. As for the other function-related principles, whether or not they are implemented may not directly affect the ability of users to access datasets. However, their implementation can enhance the ability of users to access datasets, and further contributes to the goal of achieving government transparency. Context and process-related principles are important as well since they provide guidance to ensure that the functionalities implemented are aligned with the needs of the users and the context of the open government data portals.

The design principles are formulated through the several phases conducted in this research project. However, the research project has several limitations (see section 6.2). For example, the literature used in exploring the dimensions of transparency, privacy, and information quality may not be comprehensive enough and the choice of interviewees may affect the comprehensiveness of the design requirements of open government data portals obtained. The incomprehensiveness may then be transferred to each of the phases of the research project. As a result, the design principles, which are the final result of the research project, may be incomplete. Thus, in using the design principles to design, redesign, or evaluate an open government data portals, the possibility of the incompleteness of the design principles should be taken into account.

As what have been discussed in the second chapter of this report, design principles act as general rules and guidelines in designing and developing open government data portals. They are not restrictive, but normative and prescriptive (Bharosa & Janssen, 2015). Thus, governments and portal developers should use it as such. The function-related design principles can provide guidelines on what functionalities should exist in the open government data portal. However, the implementation of the functionalities itself should be decided based on the evaluation of the context of the open government data portals, as what have been described in the context-related and process-related design principles.

In using the design principles, it can also be useful to notice that some design principles may have narrow implementation, while other design principles have wider possibility of implementation. This can be seen in the result of the design observation. Some design requirements are implemented uniformly by the observed open government data portals, while others, although implemented by the majority of the portals, have different implementation. Function-related design principles that may have narrow implementation include design principle [5], [6], [7], [8], [9], and [10]. Meanwhile, the other function-related design principles ([1], [2], [3], [4], [11], and [12]) seem to have wider and more various implementation. Context-related and process-related design principles are generic principles that can have various interpretation. Therefore, principles [13], [14], and [15] may also have wide possibility of implementation.

5.5. Conclusion of chapter 5

This chapter focuses on the formulation of the design principles which are the main objective of this research. Each of the phase of this research contributes to the formulation of the design principles. The literature review provides the dimension of transparency, privacy, and information quality which provide the frame for connecting the design principles to transparency, privacy, and information quality. The interview and design observation provide input for the design principle formulation. The formulation process of the design principles itself is a synthesis of three different inputs: 1) a set of design requirements for open government data portal, 2) additional findings from interview, and 3) findings from the design observation of existing open government data portal.

The synthesis of the three inputs results in fifteen design principles of open government data portal which accommodate transparency, privacy, and information quality assurance. They are: provide the mechanism for users to request datasets and to track whether their request has been approved or

rejected [1], provide the mechanism for users to give feedback to datasets [2], stimulate knowledge exchange between users by providing space for interaction between users [3], assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage [4], tag datasets with relevant keywords to allow faster searching of datasets [5], allow users to download and request datasets without having to register first [6], assist users in searching datasets by providing filtering and suggestion feature on the search function [7], provide API when users require to access datasets through external applications [8], provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer [9], provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets [10], provide the function to visualize datasets to allow quick viewing of datasets [11], use user friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need [12], deal with privacy issues before dataset is uploaded to the portal [13], take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities [14], and involve users in early and later development of the portal to match portal design with users' need [15]. These design principles answer the main question of this research, which is:

"*What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance?*"

Furthermore, the design principles can be divided into three categories based on how sensitive its implementation is to the context of the open government data portals. The three categories are: 1) design principles with high context sensitivity (design principles [2], [3], and [11]), 2) design principles with moderate context sensitivity (design principles [1], [5], [8], and [9]), and 3) design principles with low context sensitivity (design principles [4], [6], [7], [10], [12], [13], [14], and [15]).

Chapter 6
Conclusions and Discussions

This chapter provides the summary of the findings from each phase of this research as well as the answers to the sub-questions, followed by the answer to the main research question which concludes this research. The conclusions section will also discuss the academic contribution and practical relevance of this research. Limitations of research are presented to explain the factors that may affect the generalization of the outcome of this research. The next section will discuss future research questions related to this research. Lastly, the reflection section will discuss the reflection on the choices made during this research and the relevance between the focus of the research to the Master Management of Technology program that serves as the background in conducting this research.

6.1. Conclusions

This section will discuss the results and conclusions of this research, including what the academic contribution and practical relevance of this research are. The objective of this research is to formulate principles for designing open government data portals which incorporates transparency, privacy, and information quality assurance. In the effort to reach the objective of this research, a main research question is formulated, which is the following:

RQ: What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance?

The main research question define the focus of the research. To help in answering the main research question, three sub-questions are developed. The research follows the research design to answer each of the sub-question before arriving to the main research question.

SQ1: What are the dimensions of transparency, privacy, and information quality that are relevant to open government data portal?

Sub-question 1 focuses on the exploring the definition and dimensions of transparency, privacy, and information quality which are relevant to open government data portal design. The dimensions of transparency, privacy, and information quality is used provide a frame on what kind of factors should be focused on in developing the design principles of open government data portals. Sub-question 1 is answered using literature review.

Literature review shows that there are many dimensions of transparency, privacy, and information quality mentioned by scholars. However, not all of the dimensions are related to open government data portals. Therefore, the next task is to analyze each of the dimensions and select the most relevant dimensions of transparency, privacy, and information quality to open government data portal design. After careful analysis is done to each of the dimensions of transparency, privacy, and information quality obtained from literature, it is found that there are 17 dimensions of transparency, privacy, and information quality which are relevant to open government data portal design.

There are three dimensions of transparency that are related to open government data portal design. They are: publication of open government data [1], participation [2], and accountability [3]. There are six dimensions of privacy concern that are related to open government data portal design. Three

dimensions are related to the publication of open government data. They are: combining data [4], lack of control over personal information [5], and social sorting and discrimination [6]. Another three dimensions are related to the use of open government data portals. They are: identity privacy [7], query privacy [8], and location privacy [9]. Lastly, there are 8 dimensions of information quality that are related to open government data portal design. They are: accuracy [10], completeness [11], consistency [12], timeliness [13], accessibility [14], ease of operations [15], assistance [16], and ease of use (of h/w, of s/w) [17].

After finding the relevant dimensions of transparency, privacy, and information quality, the research continues to the next step.

SQ2: What design requirements of open government data portals are relevant to transparency, privacy, and information quality?

The second sub-question seeks to explore what kind of functionalities an open government data portal should have with relation to the dimensions of transparency, privacy, and information quality. Design requirements can determine what should or should not be in an open government data portals. In order to answer SQ1, interviews with five academic researchers from TU Delft are conducted.

The interviewees are selected using several criteria, including they should work as academic researchers in TU Delft, their research focus is about open data, have experience in using an open government data portal, and willing to participate in the interview. To be able to conduct the interview in a systematic way, an interview protocol was formulated. Before an interview is conducted, the interviewers are presented with the dimensions of transparency, privacy, and information quality. During the interview, in a semi-structured manner, the interviewees were asked several questions regarding the functionalities of open government data portals that are related to the dimensions of transparency, privacy, and information quality, as well as how they should be designed.

The interviews resulted in two findings. The first one is a set of twelve design requirements of open government data portals. The design requirements define what kind of functionalities an open government data portal should have in order to accommodate transparency, privacy, and information quality. The twelve design requirements are: dataset request [1], user feedback [2], provide space for interaction between users [3], help/assistance [4], tagging of datasets [5], no registration for download [6], search function [7], provide API for dataset access [8], use open data format [9], provide metadata [10], provide dataset visualization [11], and user friendly interface [12]. Each of the design requirements corresponds to a set of dimensions which are part of the dimensions of transparency, privacy, and information quality.

The second finding of the interview are additional findings that may contribute to the formulation of the design principles, but not directly suggest what functionality should or should not exist in an open government data portal. The second findings include: 1) privacy is an issue that has to be dealt with before the publication of open government data, 2) functionalities implemented in an open government data portal should match the context of regulation, target users, and other situation surrounding the portal, and 3) potential users should be involved in the development process of an open government data portal.

Most interviewees agree that privacy issues due to the publication of an open government data on an open government data portal are not issues that should be tackled by providing specific functionalities in the open government data portals. Instead, measures should be taken by government agencies during the collection and preparation of data to ensure that they do not publish data with personal information to public domain. Interviewees also agree that location privacy is not an issue to open government data portal, and there should be no specific functionality of open government data portal to prevent the concern due to location privacy issue.

The decision to implement functionalities on an open government data portal can sometimes depend on what regulation concerning open data affect the open government data portal. Target users of an open government data portal can also affect whether or not a functionality is suitable to be implemented in the open government data portal. Regulation and target users are part of the context surrounding an open government data portal. The interviewees agree that the functionalities implemented in an open government data portal should match the context surrounding the open government data portal.

Lastly, the interviewees agree that the development process of an open government data portal should involve the potential users of the open government data portal. The involvement of users in the development process can happen in both early stage and later stage of the development of an open government data portal. The purpose of user involvement is to gain input from the users regarding the functionalities and design of the open government data portal and thus making sure that the open government data portal is designed based on users need.

After acquiring findings from the interview, the next step is doing a design observation on existing open government data portals. This process is represented by the third sub-question of this research, which is:

SQ3: How do existing open government data portals implement the design requirements of open government data portals related to transparency, privacy, and information quality?

The third sub-question explores how existing open government data portal implement the design requirements obtained from SQ2. The purpose of the open government data design observation is to find out how existing open government data portals implement the design requirements obtained through the interviews. The observation is done by confirming whether or not the design requirements are implemented by the government data portals. To conduct the observation systematically, a design observation protocol is developed.

The design observation results in three findings. The first finding is the finding that conform with previous findings. There are design requirements which are implemented by the majority of the observed open government data portals. These design requirement include dataset request, help/assistance functionalities, dataset tagging, no registration requirement to download datasets, search function, API, open and machine-processable data format, metadata, and standardized design of user interface. The design observation result also shows that the implementation of a design requirement on an open government data portal depends on the context surrounding the open government data portal. The second finding is the finding that does not conform with previous findings. Some design requirements which the interviewee mentioned as important are not implemented on the observed open government data portals. These include user feedback, space for interaction between users, and visualization. The last finding from the design observation is the

finding that is not described by previous findings. These include the selection of languages provided on the user interface, and the requirement of registration for requesting datasets.

After the design observation is done, the research reached its last phase and has enough knowledge to answer the main research question. The last phase, the formulation of design principles, involves synthesizing the findings from previous phase of the research. The dimensions of transparency, privacy, and information quality will be used as a frame that connects the design principles to the concept of transparency, privacy, and information quality. Meanwhile, the findings from the interview and the design observation will be used as inputs for the design principle formulation process.

There are three set of inputs used in formulating the design principles. They are: 1) a set of design requirements for open government data portal, 2) additional findings from the interviews, and 3) findings from the design observation of existing open government data portal. The synthesis of these three inputs results in fifteen design principles of open government data. They are: provide the mechanism for users to request datasets and to track whether their request has been approved or rejected [1], provide the mechanism for users to give feedback to datasets [2], stimulate knowledge exchange between users by providing space for interaction between users [3], assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage [4], tag datasets with relevant keywords to allow faster searching of datasets [5], allow users to download and request datasets without having to register first [6], assist users in searching datasets by providing filtering and suggestion feature on the search function [7], provide API when users require to access datasets through external applications [8], provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer [9], provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets [10], provide the function to visualize datasets to allow quick viewing of datasets [11], use user friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need [12], deal with privacy issues before dataset is uploaded to the portal [13], take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities [14], and involve users in early and later development of the portal to match portal design with users' need [15].

Looking at the design principles, the fifteen design principles can be further divided into three categories based on how context sensitive their implementation are on an open government data portal design. The three categories are: 1) design principles with high context sensitivity (design principles [2], [3], and [11]), 2) design principles with moderate context sensitivity (design principles [1], [5], [8], and [9]), and 3) design principles with low context sensitivity (design principles [4], [6], [7], [10], [12], [13], [14], and [15]). High context sensitivity implies that how portal developer follow the guidelines from the design principles has high dependency on the context surrounding the open government data portal. Meanwhile, low context sensitivity means that how portal developer follow the guidelines from the design principles has low dependency on the context surrounding the open government data portal. Moderate context sensitivity means that the dependency to context is not too high, but not too low as well.

Finally, the set of fifteen design principles answer the main research question and is the final deliverable of this research. The deliverable takes the form of Table 13 (see section 0), which shows the design principles, examples of implementation, and which dimensions of transparency, privacy, and information quality they accommodate.

6.1.1. Academic contribution

This research contributes to the field of open data in several ways. Firstly, it fills the knowledge gap on how to design an open government data portal that can accommodate transparency, privacy, and information quality. As what has been discussed in the first chapter, existing principles of open government data are centered on the data itself and there is a need for explicit design principles of open government data portals. This research fills the knowledge gap by delivering fifteen principles of open government data portal design that accommodates transparency, privacy, and information quality.

Secondly, literature says that there is a need for both fundamental and applied open data research, especially related to the conceptualization of transparency and privacy in open data (Janssen & van den Hoven, 2015). This research contributes in enriching the body of knowledge of open government data by delivering the design principles of open government data portal design which are based on empirical evidence. Furthermore, the knowledge provided by the result of this research can be used as a stepping stone to further explore the field of open data.

Next, this research contributes to the field of open data and design science research by providing a method or approach to derive design principles. The approach used in this research is a modification of the approach used by Bharosa & Janssen (2015), Lourenço (2015), and Fedorowicz, et al. (2014). The use of the approach from the modification of the three approaches proved to be able to deliver a a set of design principles.

Finally, this research contributes to the body of literature of transparency, privacy, and information quality by identifying which dimensions of the three values are relevant to the design of open government data portals. Specifically, literature review phase of this research project identifies which dimensions of transparency mentioned by Bauhr & Grimes (2012), Hollyer et al. (2014), Kim (2009), and Matheus & Janssen (2015), which dimensions of privacy mentioned by Smith, et al. (1996), Borgesius, et al. (2015), and Martínez-Ballesté, et al. (2013), and which dimensions of information quality mentioned by Batini, et al. (2009) and compiled by Lee, et al. (2002) are relevant to the design of open government data portals.

6.1.2. Practical relevance

The deliverable of this research can have several relevance in the development of open government data portals. The design principles of open government data portal design provide guidelines for governments and portal developers in designing, redesigning, and evaluating an open government data portal that can accommodate transparency, privacy, and information quality. Firstly, it can help both governments and portal developers in identifying which functionalities they should provide in an open government data portal. As these guidelines are not strict and do not delve deep into the technical side of open government data portals, governments and portal developers have flexibility in translating the design principles into functionalities of open government data portals. Furthermore, the examples of implementation provided in the deliverable of the research can provide governments and portal developers with inspiration on how to actually translate the design principles into functionalities.

Secondly, the design principles can help governments in choosing the right open data portal platform which they can use to develop their open government data portals. Currently there are several open data portal platforms available for governments to use. As an example, two of the observed open government data portals in this research use the CKAN platform, which is open source, and one observed open government data portal uses Socrata, which is proprietary. Using an open data portal platform to develop an open government data portal can reduce the development

effort compared to building an open government data portal from scratch. If a government chooses to use an open data portal platform, it needs to focus on customizing the existing functionalities provided by the platform to suit the government's needs. The design principles of open government data can help governments select the right open data portal platform by providing a checklist of function-related principles, which are derived from the design requirements of open government data portals. This checklist can be used to identify which functionalities the government needs and check whether it is available on the open data portal platform. The more the open data portal conform to the function-related design principles, the less customization will be needed by the government. This means that the government can put less effort into customizing the open data portal to suit its need.

6.2. Limitations

The deliverable of this research is intended to be relevant both for the academic field and the practical field. However, in conducting this research, there are several limitations that may affect the generalizability of the findings of the research. Firstly, the concepts discussed in this research are all broad. To fully grasp the concept of transparency, privacy, and information quality and combine it under the perspective of open government data requires tremendous effort in studying the abundant literatures in each of the academic fields. However, this research is constrained by time, and thus, a deep exploration into the three concepts is not possible. Therefore, the use of literature in the literature review phase may be limited and it can affect the result of the literature review, specifically the dimensions of transparency, privacy, and information quality by providing an incomplete view on the dimensions of transparency, privacy, and information quality. Since the literature review is at the beginning of the research process, this limitation may eventually affect the next phase of the research.

Secondly, the research focuses more on the demand side of open government data. This is reflected in the choice of interviewees, which all are academic researchers on open data. Four out of five academic researchers selected for the interviews have little experience in the development of an open government data portal. This may cause them to take the perspective of a user of open government data in answering the interview questions. Therefore, it may be the case that the open government data portal design requirements are more focused on the users' perspective. Meanwhile, the perspective of government as the data provider and portal developers who interact with both government and users is discussed less in the interview. Government's perspective can provide more insight on the design requirements of open government data portal based on government's needs. On the other hand, a portal developer's perspective can give more insights into the design requirements based on the technical expertise of the developers. To have a complete perspective for the design requirements, government's point of view as well as portal developer's perspective must also be taken into account.

6.3. Future research

The previous subsection provides some clues about the limitation of this research, and what can be pursued in further research based on the limitations. Since this research focuses on the user perspective, it would be interesting to see how government's and open data portal developer's perspective would enrich the result of this research. Therefore, a research on how government and portal developers view the design requirements or a research on how data should be designed based on the perspective of the government or portal developers can be an interesting research. Future research may focus on the research question such as "*What principles can be used in designing open government data portals so that it leads to transparency, privacy, and information quality assurance based on governments' perspective*".

Another possibility for future research can be explored by modifying the methodology of this research. One of the phases of this research is design observation. Ideally, the design observation would be followed by validation from the interviewees. This is because the design requirements that are used as the basis for the design observation originated from the interviews. Another possibility to enhance the research would be by validating the result of the design observation though interviewing people who are involved in developing the portals. For example, one of the findings from the design observation is that there are design requirements of open government data portals which are not implemented in the majority of the portals. Interviewing people who are involved in the development of the portal can provide more insight from the developer's side regarding why the requirements are not fulfilled by the portals. Therefore, future research can focus on the research question such as *"How do portal developers view the design principles of open government data portals?"*

Lastly, it is pointed out as one of the principles that context can play a role on the implementation of the functionalities of an open government data portal. However, the extent to which context affect the implementation of an open government data portal is not fully explored. Therefore, future research can focus on a research question that explore this problem, such as *"How does the context of an open government data portal affect its functionalities?"*

6.4. Reflection

During the course of the research, several decisions are made which may affect the result of the research. This subsection will focus on the reflection of those decisions, the project process, and how this research is related to the program of Master Management of Technology.

1. Reflection on the choices made within the research

The process of conducting a research involves choosing from many possible choices that can affect the course of the research as well as the result of the research. Some of the choices made in conducting this research include:

a. Selecting portals for design observation

The third phase of this research involves observing existing open government data portals in order to gain additional insight which can contribute to the formulation of the design principles of open government data portals. Five open government data portals were selected to be observed, with a selection criteria which focus on selecting as diverse open government data portals as possible. The thought was that the diversity of the selected portals can help in generating design principles that can be applied in more diverse situations. The choice of open government data portals to observe may have contributed to some of the outcomes in this research. For example, the selection of portal allows the observation of open data portals which are different in context, and led to the findings that context does affect the implementation of functionalities in an open data portal. If the portals selected were less diverse, this kind of finding may not be found from the design observation. However, although the resulting design principles are generally applicable for most open government data portal design, the influence from the selection of portals observed may be minor at best.

b. Selecting interviewees

The interviewees were selected from the TU Delft academic researchers who work on open data. The reason for this selection is for practical reasons, as well as because TU Delft is one of the leaders in open data research. However, there were other options other than academic

researchers. The stakeholders of open government data portals may include governments, portal designers and developers, as well as users who frequently access datasets from the portals. These people can also provide inputs for the research, but due to time constrain, it was impossible take all the stakeholders' perspective, and only the academic researchers were selected as the interviewees.

The consequence of this selection is that only one perspective is taken as inputs for the research, which is the user perspective as the interviewees mostly took the perspective of the users to explain what functionalities an open government data portal should or should not have. The design observation was conducted partly to bridge the gap between users' perspective and the perspective of other stakeholders. However, it may not be enough to represent the actual perspective of the governments or portal designers and developers. Thus, the selection of the interviewees may somehow reduce the completeness of the design principles due to the incomplete perspective used as an input of the research.

c. Evaluation of dimensions of information quality which are related to open government data portal design

During the literature review process, an evaluation is conducted on the dimensions of transparency, privacy, an information quality to determine whether or not they are related to open government data portals. The evaluation include selecting only accessibility information quality dimensions as they are related to how data is being accessed, and the purpose of an open government data portal itself is to provide access to datasets. Furthermore, other information quality dimensions (intrinsic, contextual, and representational information quality) are not chosen as they are more related to the data itself instead of the access to data. However, in addition to the accessibility information quality dimensions, four other information quality dimensions including accuracy, completeness, consistency, and timeliness are included to cover the possibility that content or data also affects open government data portal design.

The resulting design principles shows that accessibility information quality does affect open government data portal design, as many of the principles are related to the accessibility information quality. Meanwhile, some design principles are also related to accuracy, completeness, consistency, and timeliness. These design principles are mostly related to the effort to help users in understanding the content of a dataset, such as principle [4] and [10]. This means that content of data does affect the functionalities of an open government data portal, but the effect is limited to the functionalities that help users in understanding the content of a dataset.

2. Reflection on project process

The graduation project is a semester long research project which is a mandatory project in order to graduate from the program of Master Management of Technology under the faculty of TPM, TU Delft. During the course of the research, there are some lessons that the author learned. The early phase of a research is one of the most important part of the research as it determines how the research will go on. Therefore, defining the scope of the research should be done carefully while taking into account the consequences that may happen in later stage of the research. However, based on the experience of the author, it is best not to stay too long and think too much about what may or may not happen in later stages of the research. This is because we have limited information about the course of the research at the beginning of the research. Instead of allowing one to get stuck in one phase of the research, it is better to continue with what we have and reorganize later in

the future when we have more complete sense and information about the direction of the research, if necessary.

Secondly, during the course of the research, problems may also occur which causes us to doubt our research direction. The author learned that during this situation, it is best to stay on track and stick to the plan that we have made in the early stage of the research. Thinking about changing the course of the research, for example by exploring other similar topic, may seem easy. However, later on, the same problem may plague us and we can have another doubt with the new direction of the research. Staying on track during doubt, while thinking of an alternative approach to the same objective, can result in better chance to complete the research more efficiently. In every situation, planning and thinking realistically should help in determining how decisions should be made and carried out.

3.   Relationship with Master Management of Technology program

The Management of Technology (MoT) program is focused on preparing future technology managers who are competent in their fields. It is essential that technology managers are able to demonstrate the knowledge that they obtain from the MoT program. One of the focus of MoT program is to give understanding on how technological products are designed and developed. This research, aiming at formulating design principles for open government data portals, is in line with the focus of MoT program, as an open government data portal is a technological product which needs to be managed and improved continuously. In the context of open government data portals, governments can be seen as corporations which are required to obtain a specific technology (an open government data portal) in order to reach its goals (increase transparency). Governments can have several choices in acquiring the technology, such as developing the open government data portal in-house or using portal templates offered by external developers. This research can help governments choose the right product from external developers and making sure that their open government data portal is designed correctly by providing the knowledge on how it should be designed.

The ICT Management and Design specialization of the Management of Technology program provides various knowledge regarding the management and design of ICT architecture. This specialization is at the foundation in developing this thesis. The design research that was taught in the ICT Design course provide basic knowledge about the design process for ICT-based artifacts, which can be implemented in designing an information system such as an open government data portal.

This research also reflects the contents of some of the courses from the MoT curriculum. The course of Social and Scientific Values discusses how ethics and moral become part of the management of technology. The effort to take into account the privacy of citizens as well as the users in the design of open government data portal stems from the knowledge gained from this course. This effort reflects the intention to value citizens' and users' rights to privacy. This also proves that technology can be value-laden, as what have been taught in the Technology Dynamics course. Part of the work done in this research is finding out how the value of transparency, privacy, and information quality can be translated into functionalities in an open government data portal. Therefore, the development of design principles that accommodates transparency, privacy, and information quality strongly reflects these teachings from the MoT curriculum.

Finally, the Research Methods course taught as part of the MoT curriculum plays a significant role in understanding what qualitative research is and how it can be done. The Research Methods course provides the knowledge regarding how to choose between qualitative and quantitative research, the various steps in doing qualitative research, and the instruments that can be used to collect data. This

knowledge is used in the planning phase of this research, and ensure that the correct research instrument is selected for this research.

Bibliography

Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis, 16*(1), 3-9.

Akatani, T. (2015). Projects of open data for Official Statistics of Japan. *Workshop on the Communication of Statistics*, (pp. 1-9). Washington DC, USA.

Ballou, D., & Pazer, H. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management Science, 31*(2), 150-162.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys, 41*(3), 16:1-16:52.

Bauhr, M., & Grimes, M. (2012). What is Government Transparency? *The Quality of Government Institute: Working Paper Series*.

Behkamal, B., Kahani, M., Bagheri, E., & Jeremic, Z. (2014). A metrics-driven approach for quality assessment of linked open data. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 64-79.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*, 264-271.

Bharosa, N., & Janssen, M. (2015). Principle-Based Design: A Methodology and Principles for Capitalizing Design Experiences for Information Quality Assurance. *Homeland Security & Emergency Management, 12*(3), 469-496.

Borgesius, F. J., Eechoud, M. v., & Gray, J. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Technology Law Journal, 30*(3), 2073-2131.

Cherns, A. (1976). The Principles of Sociotechnical Design. *Human Relations, 29*(8), 783-792.

Choenni, S., van Dijk, J., & Leeuw, F. (2010). Preserving privacy whilst integrating data: Applied to criminal justice. *Information Polity, 15*, 125-138.

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly, 31*, S10-S17.

data.gov - Wikipedia. (2016). Retrieved September 15, 2016, from Wikipedia: https://en.wikipedia.org/wiki/Data.gov

Dawes, S. S., & Helbig, N. (2010). Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency. *EGOV 2010* (pp. 50-60). Lausanne, Switzerland: Springer Berlin Heidelberg.

Delone, W., & McLean, E. (1992). Information systems success: the quest for the dependent variable. *Information systems research, 3*(1), 60-95.

Fedorowicz, J., Sawyer, S., Williams, C. B., Markus, M. L., Dias, M., Tyworth, M., . . . Schrier, R. (2014). Design observations for interagency collaboration. *Government Information Quarterly, 31*, 302-316.

Geiger, C. P., & von Lucke, J. (2012). Open Government and (Linked) (Open) (Government) (Data). *Journal of e-Democracy and Open Government, 4*(2), 265-278.

Gilb, T. (1997). Towards the engineering of requirements. *Requirements Engineering, 2*, 165-169.

Goodhue, D. (1995). Understanding user evaluations of information systems. *Management Science, 41*(12), 1827-1844.

Hart, C. (1998). *Doing a literature review: releasing the social science research imagination.* London, UK: Sage Publications.

Herriott, R. E., & Firestone, W. A. (1983). Multisite Qualitative Policy Research: Optimizing Description and Generalizability. *Educational Researcher, 12*, 14-19.

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2014). Measuring Transparency. *Political Analysis, 22*(4), 414-434.

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce, 26*, 14-40.

Housel, T. J., Sawy, O. A., & Donovan, P. F. (1986). Information Systems for Crisis Management: Lessons from Southern California Edison. *MIS Quarterly, 10*, 389-400.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly, 28*, 446-456.

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*, 363-368.

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.

Jarke, M., & Vassiliou, Y. (1997). Data warehouse quality: a review of the DWQ project. *Proceedings of the Conference on Information Quality*, (pp. 299-313). Cambridge, MA.

Kim, S. (2009). *Management Strategy for Local Governments to Strengthen Transparency in Local Governance.* United Nations, United Nations Project Office on Governance. Seoul: UNPOG Publication.

Krishnamurthy, R., & Awazu, Y. (2016). Liberating data for public value: The case of Data.gov. *International Journal of Information Management, 36*, 668-672.

Lee, G., & Kwak, Y. H. (2011). Open government implementation model: a stage model for achieving increased public engagement. *12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 254-261). College Park, Maryland: Association fo Computing Machinery.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management, 40*, 133-146.

Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science, 9*, 181-212.

Linders, D., & Wilson, S. C. (2011). What is Open Government? One Year after the Directive. *12th Annual International Conference on Digital Government Research* (pp. 262-271). College Park, MD: Associaion ofr Computing Machinery.

Lindman, J., Kinnari, T., & Rossi, M. (2014). Industrial open data: Case studies of early open data entrepreneurs. *47th Hawaii International Conference on System Science* (pp. 739-748). Waikoloa, Hawaii: IEEE.

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

Luna-Reyes, L. F., Bertot, J. C., & Mellouli, S. (2014). Open Government, Open Data and Digital Government. *Government Information Quarterly, 31*, 4-5.

Lyon, D. (2003). Surveillance as Social Sorting: Computer Codes and Mobile Bodies. In *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination* (pp. 13-30).

Martínez-Ballesté, A., Pérez-Martínez, P. A., & Solanas, A. (2013). The Pursuit of Citizens' Privacy: A Privacy-Aware Smart City Is Possible. *IEEE Communications Magazine, 51*(6), 136-141.

Matheus, R., & Janssen, M. (2015). Transparency Dimensions of Big and Open Linked Data. *14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society* (pp. 236–246). Delft: Springer International Publishing.

Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: connecting vision and voice. *International Review of Administrative Sciences, 78*(1), 10-29.

Merriam-Webster. (2016). *Portal | Definition of portal by Merriam-Webster*. Retrieved June 3, 2016, from Merriam Webster: http://www.merriam-webster.com/dictionary/portal

O'Hara, K. (2012). Data quality, government data and the open data infosphere. *AISB/IACAP World Congress 2012: Information Quality Symposium.* Birmingham, GB: The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Open Knowledge Foundation. (2016). *Open Definition 2.1*. Retrieved March 24, 2016, from Open Definition: http://opendefinition.org/od/2.1/en/

Piotrowski, S. J., & Van Ryzin, G. G. (2007). Citizen Attitudes Toward Transparency in Local Government. *The American Review of Public Administration, 37*(3), 306-323.

Principles. (2016). Retrieved March 16, 2016, from open data charter: http://opendatacharter.net/principles/

Richardson, G. L., Jackson, B. M., & Dickson, G. W. (1990). A Principles-Based Architecture: Lessons from Texaco and Star Enterprise. *MIS Quarterly, 14*, 385-403.

Rowe, G., & Frewer, L. J. (2005). A Typology of Public Engagement Mechanisms. *Science, Technology & Human Values, 30*(2), 251–290.

Scholl, H. J. (2014). The EGOV Research Community: An Update on Where We Stand. *13th IFIP WG 8.5 International Conference, EGOV 2014* (pp. 1-16). Dublin, Ireland: Springer Berlin Heidelberg.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly, 20*(2), 167-196.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

The 8 Principles of Open Government Data. (2016). Retrieved March 16, 2016, from OpenGovData: https://opengovdata.org/

TOGAF. (2011). *TOGAF 9.1*. Retrieved March 17, 2016, from OpenGroup: http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap23.html

Verschuren, P., & Doorewaard, H. (2010). *Designing a Research Project* (2nd edition ed.). The Hague: Eleven International Publishing.

Walls, J. G., Widmeyer, G. R., & Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research, 3*(1), 36-59.

Wand, Y., & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Yin, R. K. (2003). *Case Study Research: Design and Methods* (3rd ed.). Sage.

Zmud, R. (1978). Concepts, theories and techniques: an empirical investigation of the dimensionality of the concept of information. *Decision Sciences, 9*(2), 187-195.

Zuderwijk, A. (2015). *Open Data Infrastructures: The design of an infrastructure to enhance the coordination of open data use (Doctoral Dissertation).* Tehnische Universiteit Delft. Delft: Uitgeverij BOXPress.

Zuiderwijk, A., & Janssen, M. (2014). The negative effects of open government data - Investigating the dark side of open data. *The 15th Annual International Conference on Digital Government Research*, (pp. 147-152). Aguascalientes, Mexico.

Appendix A - Selected Open Government Data Portals

The following are the selected cases from the Open Government Course of the 2016 MOOC which are used in selecting the open government data portal for the open government data portal design observation of this research. The cases are presented as how they are submitted by the participants of the MOOC.

Case 1 – Fingal County Open Data Portal (Ireland)

Taken from category: 1 – Data and Information Sharing

Title: Fingal County (Ireland) Council's Open Data Portal

Objective: The objective of Fingal County Council's Open Data Portal is "enable citizens to access data relating to Fingal"

Actors & Stakeholders: citizens, specifically politicians, civil servants, businesses, researchers, educators, journalists, students, app developers, entrepreneurs, activists, economists, urban-planners, other NGOs and general public

General Description: Fingal Open Data has been created by Fingal County Council to enable citizens to access data relating to Fingal. We will provide access to data from Fingal County Council and other public sector organisations that publish Fingal related data. Fingal Open Data was launched in November 2010 and is the first Open Data website in Ireland. Fingal is located in north County Dublin, Ireland and is the third largest local authority area in the country by population.

Technologies: Internet (data accessed via browser or mobile apps), Semantic data: compliant with DCAP-AP Data Formats: CSV, XML, KML (machine readable formats)

Geographical coverage: Data on Fingal County which is located in north County Dublin

URL: http://data.fingal.ie/

Case 2 - Portal Site of Official Statistics of Japan

Taken from category: 1 – Data and Information Sharing

1) TITLE

e-Stat; Portal Site of Official Statistics of Japan

2) OBJECTIVE OF THE CASE

The Site aims to disseminate statistical information collected from ministries and agencies of government of Japan.

3) ACTORS INVOLVED

- Primary stakeholders

  - Statistics Bureau of Japan produces and coordinates e-Stat;
  - National Statistics Center manages the operation of e-Stat;

- Statistical departments of line ministries and agencies provide statistical information; and
- Government, businesses, economists, researchers, journalists and citizens use e-Stat.

- Secondary stakeholders

- Politicians, educators, etc. but they could be primary stakeholders as well.

4) GENERAL DESCRIPTION

Since 2008 e-Stat has delivered one-stop service for official statistics of the Japanese government. e-Stat provides general public with statistical data, geospatial data of small area statistics (sub municipality and gridded square), schedule of release, etc. In addition to Excel format and CSV file, e-Stat launched API service for statistical data in the late 2014.

5) GEOGRAPHICAL COVERAGE

Statistics data disseminated by e-Stat are those of Japan, and are available to worldwide in Japanese language, some of the contents are in English.

6) REFERENCES

http://www.e-stat.go.jp/

Projects of Open Data for Official Statistics of Japan, Toshihiko Akatani (National Statistics Center, Japan), Workshop on the Communication of Statistics, 27–29 April 2015, Washington D.C., USA

Case 3 - Kenya Open Data Portal

Taken from category: 1 – Data and Information Sharing

1. Title: Kenya Open Data Portal

2. Objective: To make core government development, demographic , statistical and expenditure data available in a useful digital format for researchers, policymakers, ICT developers and general public.

3. Actors: Primary stakeholders include The Government of Kenya, Kenya ICT Board, Kenya National Bureau of Statistics, World Bank and Socrata. Secondary stakeholders include researchers, policymakers, ICT developers and general public.

4. General Case Description: Making Public Government data sets accessible for free to the public in easy reusable formats, supporting the Government's drive to actively inform citizens and to be accountable.

5. Technologies: Web, visualizations, API access and multi-format downloads.

6. Geographical Coverage: All the Counties in Kenya

7. References: https://opendata.go.ke/

Case 4 – US open government

Taken from category: 1 – Data and Information Sharing

1. Title: Open Government

2. Objective: to make information about government operations more readily available and useful as a component for more efficient and transparent government

3. Stakeholders: a. Primary: US government agencies (federal, state, local), researchers, citizens b. Secondary: businesses, foreign governments, journalists/bloggers

4. General description: Data.gov, the central site for U.S. Government data, is an important part of the Administration's overall effort to open government. It allows to find and use a veritable treasure trove of data from many government agencies. The data is available as downloadable datasets as well as through API(s); a number of interactive visualizations/apps are created and available through the site.

5. Technologies: The site - Data.gov - is powered by two open source applications, CKAN and WordPress, and it is developed publicly on GitHub; some. The project follows the Project Open Data Schemata (https://project-open-data.cio.gov/v1.1/schema/), presenting a set of required fields foe every dataset available on the Data.gov. The data is downloadable in CVS and Excel formats, exposed through a series of prebuilt visualizations as well as API(s) geared towards developer community (actual technology choices are left to the dataset owners)

6. Geographical coverage: currently offers open data sets for 40 US states, 46 US cities and counties, 52 individual countries (besides the USA), 146 international regions

7. URL a. https://www.data.gov/open-gov/ b. https://open.whitehouse.gov/

Case 5 - Open data Rio de Janeiro (Brazil)

Taken from category: 1 – Data and Information Sharing

Case Title: Open data Rio de Janeiro

Case objective: • The objective of Portal Open data Rio de Janeiro is for greater openness and transparency by empowering the right to access to information and allowing the citizen monitor how Rio Municipality is spending public resources

Case Actors: • Primary stakeholders: Local government, citizens, tourists, public entities, non-profit agencies, companies, etc. • Secondary stakeholders: Developers, journalists, policy-makers, politicians, etc.

General Case Description: • The City of Rio de Janeiro's Open Data Portal, is the tool provided by the municipal government so that interested parties can find and use data easily, allowing citizens a better understanding of municipal government, access to public services, control of public accounts, participation in planning, policy development and a better knowledge of the city. • The website prides itself on simplicity and organization so that you can easily find the data and information you need. • The portal also aims to promote dialogue between actors in society and the government to think about the best use of the data for the sake of a better society. • The portal provides access to

information, among others, municipal road transport provided by the GPS installed on buses circulating in the city of Rio de Janeiro, health, education, etc.

Geographical coverage: • Rio de Janeiro, Brazil.

References: • http://data.rio/

Appendix B – Interview Protocol

Introduction

Transparency, privacy, and information quality assurance are profound issues that often emerge due to the publication or use of open data. Realizing the three values are important, and open government data portals should be designed to accommodate the values. However, a guideline on how the open government data portal may be designed with regard to the three concepts is not yet available. The objective of this research is to formulate principles for designing open government data portals which incorporates transparency, privacy, and information quality assurance.

In this research, transparency is defined as the ability to find out what is going on inside a government by observing data published by the government through open government data portals. Privacy is defined as the state of being free from identity disclosure due to the release or use of open government data and the use of an information system. Meanwhile, information quality refers to the information's fitness for use.

This protocol is written in order to provide guidelines for the research so that the objective of the research may be achieved with in a systematic way. The protocol describes the design of the interview, including interviewee selection criteria, interview preparation, and interview questions, and the process of data analysis and reporting.

Research approach

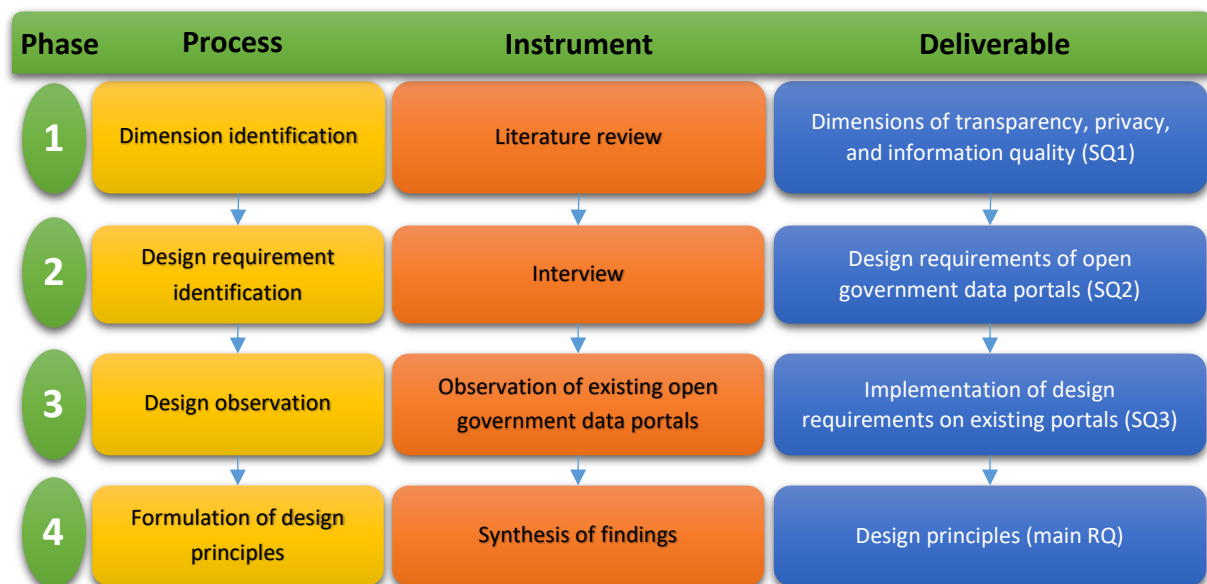| Phase	Process	Instrument	Deliverable
1	Dimension identification	Literature review	Dimensions of transparency, privacy, and information quality (SQ1)
2	Design requirement identification	Interview	Design requirements of open government data portals (SQ2)
3	Design observation	Observation of existing open government data portals	Implementation of design requirements on existing portals (SQ3)
4	Formulation of design principles	Synthesis of findings	Design principles (main RQ)

Figure 1. Research approach

This research adapts the methods used by several scholars in order to arrive to the design principles for open government data portals. Bharosa & Janssen (2015) uses interview in order to formulate design principles. Dimensions are used in order to be able to focus on the problems presented to the interviewees. A dimension can be defined as an aspect or attribute which belongs to a concept. For

example, the dimensions of information quality may include accuracy and timeliness. Lourenço (2015) performs open government data portal observation using design requirements as the focus of the observation. The design requirements are developed from characteristics or dimensions of transparency in literatures. Meanwhile, Fedorowicz, et al. (2014) uses multiple case study to arrive to design observations. Design observations describe the current condition of an architecture and provide insight into existing architectural or social conditions (Fedorowicz, et al., 2014). Furthermore, Fedorowicz, et al., 2014 also argue that design principles can arise from design observations.

In this research, design principles are derived from design observations as well as expert interviews. The expert interviews provide design requirements that will be the focus of the observations on existing open government data portals. The interview will also be used to obtain insight from experts on open data on how open government data portals should be designed with regard to transparency, privacy, and information quality. Literature review provides the dimensions of transparency, privacy, and information quality which are presented to the interviewees in order to obtain relevant design requirements.

Open government data portal selection

5 open government data portals are selected from cases submitted by the participants of the Open Government course of the 2016 Massive Open Online Course. These open government data portals will be observed in order to evaluate whether they comply with certain design requirements (obtained from interviews). The cases are selected using the following criteria:

1. Cases are from MOOC
2. Related to data sharing
3. Provides sufficient information
4. Relevant to the concept of transparency, privacy, and information quality
5. Provides sufficient diversity

The five open government data portals selected are chosen so that they have a geographical diversity and a diversity in the level of government who owns the portals. This means that open government data portals from different countries worldwide as well as different level of governments (local, regional, and national) will be selected. This is done to allow the design principles resulting from the research to have wider relevance both geographically as well as in relation to the level of government.

Data collection method

There are two primary data collection processes involved in this research. The first data collection is conducted through interview. The purpose of the interview is to find out what open government data portal design requirements are correlated to the dimensions of transparency, privacy, and information quality. The second data collection method involves observation of open government data portal design in selected open government data portals. The observation is done to discover how the design requirements obtained through the interview are implemented in the selected open government data portals.

Interview protocol

This section discusses the protocol for the interview process, including the criteria used in selecting the interviewees, the structure of the interview, the questions asked during the interview, and the preparation before conducting the interview.

Interviewee selection criteria

The interviewees are selected based on the following criteria:

5. Works as an academic researcher in TU Delft
6. Focuses on open data research
7. Has experience in using an open government data portal
8. Willing to participate in the interview

Open government data portal design requirements which are obtained through the interview are derived from the dimensions of transparency, privacy, and information quality. Therefore, in order to derive the open government data portal design requirements, understanding of the concepts of transparency, privacy, and information quality, which are some of the key concepts in open data research, is imperative. Academic researchers who focuses on open data research are chosen as the interviewees since they fit into this requirement. Moreover, since the interview result will be used to observe cases from diverse countries and government levels, a selection of interviewees' with experience on different open government data portals each is also desired in order to have more diverse view on design requirements of open government data portals.

TU Delft has many experts in open data research and it is one of the highest contributing institutions in open data research (Scholl, 2014; Hossain, Dwivedi, & Rana, 2016). Due to the expertise of the researchers of TU Delft as well as practical reasons, one of the criteria used in selecting interviewees is that they work as an academic researcher in TU Delft. Some academic researchers may also have experience in open government data portal development or design. In this case, the experience of the academic researchers may provide a valuable input in reaching the objective of the interview. Therefore, if possible, the selection of interviewees will focus on academic researchers who have prior experience in designing or developing an open government data portals. However, this is not a strict criterion, and as such, will not be included in the interviewee selection criteria.

The number of interviews involved in this research is not fixed. Although the interview process follows the initial list of potential interviewees, the interview process may be stopped or a number of interviews may be added depending whether saturation has been achieved or not. However, to ensure that sufficient results are obtained, a minimum number of four interviews is required for this research.

Interview preparation

These steps are taken in order to prepare for the interview:

1. Conducting literature review on transparency, privacy, and information quality in order to gain proper understanding on the concepts as well as their dimensions.
2. Finalizing interview questions.
3. Building a list of potential interviewee based on the interviewee selection criteria.
4. Contacting the potential interviewees and setting up interview schedule.

5. Preparing tools for interview, including a sound recorder, printed interview questions, printed list of transparency, privacy, and information quality dimensions, as well as a notebook and a pen for taking notes during the interview.

Interview structure and questions

The interview will take around a one hour time and will be conducted in a semi-structured manner. The interview is divided into four sections: introduction of the interviewer, introduction of the interviewee, transparency section, privacy section, and information quality section. At the beginning of the transparency, privacy, and information quality sections, the definition of the corresponding concept (i.e. transparency, privacy, or information quality) will be explained to the interviewee so that the interviewee will have the same perspective on the concept as the interviewer. The interview will be recorded using a recording device under the permission of the interviewee.

Introduction of the interviewer

The interview will start with the introduction of the interviewer. The following subjects will be explained by the interviewer:

1. Name of the interviewer
2. Brief introduction to the interview topic and objective
3. Confidentiality agreement
4. Asking interviewee's permission to record interview

Section 1: Introduction of the interviewee

1. What is your name?
2. What is your age?
3. From which country are you?
4. What research field are you primarily associated with?
5. How long have you been working on the field of open data?
6. How do you associate yourself with the field of open data?
7. What experience do you have in open government data portal design?

Section 2: Transparency

The interview will focus on and iterate each of the dimensions of transparency. For each of the dimensions, the definition of the dimension will be explained and the interviewee will be asked the following questions:

1. Identification of transparency dimensions on open government data portals
   a. Are you familiar with the concept of <name of dimension>?
   b. Which features and functionality of an open government data portal are related to <name of dimension>?
2. Open government data portal design principles
   a. What design requirements should an open government data portal comply with in regard to <name of dimension>
   b. What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 2, the following question will be asked to the interviewee:

3. Do you have any more comments to add regarding transparency dimensions?

Section 3: Privacy

The interview will focus on and iterate each of the dimensions of privacy concerns. For each of the dimensions, the definition of the dimension will be explained and the interviewee will be asked the following questions:

1. Identification of privacy concern dimensions on open government data portals
    e. Are you familiar with the concept of <name of dimension>?
    f. Which features and functionality of an open government data portal is related to <name of dimension>?
2. Open government data portal design principles
    g. What design requirements should an open government data portal comply with in regard to <name of dimension>
    h. What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 3, the following question will be asked to the interviewee:

3. Do you have any more comments to add regarding privacy dimensions?

Section 4: Information quality

The interview will focus on and iterate each of the dimensions of information quality. For each of the dimensions, the definition of the dimension will be explained and the interviewee will be asked the following questions:

1. Identification of information quality dimensions on open government data portals
    e. Are you familiar with the concept of <name of dimension>?
    f. Which features and functionality of an open government data portal is related to <name of dimension>?
2. Open government data portal design principles
    g. What design requirements should an open government data portal comply with in regard to <name of dimension>
    h. What kind of problems can arise when an open government data portal does not comply with the design requirement?

At the end of section 4, the following question will be asked to the interviewee:

3. Do you have any more comments to add regarding information quality dimensions?

Data analysis and reporting

Transcription

After the interviews are done, the recording of the interviews will be transcribed. Depending on the interviewee's preference, the interviewee identity may be anonymized in the interview transcript. The transcript of the interview will be shared with the interviewees in order to validate the interview and to gain feedback or comments. The interviewees are given a chance to change their statements before the interview result is finalized.

Coding process

The coding process involves analyzing interview transcript with the help of researcher's note. Open and axial coding will be used in order to gain insight from the interviewees' responses during the interview. Open coding provides the ability to give meaning to a section of the responses. Meanwhile, axial coding can reveal the relationships between the codes. In addition, the dimensions of transparency, privacy, and information quality will be used as codes in order to label which section of the interviewees' responses are related to the transparency, privacy, and information quality.

Software used in analysis

Data analysis will be done with the help of ATLAS.ti. ATLAS.ti is a software commonly used in qualitative research and qualitative data analysis. The software helps in analyzing qualitative data and provides assistance in coding. As this research is a qualitative research and gathers qualitative data, the use of ATLAS.ti can help in the process of coding and data analysis by providing tools for speeding up coding process and tools for visualizing the relationship between codes.

Reporting

The findings from the interviews will be reported in the master thesis document which the interviewer is writing as a requirement to graduate from the faculty of Technology, Policy, and Management at TU Delft. The final document containing the findings will be shared to the interviewees when it is ready to be published. In default, the interviewees will be asked to allow the publication of the interview findings using their real name and profile. However, if an interviewee does not give consent on publishing his or her identity, then any information that can personify him or her (e.g. name) will be kept confidential during the publication of the interview findings.

References

Ballou, D., & Pazer, H. (1985). Modeling data and process quality in multi-input, multi-output information systems. Management Science, 31(2), 150-162.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, 41(3), 16:1-16:52.

Bauhr, M., & Grimes, M. (2012). What is Government Transparency? The Quality of Government Institute: Working Paper Series.

Bharosa, N., & Janssen, M. (2015). Principle-Based Design: A Methodology and Principles for Capitalizing Design Experiences for Information Quality Assurance. Homeland Security & Emergency Management, 12(3), 469-496.

Borgesius, F. J., Eechoud, M. v., & Gray, J. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. Berkeley Technology Law Journal, 30(3), 2073-2131.

Delone, W., & McLean, E. (1992). Information systems success: the quest for the dependent variable. Information systems research, 3(1), 60-95.

Fedorowicz, J., Sawyer, S., Williams, C. B., Markus, M. L., Dias, M., Tyworth, M., . . . Schrier, R. (2014). Design observations for interagency collaboration. Government Information Quarterly, 31, 302-316.

Goodhue, D. (1995). Understanding user evaluations of information systems. Management Science, 41(12), 1827-1844.

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2014). Measuring Transparency. Political Analysis, 22(4), 414-434.

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. Journal of Organizational Computing and Electronic Commerce, 26, 14-40.

Jarke, M., & Vassiliou, Y. (1997). Data warehouse quality: a review of the DWQ project. Proceedings of the Conference on Information Quality, (pp. 299-313). Cambridge, MA.

Kim, S. (2009). Management Strategy for Local Governments to Strengthen Transparency in Local Governance. United Nations, United Nations Project Office on Governance. Seoul: UNPOG Publication.

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

Lyon, D. (2003). Surveillance as Social Sorting: Computer Codes and Mobile Bodies. In Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination (pp. 13-30).

Martínez-Ballesté, A., Pérez-Martínez, P. A., & Solanas, A. (2013). The Pursuit of Citizens' Privacy: A Privacy-Aware Smart City Is Possible. IEEE Communications Magazine, 51(6), 136-141.

Matheus, R., & Janssen, M. (2015). Transparency Dimensions of Big and Open Linked Data. 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society (pp. 236–246). Delft: Springer International Publishing.

Rowe, G., & Frewer, L. J. (2005). A Typology of Public Engagement Mechanisms. Science, Technology & Human Values, 30(2), 251–290.

Scholl, H. J. (2014). The EGOV Research Community: An Update on Where We Stand. 13th IFIP WG 8.5 International Conference, EGOV 2014 (pp. 1-16). Dublin, Ireland: Springer Berlin Heidelberg.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly, 20(2), 167-196.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems, 12(4), 5-33.

Zmud, R. (1978). Concepts, theories and techniques: an empirical investigation of the dimensionality of the concept of information. Decision Sciences, 9(2), 187-195.

Dimensions of transparency, privacy, and information quality

Presented below are the dimensions of transparency, privacy, and information quality which are relevant to open government data portals and their definition. These dimensions are obtained from existing literature regarding the three concepts and will be used as the focus of the interview in deriving design requirements of open government data portals.

Dimensions of transparency

The following table lists the dimensions of transparency relevant to open government data portals.

Table 1. Dimensions of transparency

Dimension of transparency	Definition
Publication of open government data (Bauhr & Grimes, 2012; Hollyer, Rosendorff, & Vreeland, 2014; Kim, 2009; Matheus & Janssen, 2015)	"The extent to which governments publish information electronically, as well as their clarification of reasons for restricting information when the wider public interest clearly demands it" (adapted from Bauhr & Grimes, 2012, p. 7, and Kim, 2009, p. 11)
Participation (Kim, 2009)	"The practice of involving members of the public in the agenda-setting, decision-making, and policy-forming activities of organizations/institutions for policy development" (Rowe & Frewer, 2005, p. 253)
Accountability (Matheus & Janssen, 2015)	"Answerability for the government's actions or inactions and the responsibility for their consequences" (Matheus & Janssen, 2015, p. 238)

Dimensions of privacy

The following table lists the privacy concerns that may arise due to the publication of open government data.

Table 2. Dimensions of privacy concerns due to publication of open government data

Dimension of privacy concern	Definition
Combining data (Smith, Milberg, & Burke, 1996)	"Concern that personal data in disparate databases may be combined into larger databases, thus creating a 'mosaic effect'" (Smith, et al., 1996, p. 172)
Lack of control over personal information (Borgesius, Eechoud, & Gray, 2015)	"The lack of control over one's personal information if that information is released as open data" (Borgesius, et al., 2015, p. 2089)
Social sorting and discrimination (Borgesius et al., 2015)	"Classification of people and populations according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access, and so on" (Lyon, 2003, p. 30)

The following are privacy concerns that may arise do to the use of open government data portals.

Table 3. Dimensions of privacy concerns due to the use of open government data portals

Dimension of privacy concern	Definition
Identity privacy (Martínez-Ballesté, Pérez-Martínez, & Solanas, 2013)	The possibility of disclosure of a data portal user's identity, thus allowing his/her activities in the data portal to be correlated to his/her identity (Martínez-Ballesté, et al., 2013)
Query privacy (Martínez-Ballesté, et al., 2013)	The possibility that the queries made by users to services are saved and thus allowing users to be profiled and information about their habits can be obtained (Martínez-Ballesté, et al., 2013)
Location privacy (Martínez-Ballesté, et al., 2013)	The possibility that a user's physical location can be disclosed (Martínez-Ballesté, et al., 2013)

Dimensions of information quality

The following table lists information quality dimensions.

Table 4. Dimensions of information quality

Dimension of information quality	Definition
Accuracy (Batini, Cappiello, Francalanci, & Maurino, 2009)	"The extent to which data are correct, reliable, and certified free of error" (Wang & Strong, 1996, p. 31)
Completeness (Batini, et al., 2009)	"The degree to which a given data collection includes data describing the corresponding set of real-world objects" (Batini, Cappiello, Francalanci, & Maurino, 2009, p. 16:7)
Consistency (Batini, et al., 2009)	The degree to which the semantic rules defined over a set of data items are not violated by the data (Batini, et al., 2009)
Timeliness (Batini, et al., 2009)	"The extent to which the age of data is appropriate for the task at hand" (Wang & Strong, 1996, p. 32)
Accessibility (Wang & Strong, 1996; Jarke & Vassiliou, 1997)	"The extent to which data are available or easily and quickly retrievable" (Wang & Strong, 1996, p. 32)
Ease of operations (Wang & Strong, 1996)	"The extent to which data are easily managed and manipulated (i.e. updated, moved, aggregated, reproduced, customized)" (Wang & Strong, 1996, p. 32)
Assistance (Goodhue, 1995)	"The extent to which help is provided in accessing and understanding the data" (Goodhue, 1995)
Ease of use (of h/w, of s/w) (Goodhue, 1995)	The extent to which the open government data portal is easy to learn and use (Goodhue, 1995)