Citizen Engagement with Open Government Data
A Model for Analyzing Factors Influencing Citizen Engagement

Purwanto, A.

DOI
10.4233/uuid:60bf122a-6ce7-48b4-b094-986d05c48428

Publication date
2021

Document Version
Final published version

Citation (APA)
Purwanto, A. (2021). *Citizen Engagement with Open Government Data: A Model for Analyzing Factors Influencing Citizen Engagement* . [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:60bf122a-6ce7-48b4-b094-986d05c48428

CITIZEN ENGAGEMENT WITH OPEN GOVERNMENT DATA

A MODEL FOR ANALYZING FACTORS INFLUENCING CITIZEN ENGAGEMENT

ARIE PURWANTO

Citizen Engagement with Open Government Data: A Model for Analyzing Factors Influencing Citizen Engagement

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Friday 10 December 2021 at 10:00 o'clock

by

Arie PURWANTO
Master of Science in Accounting, University of Gadjah Mada, Indonesia
born in Wonosobo, Indonesia

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.dr.ir. M.F.W.H.A. Janssen	Delft University of Technology, promotor
Dr. A.M.G. Zuiderwijk- van Eijk	Delft University of Technology, copromotor

Independent members:

Prof.dr.ir. N. Bharosa	Delft University of Technology
Prof.dr. F.E. Bannister	Trinity College Dublin
Prof.dr. J. Edelenbos	Erasmus University Rotterdam
Dr.ir. B. Enserink	Delft University of Technology
Dr. I.J. Mulder	Delft University of Technology
Prof.dr.ir. C.G. Chorus	Delft University of Technology, reserve member

Keywords: open data, open government data, citizen engagement, citizen-led initiative, hackathon, factors, drivers, inhibitors, motivation, intention

This research was funded by the Indonesia Endowment Fund for Education (LPDP).

ISBN: 978-94-6366-471-4
Printed by IPSKAMP Printing (https://proefschriften.net/)
An electronic version of this dissertation is available at https://repository.tudelft.nl/

Cover design by: Christiana Maya Kartikasari

Preface and Acknowledgements

Studying for a doctoral degree has been one of my dreams after graduating with my bachelor's degree. But I almost gave up when a lady representing the Delft University of Technology in the European Higher Education Fair told me that it would be impossible to be accepted at the university’s doctoral program because my master's degree was not in engineering. The turning point came slightly more than six years ago when I signed up to ResearchGate and requested Marijn to send me a paper about comparing open data benchmarks used to evaluate the progress of open data adoption (Susha, Zuiderwijk, Janssen, & Grönlund, 2015). I downloaded and read the paper, but I also became interested in open data and expressed my interest to do a Ph.D. under his supervision, something that would change the course of my life until today.

Doing the Ph.D. research, mainly writing it up as a dissertation, was never easy. I had to adjust to many challenges and difficulties; to some extent, I had to make a serious effort to come out of my comfort zone and learn more about myself. Nevertheless, these experiences have proved invaluable during my Ph.D. journey. This research would have never been completed without the support of many people who contributed to it. Therefore, I would like to thank them for this achievement.

My very first and most thanks are for Marijn Janssen, who had opened up an excellent opportunity for me to conduct this Ph.D. research at the Delft University of Technology. Marijn, you have always been academically and personally very supportive to me. Your critical, constructive feedback about the substance of my research comes timely. Discussing the study with you made me felt motivated and convinced that I can finish my Ph.D. Your practical, valuable tips on writing academic articles (e.g., deleting is writing) surprised me at first, but I always remember and apply them every time I write. You have also opened up many opportunities and avenues for my research direction. Beyond that, you had also shown considerable sympathy towards my personal and family situation and passed on helpful advice when vexing problems that delayed my dissertation progress.

I would also like to give a significant and warmest thanks to Anneke Zuiderwijk-van Eijk. With you and Marijn, I felt that I was really in good hands. Anneke, although you are ten years younger than me, your maturity in both academics and supervision really helped me move through a lot in each phase of my Ph.D. research. You were my supervisor and my mentor; until I finished this dissertation, I never met the mentor officially assigned to me, but I had you. During our discussions and collaboration in writing journal and conference

papers, you always have a different critical point and solutions to the research problems, from which I learned a lot.

I felt fortunate to join the Information and Communication Technology (ICT) section in the Engineering Systems and Services Department of Technology Policy and Management (TPM) Faculty. Yuxin Wang and Fernando Kleiman, thank you for being great office mates and, to some extent, friends. I enjoyed our chats and informal discussions about politics and culture very much. Particularly, Fernando, thank you very much for your valuable feedback and contribution in pretesting the quantitative survey instruments. Yi Yin, Dhata Praditya, Ahmad Luthfi, and other colleagues helped to test the same instruments and substantially improve the instruments. Jolien Ubacht, thank you very much for our discussions and your introduction to the qualitative approach using specialized software and its operation for my research. Mark de Reuver, a big thanks to you for spending your time to discuss the quantitative approach taken in my study; those discussions helped me make an important decision during the research phase. Ricardo Matheus and Seyed Latifi Rostami, thank you for sharing our laughs during so many occasions that we participated together, both formal and informal. Many other ICT colleagues and TPM's master students also contributed to my research. Iryna Susha, Boriana Rukanova, Sélinde van Engelenburg, Klara Pigmans, Clara Matthijs, Agung Wahyudi, Rina Elvira, and Lina Afriana to name a few. I am also very thankful to the helpful and responsive ICT's secretaries Laura Bruns and Jo-Ann Karna for their support during the study.

I would also like to express my gratitude to Hack de Valse Start participants from Team Pac Man: Johannes Visser, Cristina-Lavinia Colchior, Roman Lyapin, and Femmelien; and Team Next STEP: Thomas Gomez and Steve Thijssen. Although we interacted in such a short time, I really enjoyed working together, particularly with Team Pac-Man. A big thanks to Alexander Slaghuis from Open State Foundation, who supported our team in the hackathon. I would also like to extend thanks to Arnold Jonk, the inspector of Onderwijs Inspectie, and Gabriëlle Remerij, the Information Manager of Onderwijs Inspectie. Thank you all for your willingness and valuable time to participate in the interviews for my case study. Your information has substantially contributed to my research, helped me progress, published conference articles, and ultimately wrote this dissertation. Other big thanks to the regional Open Knowledge Foundation mailing list members from 29 countries who have helped distribute my survey and participated in it. I would also like to thank Linda Haartsen, Josien Kapma, and Anne Bruinsma from FarmHack.nl. You have provided me with anonymized data sets of farm hackathons, which were invaluable and helped me gain more insights into my Ph.D. research.

I would also like to thank the Indonesian communities and friends in the Netherlands and Indonesia who had helped and supported my research intellectually, practically, or spiritually. Kawal Pemilu members such as Dian Lestari, Ainun Najib, Yoice Stefano, to name a few, thank you very much for recommending me to most of the interviewee candidates for the qualitative research and participating in the interviews. Elina Ciptadi, another Kawal Pemilu member, thank you very much for helping me distribute my survey and participating in it. I would also like to thank I Dewa Kade Wiarsa Raka Sandi, Sidharta Wisnu Wardhana, and Eka Wisnu Wardhana, the commissioners of Komisi Pemilihan Umum. You have participated in our interviews and provided valuable and vital information. The big family of the Indonesian Church in the Netherlands (GKIN), thank you very much for your spiritual support for my family during the Ph.D. research. Notably, the Hasibuan family (Mas Davied and Mbak Hesti) was always there for us in our ups and downs for almost four years living in Delft. Many other families had also supported us. Just to mention a few, Tjahjadi (Pak Stanley and Bu Santi), Susanto (Pakdhe Eddy and Budhe Inne), Rumondor (Oom Eric and Tante Yvone), Lumingkewas (Mas Lucky and Mbak Lucy), Situmorang (Bang Baput and Mbak Yanti), Parlindungan (Bang Salmon and Kak Sonya), Wahjuwibowo (Mas Octo and Mbak Lisa), van Duijn (Peter and Gita), Vergeer (Mbak Ellen and Arie), Silitonga (Nelson and Pinta), Steller (Wilbert and Sarah), Kusumadjaja (Anton and Monica), Madijokromo (Milton and Manik), Melanie, and Vanessa Sidharta. I also greatly thank the Yang Lurus group, Mbak Evabrantie Wartono, Mbak Mira Sari, and Mas Bambang Ponco. You all have been there for my family in our ups and, mainly, our downs. Our relationships formed not only solid friendships but also have become a new family for me.

Finally, I owe the most gratitude to my family. Ibuke (mom) and Bapake (dad), thank you for your encouragement and patience during my Ph.D. research. Special thanks go to my dearly loved wife, Maya, and my two lovely children, Khya and Dayu. Moving to the Netherlands has disrupted our lives and forced us to get out of our comfort zones. Yet, you all have been willing to face the challenges of living in a stranger’s land and very supportive during the Ph.D. research and dissertation writing. I could not have finished the study and completed the dissertation without you, and therefore, in the end, I am greatly indebted to you.

Delft, August 17, 2021.

Table of Contents

1. Introduction

Nowadays, citizens have access to more and more public data provided by governmental organizations at different administrative levels worldwide (Janssen, 2011; McDermott, 2010; Ubaldi, 2013). Researchers refer to this type of data as Open Government Data (OGD). We can define OGD as *non-privacy-restricted, non-confidential data made publicly available on the internet by governmental organizations and can freely be used, reused, and distributed by anyone without any restrictions* (adapted from Ubaldi, 2013, p. 6). OGD encompass many domains such as business, geography, legislation, climate and weather, economics, employment, health, population, public administration, and transportation (Ubaldi, 2013). Governmental organizations typically publish OGD on a portal that offers basic functionalities such as finding, browsing, downloading data sets, and more advanced features for comparing and visualizing data.

Society can benefit from OGD when used to create artifacts, such as facts, applications, and visualizations that contribute to solving societal issues (Kuk & Davies, 2011; Susha, Grönlund, & Janssen, 2015). For example, in Singapore, a community of citizen-developers built mobile applications on top of the dengue cases, and dengue clusters data sets publicly available on the Internet via Application Programming Interface (API) (Young, Sangokoya, & Verhulst, 2016). The members of society in the affected areas can use these applications to protect themselves from mosquito bites and remove mosquito breeding areas. Another example concerns a crowdsourcing application built by Indonesian citizens on top of the open election data (Graft, Verhulst, & Young, 2016). The public can use the application to identify and report suspicious election results and prevent corruptive acts from taking place.

Various skills and capabilities are needed to create such an artifact utilizing OGD (Janssen, Charalabidis, & Zuiderwijk, 2012). Creating different OGD artifacts will likely lead to various complexity levels of tasks and activities (Susha, Grönlund, et al., 2015). For example, developing an OGD-based application would require advanced computer programming and cybersecurity capabilities. Performing activities to create OGD-based artifacts requires resources, such as time, money, and skills. Furthermore, collaboration is sometimes inescapable when creating complex objects such as web applications. Despite having the required expertise, average citizens might not be willing to develop artifacts utilizing OGD (Lourenço, 2015). As a result, OGD programs' success is contingent on, among other factors, active and *engaged* citizens (Dietrich, 2015).

Citizen engagement with OGD has various potential benefits. It can improve the citizen-government relationship, facilitate policy implementation, and generate ideas, information, and service innovation (Nam, 2012). As a result, governmental organizations aim to increase citizens' opportunities to engage in policymaking through OGD (Obama, 2009). Engaging citizens, in turn, will benefit the government with their collective expertise and information. However, previous research shows that governmental organizations providing OGD have very limited or no knowledge about who engages with their data (Susha, Grönlund, et al., 2015) and why they do so (Johnson & Robinson, 2014). The lack of insights into the actual users and their motivation may lead to poor judgment of the benefits of the OGD programs.

The current literature does not contain many insights into factors that explain OGD citizen engagement; the open data research needs to develop a comprehensive behavioral adoption model (Hossain, Dwivedi, & Rana, 2016). Having insights into the factors help identify essential and relevant antecedents of OGD engagement and subsequently enables data providers to focus on addressing specific areas that can improve the engage-ability of their OGD programs in practice. The model can also help OGD providers design new programs that take citizens' perspectives into account. Such a model should consider the profiles of citizens who would act as potential users of OGD, the characteristics of the OGD provision, conceivable outcomes of OGD engagement, and factors that influence citizen engagement with OGD.

This research aims *to develop a model that explains what factors contribute to citizen engagement with OGD*. In an OGD ecosystem, three types of actors exist *OGD providers*, *OGD users*, and *OGD end-users* or *beneficiaries* (Dawes, Vidiasova, & Parkhimovich, 2016). Section 1.1 provides a more detailed explanation of the actors in the OGD ecosystem. This research focuses on a particular type of OGD user: citizens. More specifically, since engaging with OGD requires the possession of relevant technological skills, the citizens this study focuses on are assumed to be digitally literate or technologically skilled (see Figure 1.1).

Moreover, this research focuses on citizens who are not government officials; although civil servants are also citizens, they are outside the scope of this research. Also, outside the scope of this research are OGD providers such as governmental organizations, OGD users from the public and private sectors, and OGD beneficiaries (end-users). The members of the public sector that generally use OGD may include civil society organizations or non-governmental organizations, while the private sector includes companies.

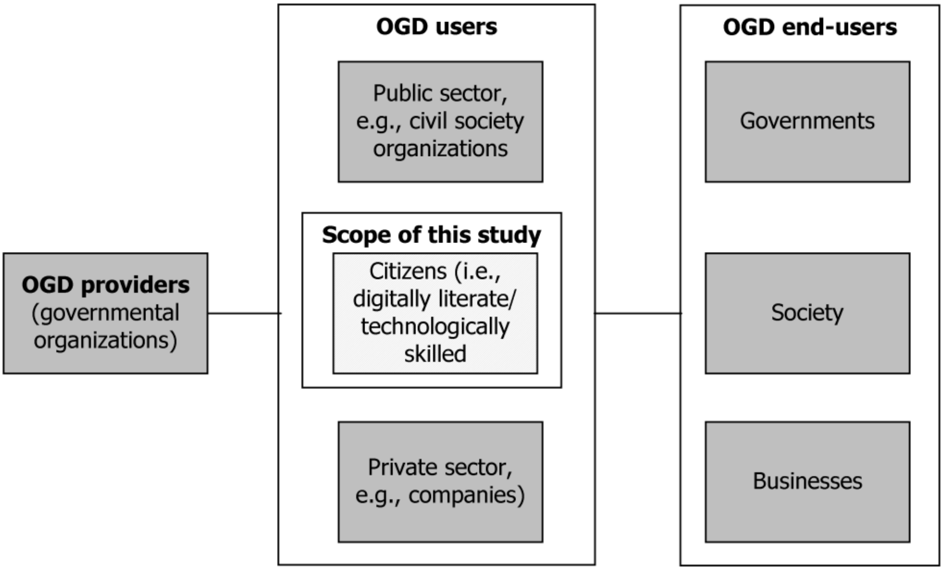


Figure 1.1. The focus of this research.

The structure of this chapter is as follows. First, Section 1.1 illustrates and describes the focus of this research within an OGD ecosystem context, followed by Section 1.2, which explains how citizens can act as users of OGD in the ecosystem. The descriptions of the knowledge gaps in the current open data literature and the problem statement formulation are presented in Section 1.3. Next, Section 1.4 discusses the scientific and societal contributions of this study. Subsequently, the objective of this study and the research questions are elaborated in Section 1.5. Finally, Section 1.6 provides an outline of this dissertation.

1.1. OGD ecosystem

Different types of independent actors can use OGD provided by governmental organizations using OGD technologies and create something out of them, and the outcomes of the usage can benefit various types of actors (Safarov, Meijer, & Grimmelikhuijsen, 2017). Generally, these actors encompass two broad groups of users who are revenue-driven (e.g., developers and companies) and oriented toward public value creation (e.g., journalists, researchers, and citizens) (Lassinantti, Bergvall-Kåreborn, & Ståhlbröst, 2014). Knowing these actors is essential because various stakeholders hold different interpretations of and interests in OGD (Gonzalez-Zapata & Heeks, 2015).

The *ecosystem* term, borrowed from biology, generally refers to the illustration of complex causal interaction between groups of actors who depend on each other's activities (Dawes et al., 2016; Jacobides, Cennamo, & Gawer, 2018). Therefore, the ecosystem metaphor is particularly relevant to describing and emphasizing this research's scope: who are the citizens engaging with OGD and how they interact with other human and non-human actors to create OGD-based artifacts. An OGD ecosystem can explain the interactions of socio-technical systems between different groups of actors at multiple interdependent levels (Dawes et al., 2016; Harrison, Pardo, & Cook, 2012; Zuiderwijk, Janssen, & Davis, 2014). Figure 1.2 illustrates such an ecosystem under study and underscores the focus of this research, concentrating on the interrelationships between human actors, i.e., citizens and non-human actors such as government data, portals, tools, services, and OGD-based artifacts.

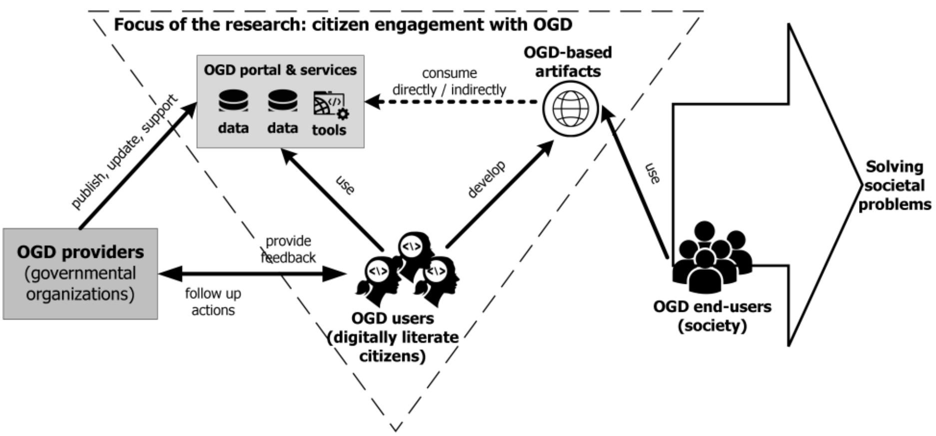


Figure 1.2. An OGD ecosystem model synthesized from Harrison et al. (2012), Zuiderwijk, Janssen, et al. (2014), and Dawes et al. (2016).

During the daily operation of governmental organizations' business processes, a large volume of data are captured, stored, processed, and created (Alexopoulos, Zuiderwijk, Charalabidis, Loukis, & Janssen, 2014). Governmental organizations typically publish OGD after undergoing anonymization and declassification procedures. The procedures ensure the exclusion of privacy-related and confidential information. Government organizations that make the data publicly available on the internet for free use, reuse, and distribution by anyone can be referred to as *OGD providers*.

Generally, OGD providers publish open data, including metadata, on a website portal and offer users *tools* equipped with necessary features such as interactive data browsers or even advanced functionalities, including data visualization and comparison (Zuiderwijk, Janssen, et al., 2014). In addition, although it is uncommon, some OGD providers also offer particular *support* or *services* to users for assisting them in accessing and using the opened data sets (Rudmark, Arnestrand, & Avital, 2012).

Utilizing government data and services provided in the portal, *OGD users* (i.e., citizens) can develop *OGD-based artifacts* that OGD end-users can use to solve societal problems. OGD-based artifacts refer to the outputs of the processes carried out by OGD users when using the data; different uses of OGD can generate multiple outcomes (Davies, 2010). These artifacts can be facts, information, other data, interface, or service (Susha, Grönlund, et al., 2015). For example, the X-Dengue application is an OGD-based artifact built by a community of citizen-developers based on the dengue cases and dengue clusters data sets opened by the Singaporean National Environment Agency (Young et al., 2016). The application represents multiple outputs of OGD use. For instance, it provides information on whether a particular location is at risk, offers an interface to search for the occurrence of the disease, and provides service for the citizens to report danger zones. Such artifacts contribute to solving societal problems such as taking preventive measures to protect themselves from mosquito bites and removing mosquito breeding areas. The use of OGD by citizens to create artifacts is the focus of this research.

In the ecosystem, *OGD end-users*, or *OGD beneficiaries*, are society members who use OGD-based artifacts that contribute to societal problems such as preventing corruption and at the same time receive the benefits from public values created or added by the solutions (Gonzalez-Zapata & Heeks, 2015). For example, widely shared information about irregular items found in open budgeting data on social media platforms would help the government revoke the budget and prevent corruption of civil servants. However, OGD end-users and how they use and benefit from OGD-based objects are excluded in this research.

1.2. Engaging citizens as OGD users

Citizens are typically viewed merely as the beneficiaries of OGD who receive the benefits or value created by the use of OGD (Safarov et al., 2017). Most open data scholars support this view (Harrison et al., 2012; Parycek, Höchtl, & Ginner, 2014). However, empirical evidence shows that citizens can also be the direct users of OGD or infomediaries who create new solutions to solve societal issues. For example, citizens comprised three-fifths of open transportation data hackathon participants in Sweden (Juell-Skielse, Hjalmarsson, Johannesson, & Rudmark, 2014). Another example from the Netherlands shows that citizens constituted 48% of the unique users of the Netherlands Land Registry and Mapping Agency's open geographical data portal (van Loenen, Ubacht, Labots, & Zuiderwijk, 2017).

Citizen engagement with OGD can manifest in different initiatives, whether they are led and organized by citizens (citizen-led) or by governmental organizations as OGD providers (government-led). For example, hackathons or innovation contests, typically sponsored by governmental organizations, are examples of joint efforts to engage citizens as OGD users in a collaborative setting. This government-led type of engagement has been relatively well-studied (e.g., Briscoe & Mulligan, 2014; Concilio, Molinari, & Morelli, 2017; Gama, 2017; Hartmann, Mainka, & Stock, 2016; Juell-Skielse et al., 2014). Nevertheless, a different type of OGD engagement initiative wholly led and organized by citizens also exists (e.g., Graft et al., 2016). However, the open data literature barely studies the latter type of engagement.

1.3. Problem statement

This study is not among the first to address the citizen-led and government-led citizen engagement with OGD. Previous research has investigated the topic in different contexts. For example, Brajawidagda and Chatfield (2014) studied social media's roles in citizen-led engagement with open election data initiatives. Another example is Hjalmarsson, Johannesson, Juell-Skielse, and Rudmark's (2014) work, which explored innovation barriers in an open data innovation contest sponsored by a governmental transportation agency. The following knowledge gaps concerning citizen led-engagement with OGD and formulated problem statements related to the open data literature gaps are identified.

First, citizens are barely seen as direct users of OGD and typically viewed solely as indirect OGD users, namely as the beneficiary of values created or added by the result of the OGD use (e.g., Harrison et al., 2012; Parycek et al., 2014). However, in the OGD ecosystem, citizens as direct OGD users can act like an intermediary who bridges the society and governmental organizations (e.g., Juell-Skielse et al., 2014; van Loenen et al., 2017). Nevertheless, open data scholars rarely take this role into account when addressing engagement with OGD. Most scholars study OGD intermediaries from the perspective of companies or civil society organizations and their business models (e.g., da Silva Craveiro & Albano, 2017; Janssen & Zuiderwijk, 2014; Kassen, 2018; Mercado-Lara & Ramon Gil-Garcia, 2014). Therefore, empirical research is needed to enrich insights into citizens' roles in real-life OGD engagement cases.

Second, the current literature's widely known form of citizen engagement with OGD is open data hackathon or innovation contest (e.g., Briscoe & Mulligan, 2014; Concilio et al., 2017; Gama, 2017; Hartmann et al., 2016; Juell-Skielse et al., 2014). This type of engagement is typically initiated, led, sponsored, and organized by governmental organizations that provide OGD. However, entirely independently, citizen-led engagement also exists in practice (e.g., Graft et al., 2016). The open data literature barely investigates citizen-led engagement initiatives. More insights from concrete citizen-led initiatives are needed to understand OGD engagement better.

Third, no comprehensive overview of factors influencing citizens' intention to engage with OGD is available in the current open data literature (Hossain et al., 2016). A handful of studies have investigated the antecedents of intention to use open government or open data technologies (e.g., Jurisch, Kautz, Wolf, & Krcmar, 2015; Zuiderwijk, Janssen, & Dwivedi, 2015). However, they examined different antecedents and offered different mixed conclusions. The current open data research rarely extends or consistently builds upon these already studied antecedents, and thus, insights into the factors that influence citizen engagement with OGD are fragmented. A model that integrates relevant factors is needed to explain citizens' intentions as OGD users to engage with OGD.

Fourth, previous open data research used and applied various high-level theories, such as the Unified Theory of Acceptance and Use of Technology (UTAUT), Technology Acceptance Model (TAM), or Information System (IS) Success Model. Although scholars have widely used these theories in the IS research domain, they are only partly applicable in the context of OGD. Moreover, the development of open data theory is still in its infancy (Hossain et al., 2016). Thus, there is a need to develop more specific theories or theoretical models adapted to particular open data contexts.

1.4. Scientific and societal contributions

The scientific contributions of this research are as follows. First, while previous research mainly focused on the roles of companies or civil society organizations as OGD users, this research provides insights into citizens' roles as direct users of OGD. Thus, this research contributes to studying the prospective roles of citizens as direct users of OGD and, subsequently, OGD intermediaries. Second, top-down initiatives such as open data hackathons predominate the understanding of OGD engagement in current literature. Thus, this research sheds light on citizen-led OGD engagement and complements the recent insights predominated by government-led engagement. Third, since no comprehensive overview of factors that influence citizens' intention to engage with OGD exists, this research contributes to the literature by providing a unified theoretical model of OGD citizen engagement. Fourth, while previous research adopts particular theories and theoretical models from the IS domain that are only partially applicable to the OGD contexts, this research develops and evaluates a theoretical model grounded in and synthesized from the OGD literature. Finally, this research contributes to the literature by providing an empirical assessment of factors currently assumed to influence citizens' intention to engage with OGD.

This research also offers practical contributions to practitioners, i.e., civil servants in charge of OGD provision and open data policymakers and governmental organizations responsible for auditing open data programs. The proposed OGD citizen engagement model can guide open data managers to improve ongoing open data programs or design new programs. The managers can use such a model to evaluate whether the programs have considered the profiles of citizens using OGD and factors that positively impact this use. Policymakers can utilize the model for developing segmented open data policies that deal with specific user characteristics and factors to increase public engagement with OGD. Auditors of national audit institutions can apply the model for developing audit criteria of OGD programs' performance. The program evaluation results can provide recommendations to improve the engage-ability of an OGD program and ultimately lead to public value creation.

1.5. Research objective and research questions

Citizen engagement and the subsequent expected value creation are among the main reasons governments worldwide open their data. However, it is not clear which factors influence citizen-led engagement with OGD in comparison to government-led OGD. Therefore, taking into account the identified gaps within the open data literature (see Section 1.4), the objective of this research is:

***Research objective:* to develop a model for understanding factors contributing to citizen engagement with OGD**

Three research questions (RQs) are formulated to achieve the research objective. This research is exploratory, aiming to understand the citizen engagement with OGD and subsequently develop a model explaining the factors influencing the engagement. Therefore, the investigation begins with exploring the phenomenon under study to understand OGD citizen engagement better, and in the end, evaluate the examined factors in larger samples of digitally literate citizens. This research is divided into three phases: 1) understanding the phenomenon, 2) exploration and identification of factors, and 3) factor analysis and model validation.

1.5.1. RQ1: What drivers and inhibitors for citizen engagement with OGD have been identified in previous research?

The first phase of this research is understanding the phenomenon under study, citizen engagement with OGD, and factors that drive and inhibit it, to obtain better insights from the current knowledge for three purposes. First, the phase enables researchers to position their study corresponding to the current knowledge, propose areas for further study, and expand on it. The second purpose of this phase is to synthesize existing evidence concerning factors that drive individual citizens to engage with OGD or inhibit them from engaging with OGD. Lastly, the primary purpose of this research phase is to develop a theoretical framework that explains different driving and inhibiting factors that influence and moderate citizen engagement with OGD.

Citizen engagement with OGD is expected to be influenced (driven or inhibited) and moderated by different factors. Such influencing factors include personal factors such as citizens' motivations, technical factors concerning the quality of OGD, social factors such as influence from social relationships, and political factors. Factors such as citizens' profiles (e.g., capabilities) may also play a role in moderating the relationships between drivers and inhibitors and citizen engagement with OGD. A systematic literature review is conducted to answer the first research question. A theoretical framework for analyzing the driving and inhibiting factors that influence citizen engagement with OGD is developed as the outcome of the research phase to assist the proceeding phase.

1.5.2. RQ2: Why do citizens engage with OGD in existing government-led and citizen-led OGD initiatives?

Using the theoretical framework developed from the first phase (RQ1), factors influencing citizen engagement with OGD in real-life cases in the second research phase (RQ2) are explored and identified. The answer to the second research question helps better understand whether the framework has included all the necessary factors influencing citizen engagement. Multiple case study research is used to investigate whether factors proposed in the theoretical framework exist in practice and whether factors are missing from the framework. The primary purpose of this research phase is to apply the proposed theoretical framework in real-life settings and evaluate factors that emerge in citizen engagement cases.

Citizen-led and government-led are two opposing types of OGD engagement initiatives known in the current literature; this research phase focuses on real-life cases replicating these engagement types. It is expected that different factors influence citizens in different kinds of OGD engagement. The diverse backgrounds of citizens are also assumed to moderate different types of engagement. The cases selected concern one government-led engagement initiative in open education hackathons and one citizen-led engagement initiative with open election data. The outcomes of this research phase are propositions that lay the ground for developing the hypotheses that are assessed in the subsequent research stage.

1.5.3. RQ3: What model explains citizens' intention to engage with OGD?

This third research phase evaluates the hypotheses (and research model) formulated from the propositions generated in the previous research phase (RQ2). The research model is developed based on the hypotheses that predict causal relationships between factors and citizens' intention to engage with OGD. The final phase of this study involves quantitative assessments of the research model using more extensive samples of digitally literate citizens engaged with OGD. A questionnaire is developed to measure citizens' intention to engage with OGD. The questionnaire is created in the English and Indonesian languages and distributed to various OGD user groups. The evaluation of hypotheses and research models is central to the third research question and performed using a partial least square-structural equation modeling approach (PLS-SEM). The approach examines the measurement model, i.e., the relationships between variables and the underlying factors, and the structure model, i.e., the relationships among factors. The outcomes of this research question resemble the final models of factors influencing citizen engagement with OGD.

1.6. Dissertation outline

This dissertation is organized as follows (see Figure 1.3). First, Chapter 2 describes a more detailed account of the approaches used in this research. Chapter 3 exhibits the theoretical and empirical perspectives on citizen engagement factors based on the current open data knowledge. Next, Chapter 4 describes the case studies to identify and explore factors influencing citizen engagement with OGD. This chapter also reports the case study analysis. Subsequently, Chapter 5 explains the development of the questionnaire and the quantitative data collection and analysis to explore the structures of factors comprising the OGD Citizen Engagement (OGDCE) model. This chapter also validates the OGDCE factor structures and evaluates their measurement and structure models based on a CFA approach. Finally, Chapter 6 discusses the overall research conclusion, research limitations, and future research.

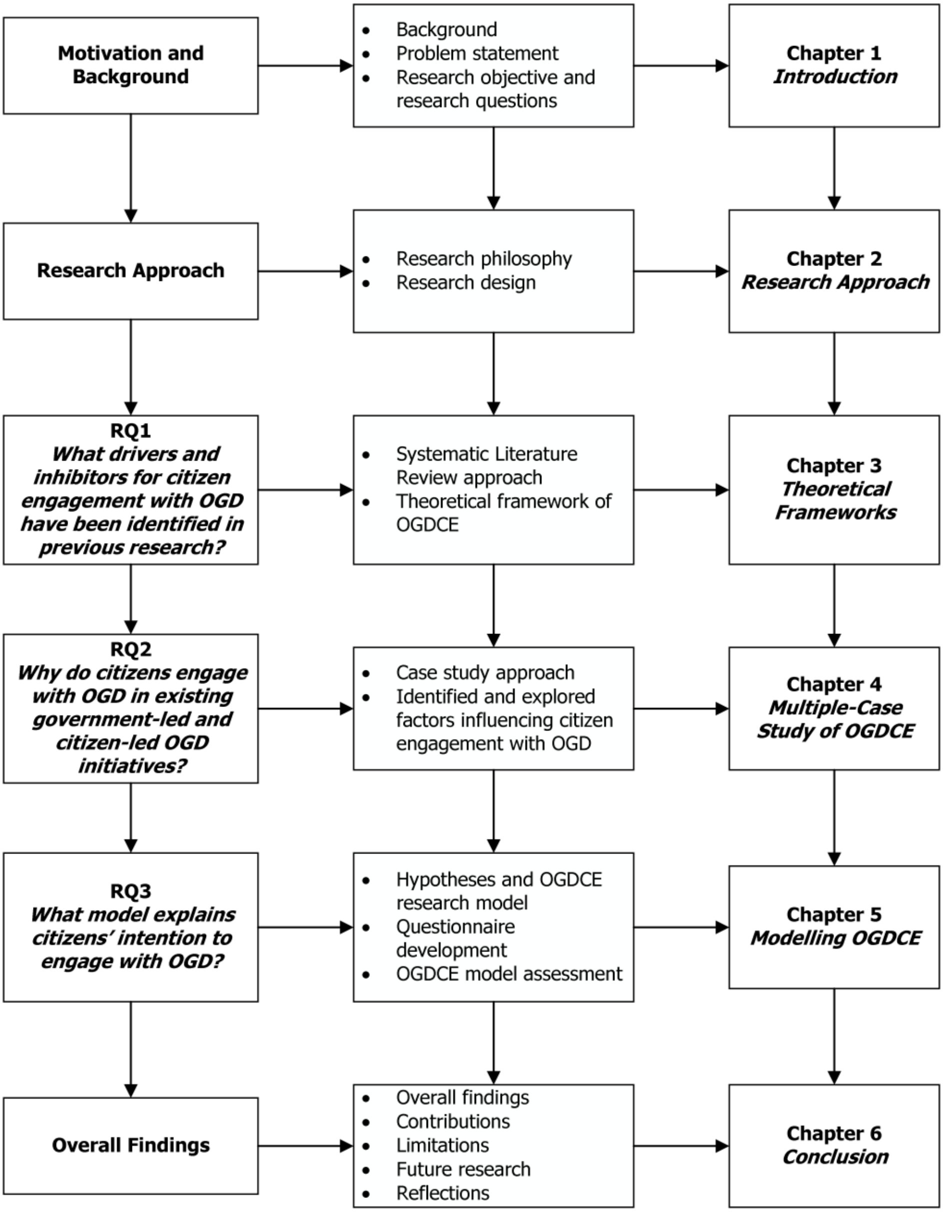


Figure 1.3. The outline of the dissertation.

2. Research Approach

The previous chapter presents the background and motivation of this research, the knowledge gaps this study aims to fill, the research objectives, and the research questions formulated for attaining the objectives. This chapter describes the design and strategies for conducting the research. The research design largely depends on the researcher's view on reality, how it can be studied, and how knowledge about reality can be constructed and validated. Therefore, this chapter presents beliefs about reality and knowledge creation and this research's stance, *pragmatism*. The chapter then discusses the mixed methods approach used in this research as a natural consequence of embracing a pragmatist view. Next, it describes the mixed methods research design used to combine qualitative and quantitative approaches. Finally, it explains the phases of this research and the methods used in each stage.

2.1. Research philosophy

Knowledge is a social artifact about people, their physical and social environment, and their relationship (Chua, 1986). In knowledge production, rules or beliefs that guide researchers delimit them in performing their works (Burrell & Morgan, 1979; Chua, 1986; Orlikowski & Baroudi, 1991). These beliefs constitute the philosophical viewpoints or assumptions that researchers adopt when researching and perceiving the world (Orlikowski & Baroudi, 1991). Burrell and Morgan (1979) conceptualize these assumptions based on four aspects of reality and how it may be studied: ontology, epistemology, human nature, and methodology.

First, ontological assumptions concern the essence of phenomena under study, which is how physical and social reality is considered. Burrell and Morgan (1979) recognize two opposing ontological views in social science: objective and subjective. A researcher may assume that reality is objective and independent and externally exists without human intervention. In contrast, subjectivism believes that reality exists only through human action.

Second, epistemology is associated with ontological issues. Whichever a researcher adopts the ontological view, it may lead to different epistemological beliefs. Epistemological assumptions concern the criteria for constructing knowledge and evaluating the validity of knowledge about a phenomenon. As an example, Chua (1986) writes that "an epistemological assumption might state that a theory is considered true if it is repeatedly not falsified by empirical events" (p. 604). On the other hand, anti-positivism is in the extreme position opposite positivism, assuming that knowledge is constructed through unique personal experience and insight (Burrell & Morgan, 1979).

Third, the assumptions on human nature concern the relationship between human and their environment. They are also related to the ontological and epistemological issues, though conceptually separated from them (Burrell & Morgan, 1979). Two instances of extremely opposing views on human nature are determinism and voluntarism (Burrell & Morgan, 1979). Burrell and Morgan (1979) identify the propensity of the deterministic view to believe that "human beings and their experience are regarded as products of the environment" (p. 2), and "humans are conditioned by their external circumstances" (p. 2). Voluntarism contradicts this view and believes that humans are the creators of their environment. Furthermore, voluntarism assumes that free will is the foundation of humans' actions in responding to the situations encountered.

Fourth, methodology indicates selecting appropriate research methods and techniques for collecting valid empirical evidence (Chua, 1986; Orlikowski & Baroudi, 1991). The stances of ontology, epistemology, and human nature views subscribed by a researcher will implicate specific sets of available methodologies. "Different ontologies, epistemologies and models of human nature are likely to incline social scientists towards different methodologies" (Burrell & Morgan, 1979, p. 2). Bryman (2012) suggests that quantitative and qualitative methodologies are instances of the extreme positions of positivist and anti-positivist research, respectively.

2.2. Philosophical paradigms

Research in social science has been primarily conducted under two dominant, opposing philosophical paradigms: positivism and interpretivism (Orlikowski & Baroudi, 1991; Wynn & Williams, 2012). However, alternative philosophical streams have emerged as agnostic responses to both paradigms, including pragmatism (Van de Ven, 2007; Venkatesh, Brown, & Bala, 2013). Table 2.1 provides an overview of the stances of the assumptions towards ontology, epistemology, human nature, and research methods, according to Orlikowski and Baroudi (1991), Burrell and Morgan (1979), Guba and Lincoln (1994), and Van de Ven (2007). The following sections explain the paradigms of positivism, interpretivism, and pragmatism.

Table 2.1. Overview of research paradigms adapted from Orlikowski and Baroudi (1991), Burrell and Morgan (1979), Guba and Lincoln (1994), and Van de Ven (2007).

Beliefs about	**Positivism**	**Interpretivism**	**Pragmatism**
Ontology	Objective reality exists independent of human	Reality is constructed in the human mind (or socially constructed)	Reality can only be imperfectly understood

Beliefs about	**Positivism**	**Interpretivism**	**Pragmatism**
Epistemology	Falsification of hypotheses and theories	Understanding of social practices and meanings constituted and influenced by the language and tacit rules	Non-falsification of hypotheses and theories (triangulation across multiple fallible perspectives)
Human nature	The environment ultimately shapes human actions	Humans are entirely autonomous	Dialectical interactions between humans and the environment
Methodology	Often quantitative methods	Often qualitative methods	Pluralistic or mixed methods (qualitative and quantitative)

2.2.1. Positivism

Positivism refers to the belief that social science research should model research in the natural sciences (Lee, 1999; Lee & Baskerville, 2003). Orlikowski and Baroudi (1991) identify that "ontologically, positivist information systems researchers assume an objective physical and social world that exists independent of humans, and whose nature can be relatively unproblematically apprehended, characterized, and measured" (p. 9). They consider that social reality is as objective as a physical reality and that researchers and reality are separate (Weber, 2004). Hence, they believe that objective research is value-free or unbiased (Guba & Lincoln, 1994). Epistemologically, positivists rely upon the empirical testability of hypotheses and theories to conclude whether they are "true" or "false" (Chua, 1986). Researchers cannot claim a theory as knowledge if they cannot expound it unambiguously and confirm it by scientific inquiry (Chua, 1986). On the models of human nature, positivist researchers believe that the environment entirely determines human beings and their actions (Burrell & Morgan, 1979). Based on these beliefs, a positivist is likely to use a nomothetic approach, law-like generalizations independent of time or context, to social science (Orlikowski & Baroudi, 1991). The nomothetic method heavily depends on quantitative techniques for data analysis (Lee & Baskerville, 2003). For example, researchers often regard sampling-based statistical analysis as a positivist method (Lee & Baskerville, 2003). A revised form of positivism, *post-positivism* (Guba & Lincoln, 1994; Teddlie & Tashakkori, 2009) or *neo-positivism* (Alvesson, 2003), acknowledges that the positivist researchers' value systems play an essential role in the research conduct and interpretation of data. Nevertheless, neo-positivism (or post-positivism) emphasizes quantitative methods (Teddlie & Tashakkori, 2009).

2.2.2. Interpretivism

Interpretivism refers to the belief that knowledge, in social science research, is obtained merely through social constructions (Klein & Myers, 1999; Mingers, 2004), which cannot be studied with the natural sciences (Lee & Baskerville, 2003). Ontologically, interpretive IS researchers believe that the existence of social reality is not “given” but rather a product of humans’ actions and interactions (Orlikowski & Baroudi, 1991). Therefore, social reality can only be viewed subjectively by interpreting human experience (Klein & Myers, 1999). Multiple realities exist and cannot be discerned, characterized, and measured objectively or universally. Epistemologically, interpretivism assumes that deducting hypotheses and theories cannot construct knowledge about social reality (Orlikowski & Baroudi, 1991). Instead, researchers capture knowledge through involvement in a social process that narrates how everyday social practices and meanings are constituted and influenced by humans' language and unspoken rules working toward common goals (Orlikowski & Baroudi, 1991). An interpretive researcher believes that humans are wholly autonomous and free-willed (Burrell & Morgan, 1979). Methodologically, directed by the stances above, an interpretivist tends to use an idiographic approach to social science (Orlikowski & Baroudi, 1991). Burrell and Morgan (1979) pinpoint the emphasis of the idiographic on “the detailed analysis of the insights generated by such encounters with one’s subject and the insights revealed in impressionistic accounts found in diaries, biographies and journalistic records” (p. 6). The idiographic approach, hence interpretive research, is typically contingent on qualitative techniques (Burrell & Morgan, 1979). Interpretive research usually involves actual case studies involving real people in real situations and is conducted in real-world settings (Hirschheim & Klein, 2003).

2.2.3. Pragmatism

Pragmatism refers to the belief that natural sciences and social sciences (and other types of sciences) are not fundamentally or categorically distinct. Each is a narrative of how researchers obtain knowledge (Wicks & Freeman, 1998). Therefore, pragmatist researchers are not committed to any one type of science (Creswell & Creswell, 2018). Pragmatists believe that an objective reality exists externally to humans (Goles & Hirschheim, 2000). However, this reality is grounded in each individual’s environment and experience and cannot be perfectly discerned (Goles & Hirschheim, 2000; Guba & Lincoln, 1994). Guba and Lincoln (1994) argue that human intellectual mechanisms are flawed, and the nature of phenomena is fundamentally intractable (Guba & Lincoln, 1994). Hence, pragmatism views knowledge as “being both constructed and based on the reality of the world one experiences and lives in” (Teddlie & Tashakkori, 2009, p. 70). Epistemologically, pragmatist IS researchers assume the naturalistic and process-oriented organism-environment transaction, and therefore, endorse fallibilism (Johnson & Onwuegbuzie, 2004). Since a pragmatist researcher cannot be entirely objective and independent when trying to understand reality, bias is an inherent characteristic of the researcher (Guba & Lincoln, 1994). Therefore, methodological pluralism, typically involving different data collection and analysis forms, is endorsed in pragmatist research, based on what is useful and what works (Creswell & Creswell, 2018; Teddlie & Tashakkori, 2009; Wicks & Freeman, 1998).

The philosophical stance of this research is pragmatism based on the following arguments. First, the research domain is multidisciplinary and positioned at the intersection of Public Administration (PA) and IS disciplines. PA is part of the political studies domain. In this domain, researchers explore the political nature of PA management and the public policymaking process (Osborne, 2006). Indeed, opening government data online invites citizens to participate in public policy discourses. On the other hand, IS is an applied research discipline that frequently applies theories from different fields such as computer science and the social sciences (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

Second, the focus of this study is citizens, who are human beings with different purposes and experiences in OGD engagement. They have to dialectically interact between themselves and with their environment to achieve their goals. Assumingly, the individual citizen's perspective and experience, as well as the perspectives of citizens as collective groups, are crucial to better understanding factors that influence them to engage with OGD.

Third, the experience of individual citizens who engage with OGD in actual, real engagement initiatives is essential to grasp and analyze factors that influence citizen engagement with OGD. Aggregating these experiences is also crucial to assess whether the factors apply to a more significant number of citizens who engage with OGD. Collecting and analyzing data for these two purposes requires a pluralistic, mixed-methods approach, combining qualitative and quantitative approaches. Pragmatism is open to different, mixed methods as an alternative to positivism and interpretivism (Johnson, Onwuegbuzie, & Turner, 2007; Teddlie & Tashakkori, 2009; Venkatesh et al., 2013).

2.3. Mixed methods research

Mixed methods research (MMR) can be defined as "research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study" (Tashakkori & Creswell, 2007, p. 4). In general, MMR is suitable for research problems in which one particular type of data source may be insufficient (Creswell & Plano Clark, 2018). Social science researchers have widely applied and developed MMR (Small, 2011), including in the IS domain (Venkatesh et al., 2013).

The need for a particular MMR design should emerge from the research questions and problems that elaborate on the research's aim or purpose (Creswell & Plano Clark, 2018). According to Venkatesh et al. (2013), MMR may serve one or more of the following purposes: *complementarity*, *completeness*, *developmental*, *expansion*, *confirmation*, *compensation*, and *diversity* (see Table 2.2). The design of an MMR is heavily contingent on the purposes of the research. Researchers should be aware of the different objectives and explicitly outline their goals to understand better their research outcomes (Venkatesh et al., 2013).

Table 2.2. Purposes of MMR; adapted from Venkatesh et al. (2013).

Purposes	**Description**
Compensation	Researchers use mixed methods to compensate for the weaknesses of one approach by using the other.
Complementarity	Researchers use mixed methods to gain complementary perspectives about similar phenomena or relationships.
Completeness	Researchers use mixed methods designs to ensure that they can obtain a complete picture of a phenomenon.
Corroboration/ Confirmation	Researchers use mixed methods to assess the credibility of inferences obtained from one approach.
Developmental	Researchers use an approach's inferences to create research questions for the proceeding approach or test hypotheses that emerged from a previous approach in the next one.
Diversity	Researchers use mixed methods to obtain divergent views of the same phenomenon.
Expansion	Researchers use mixed methods to extend or expand the understanding obtained in a previous approach of a study.

MMR is chosen as the approach to this research based on the following reasons. First, MMR yields more robust inferences than a single method (Teddlie & Tashakkori, 2009). Collecting both quantitative and qualitative data helps researchers make accurate and better conclusions that integrate findings from both quantitative and qualitative strands. Second, based on Venkatesh et al.'s (2013) classification, the purpose of this research is mainly *developmental*, i.e., developing a model that researchers can use for investigating factors that influence citizen engagement with OGD. To better create the contextualized model, as explained in Section 1.5, qualitative data are initially collected and analyzed, and then quantitative data are administered to assess the model to a sample.

In summary, the pragmatism belief advocates using the MMR approach, and the approach is deemed appropriate to attain the objective of this research. Since the nature of this study is developmental, the *exploratory sequential mixed methods design* (Creswell & Creswell, 2018) involving a sequence of qualitative research followed by quantitative research is applied. The following section describes the research design, illustrates the strategies, including research phases, and explains how each study stage is conducted.

2.4. Research design

When developing a research strategy for MMR design, the mixed methods design should be carefully selected based on its suitability for the research questions, objectives, and contexts (Venkatesh et al., 2013). As explained in Section 1.5, this research aims to develop a model to investigate factors that influence citizen engagement with OGD. According to Venkatesh et al. (2013) classification of MMR purposes (see Table 2.2.), this research's objective is mainly *developmental*. Furthermore, this research's nature is exploratory; factors that influence citizen engagement with OGD, positively (drivers) or negatively (inhibitors), have only been patchily investigated to date. Therefore, this research is initially designed to identify the influencing factors from current literature and explore both a priori and emerging factors in real-life OGD engagement cases. Subsequently, this research is intended to explore further the factors in a larger sample of citizens who have experience engaging with OGD and provide hypotheses of factors and citizens' intention to engage with OGD. Finally, this research is designed to test the hypotheses using these samples of citizens. Overall, this design fits the developmental purpose of the study.

According to Creswell and Creswell (2018), the sequence of approaches used in this research is exploratory sequential mixed methods design. As depicted in Figure 2.1, the design is adopted and modified to fit the three phases of this research. The first phase of this research aims to better understand citizen engagement with OGD by systematically analyzing the current literature, identifying factors that drive and inhibit citizen engagement, and developing a theoretical framework of factors. The second phase explores factors identified in the theoretical framework in real OGD engagement cases and identifies emerging factors missing from the framework. The second research phase aims to formulate hypotheses of factors and citizens' intention to engage with OGD. The purpose of the third research phase is to assess and validate the hypotheses and the research model that predicts the factors influencing citizens' intention to engage with OGD. This three-phase research design is primarily a qualitative research sequence, conducted in the first two research phases, followed by quantitative analysis in the last stage. These early two stages emphasize the qualitative approach of the second research phase because its outcomes serve as the basis for developing quantitative instruments used in the last research stage. The following sections briefly describe the three phases of exploratory sequential MMR design.

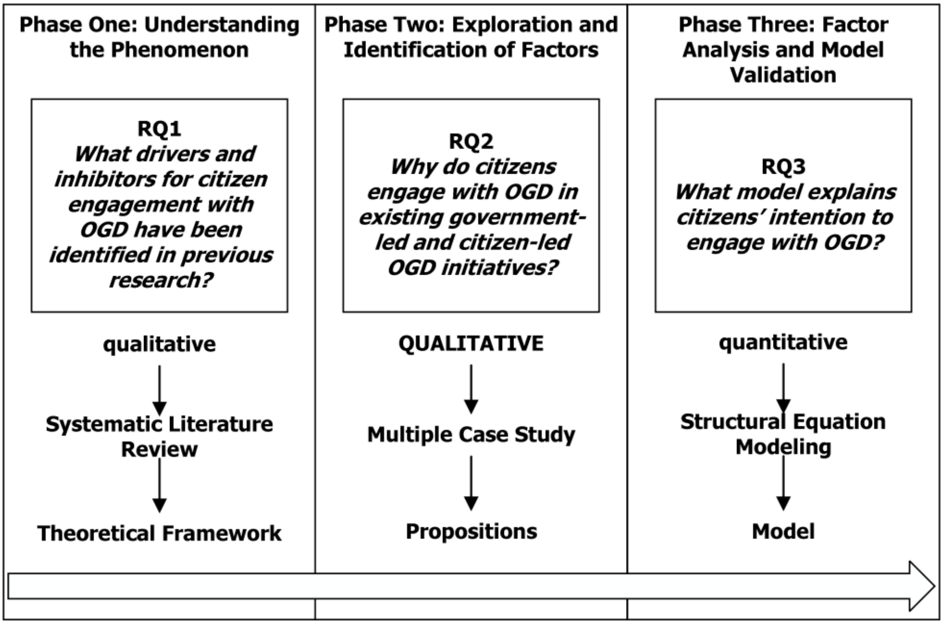


Figure 2.1. The exploratory sequential mixed methods design of this research.

2.4.1. Phase one: Understanding the phenomenon

The first research phase answers the first research question (RQ1), i.e., "what factors drive and inhibit citizen engagement with Open Government Data?" A literature review method is utilized to answer this question. Hart (1998) defined literature review as "the selection of available documents 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" (p. 13). A useful literature review produces a sound basis for progressing knowledge (Webster & Watson, 2002). This study mainly follows Kitchenham and Charters' (2007) Systematic Literature Review (SLR) method combined with Webster and Watson's (2002) backward and forward search. Section 3.1 describes the way this approach is applied.

The current literature does not provide good insights into factors that drive citizens to engage with OGD and inhibit them from engaging with OGD, and scholars rarely integrate these factors into a comprehensive model (Hossain et al., 2016). Therefore, in the conduct of the SLR, the following themes around citizen engagement with OGD are explored: 1) definition, 2) citizens' profiles, 3) typical manifestation, 4) drivers, and 5) inhibitors. Based on these elements, a unified theoretical framework for analyzing factors that influence citizen engagement with OGD is developed as an outcome of the research phase. This theoretical framework is used to create a case study protocol for the second research phase. Section 3.5 explains the theoretical framework.

2.4.2. Phase two: Exploration and identification of factors

The second research question (RQ2), i.e., "why do citizens engage with Open Government Data?" is addressed in this research phase. The purpose of this research phase is to investigate the reasons and motivations behind citizen engagement with OGD. A multiple case study research method is employed to answer question RQ2. Researchers can use case studies to examine real-world situations over which they have little or no control (Yin, 2014). The multiple-case study design is selected because scholars usually consider its evidence more convincing, and the overall study is stronger than a single-case study (Herriott & Firestone, 1983). This study follows Yin's (2014) recommendations on conducting a multiple case study research method (see Section 4.1 for a complete description).

In this research phase, factors identified in the theoretical framework, developed in the previous research phase, are explored, and missing factors are identified in real OGD engagement cases. Eisenhardt (1989) argues that case study research focuses on understanding the dynamics within single settings. Therefore, a case protocol built on the previous research phase's theoretical framework primarily guides the cases investigated in this research phase. Currently, the common manifestations of OGD citizen engagement are government-led engagement such as hackathons or innovation contests (e.g., Concilio et al., 2017; Juell-Skielse et al., 2014). However, in practice, citizen-led OGD engagement exists (e.g., Graft et al., 2016). In the case protocol, selecting a case is mainly contingent upon these two different types of OGD engagement.

The outcomes of this research phase are propositions derived from the theoretical framework developed in the previous research phase and reconciled with those emerging in the cases and missing from the framework. In the following research phase, these factors are evaluated using a larger sample of citizens who engage with OGD.

2.4.3. Phase three: Factor analysis and model validation

The third research phase answers the third research question (RQ3), i.e., "Which factors influence citizens' intention to engage with Open Government Data?" This research phase aims to test the hypotheses and assess the OGD citizen engagement model formulated from the propositions proposed in the previous research phase. The test and the assessment are conducted using a larger sample of citizens who have experience in OGD engagement responding to the survey developed in this research phase. To answer question RQ3, a Structural Equation Modeling (SEM) technique, namely Partial Least Squares (PLS)-SEM, is applied using the computer program "SmartPLS 3" (Ringle, Wende, & Becker, 2015). SEM is a technique that simultaneously examines numerous types of associations between dependent and independent variables (Gerbing & Anderson, 1988; Tabachnick & Fidell, 2019; Urbach & Ahlemann, 2010). SEM allows researchers to validate the hypothesized causation among a group of dependent and independent constructs (structural model) and the loadings of observed items on their expected latent variables (measurement model) (Gefen, Straub, & Boudreau, 2000; Hair, Black, Babin, & Anderson, 2014). Hair, Black, et al. (2014) describe factor analysis as "an interdependence technique whose primary purpose is to define the underlying structure among the variables in the analysis" (p.92). The purpose of this final research phase is to examine the relationships between factors and citizens' intention to engage with OGD. As a result, both the structure and the measurement of the OGD citizen engagement model candidates must be assessed. Therefore, SEM is used.

An online survey based on the list of factors produced in the previous research phase is developed. Groves et al. (2009) describe a survey as "a systematic method for gathering information from (a sample of) entities for the purposes of constructing quantitative descriptors of the attributes of the larger population of which the entities are members." The survey is distributed among citizens from different OGD user communities to collect data from those experienced in engaging with OGD.

This study follows the guidelines for using the SEM technique proposed by different scholars such as Gefen et al. (2000) and Gefen, Rigdon, and Straub (2011), and particularly Hair, Hult, Ringle, and Sarstedt (2017) for PLS-SEM. The outcome of the final research phase is the assessed model of OGD citizen engagement that can explain factors influencing citizens' intention to engage with OGD. The model contributes to open data literature because it operationalizes the factors influencing citizens to engage with OGD in different settings: government-led and citizen-led initiatives. The model also contributes to practitioners and open data policymakers by offering valuable insights into the influencing factors of citizen engagement that should be considered when designing OGD programs.

3. Theoretical Frameworks

In the previous chapter, the design and strategies for carrying out the study, consisting of three research phases, have been laid out. This chapter reports the first research phase that aims to answer the first research question: *what drivers and inhibitors for citizen engagement with OGD have been identified in previous research?* Kitchenham and Charters' (2007) Systematic Literature Review (SLR) approach is used in this phase to understand better what has already been known about the phenomenon under study: citizen engagement with OGD. Notably, the purpose of the SLR approach is to identify factors that drive and inhibit citizen engagement and develop an integrated theoretical framework of factors. Firstly, this chapter describes the particular SLR approach. Subsequently, the theoretical background of this research, encompassing the definitions of OGD citizen engagement, profiles of citizens who engage with OGD, types of OGD citizen engagement, and intention-based theories used to study OGD citizen engagement, is introduced. Next, the factors that drive citizens to engage with OGD and those that inhibit citizens from engaging with OGD are described and summarized. Then, the proposed theoretical framework for studying the factors that influence citizen engagement with OGD based on the literature review synthesis is introduced and explained. Finally, the research phase is concluded in the final section of this chapter. We have published parts of this chapter in Purwanto, Zuiderwijk, and Janssen (2020a).

3.1. Systematic literature review approach

A literature review is an integral part of the research process, including research conducted through the view of pragmatism (Teddlie & Tashakkori, 2009). The researcher followed Kitchenham and Charters' (2007) SLR guidelines and Webster and Watson's (2002) backward and forward search strategy. The researcher developed a review protocol functioning as the blueprint for carrying out the review. A review protocol "specifies the research question being addressed and the methods that will be used to perform the review" (Kitchenham et al., 2009, p. 4). Developing the protocol is a critical step required before the work of an SLR can start (Okoli & Schabram, 2010). Although the processes and approaches planned in the protocol can be amended during the review, developing a protocol is crucial for an SLR to minimize researcher bias. A review protocol mainly consists of search strategies, inclusion and exclusion criteria, data collection and analysis, and synthesis (see Figure 3.1). The elements of the protocol are presented in the following sections.

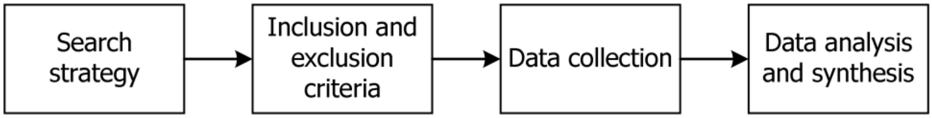


Figure 3.1. The adopted SLR protocol for this research.

In this research, the researcher used an SLR approach for the following purposes. First, the SLR approach aims to position this research relative to current knowledge and expand on this knowledge. For this purpose, the contexts of citizen engagement with OGD investigated by previous research need to be understood. Specifically, the contexts mainly related to the profiles of citizens who engaged with OGD (e.g., demographic background, capabilities, and roles) and the types of OGD citizen engagement (e.g., citizen-led, government-led). The engagement context is crucial for understanding different OGD actual settings that might drive or inhibit citizen engagement. Insights into the profiles of citizens that include the demographic background (e.g., age) and capabilities (e.g., occupation) are also important to comprehend who the citizens are (Sieber & Johnson, 2015; Susha, Grönlund, et al., 2015). For instance, citizens from a particular background (e.g., highly educated young men) with specific capabilities (e.g., programming skills) may engage with OGD to build an application. Moreover, our knowledge of citizen engagement with OGD is currently typically limited to hackathons or innovation contests (e.g., Juell-Skielse et al., 2014). On the other hand, practitioners recognize the existence of citizen-led OGD engagement in real life (e.g., Young & Verhulst, 2016). Some citizens may prefer to participate in a hackathon than a citizen-led OGD engagement.

Second, the purpose of the SLR is to accumulate empirical evidence relating to individual citizens' drivers and inhibitors of OGD engagement and summarize it. Consequently, the researcher reviews previous studies investigating the factors that drive individual citizens to engage with OGD and those that examine the factors that inhibit individual citizens from engaging with OGD. The insights obtained from the review enable the development of a theoretical framework that guides the subsequent research phases in this study.

3.1.1. Search strategy

Kitchenham and Charters (2007) suggest that researchers take the following three steps into account when developing a *search strategy*. First, researchers must consider all possible words with similar meanings when formulating the literature search terms. Second, researchers have to select relevant academic publication databases to carry out the literature search. Third, researchers must adopt and modify the wording of the search strings according to each database's field codes. Table 3.1 summarizes the overall search terms applied in the SLR. A search string constituting words that have a relatively similar meaning to "engagement" and "open government data" was developed (see Figure 3.2 for a complete overview).

("open government data" OR "public sector information" OR "open data" OR "public data" OR "public government data" OR "open public sector data" OR "open public data" OR "big open data" OR "big open public sector data" OR "open public sector information" OR "open government information") AND (accept* OR use OR usage OR using OR adopt* OR engag* OR participat* OR involv*)

Figure 3.2. The complete query of search strings used in the SLR approach.

The terms *participation* and *involvement* were included in the search string because *user engagement* is also closely related to *user participation* and *user involvement* in the IS research domain (Hwang & Thorn, 1999; Kappelman & McLean, 1992). Citizen engagement with OGD implies that an individual citizen has to adopt OGD – a process that starts from accepting OGD and ends with making full use of it (Renaud & van Biljon, 2008). Therefore, words, such as *acceptance*, *adoption*, and *use*, resembling engagement, were added.

Table 3.1. The search terms used in the SLR.

Engagement	**Open Government Data**
*Engag** (including *engage*, *engaging*, *engagement*)	*Open government data*
*Participat** (including *participate*, *participating*, *participation*)	*Public sector information*
*Involv** (including *involve*, *involves*, *involving*, *involvement*)	*Open data*
*Accept** (including *accept*, *accepting*, *acceptance*)	*Public data*
*Adopt** (including *adopt*, *adopting*, *adoption*)	*Public government data*
Use, *usage*, *using*	*Open public sector data*
	Open public data
	Big open data
	Big open public sector data
	Open public sector information
	Open government information

Scopus and Web of Science databases were used to collect information about papers relevant to the SLR. Scopus is known to index prominent publishers of peer-reviewed academic articles, including ACM, Cambridge University Press, Emerald, IEEE, Oxford University Press, Sage, ScienceDirect (Elsevier), Springer, Taylor & Francis, and Wiley-Blackwell. The reported search string above (see Figure 3.2) was applied to the publications' titles, abstracts, and keywords. Google Scholar could also have been used as a database for searching relevant publications. However, it was not employed in the SLR for the following reasons. First, Google Scholar limits the length of search strings to 256 characters. We cannot search for articles using the terms specified in Table 3.1 because their size exceeds 300 characters; Boolean operators needed to formulate the complete search string are not even included in this maximum. Second, Google Scholar cannot differentiate the search results based on the type of publication (i.e., journal articles, books, conference papers, government reports) without examining them individually. Third, Google Scholar does not allow us to export its search results in bulk as structured citation records. We have to extract the citation information of each result; it is time-consuming, especially when we are working with thousands of results.

3.1.2. Selection criteria

The review was started on the literature published from 2009 onwards. This decision was made because open data researchers have revealed that there has been a sharp increase in the number of publications using the term *open data* since 2009 (Zuiderwijk, Helbig, Gil-García, & Janssen, 2014). This increase is likely because of President Obama's (2009) first executive order on *Transparency and Open Government* at the beginning of 2009, inspiring the adoption of open data provision programs worldwide (Huijboom & Van den Broek, 2011). Therefore, the decision to use 2009 as a selection criterion is deemed defensible.

On the one hand, publications that employ empirical research methods and provide a clear explanation about the research methods (e.g., case study, experiment, survey) were included. On the other hand, document analyses such as literature reviews and conceptual articles were excluded. Papers that focus on citizen engagement with OGD were included, and those investigating OGD provision or OGD usage by businesses, governmental organizations, or civil society organizations were excluded. Irrelevant technical articles from various research fields, including astronomy, medical ethics, or physics, were also excluded.

3.1.3. Metadata extraction

Table 3.2 summarizes the data extracted from each paper included in the literature review.

Table 3.2. The data extracted from each paper included in the SLR.

Publication data	Publication title, name of the authors, abstract, keywords, type of publication (journal or conference paper), name of publication outlet, publication year, research approach, data collection, and analysis method
Context of the study	Country or city understudy, period of data collection, and OGD domain
Citizen's profiles	Respondent demographics and capabilities/occupation
Types of engagement	The setting of the OGD engagement, engagement activities, and outputs or outcomes of OGD engagement
Driving factors (drivers)	Empirical evidence related to demand, factors, interests, motivations, needs, or purpose that drive individual respondents to engage with OGD
Inhibiting factors (inhibitors)	Empirical evidence related to barriers, challenges, difficulties, impediments, or problems that inhibit individual respondents from engaging with OGD

Initially, 8450 papers from Scopus (n=2589) and Web of Science (n=5861) databases were obtained. Six inclusion/exclusion stages were carried out upon the retrieved publications. First, irrelevant papers were filtered out by examining the publication source and title, followed by removing duplicate entries from the two databases; 7412 articles were eliminated, and 1038 articles were retained in this step. Second, the retrieved publications were examined to determine whether they were empirical research papers; those that were not were excluded, including conceptual articles, literature reviews, research notes, and technical reports. Sixty-eight publications were dropped, and 970 publications were held in this step. Third, the relevancy of the publications' abstracts was evaluated, and papers containing irrelevant subject matters, including blockchain, machine learning, and virtual reality-based participation, were excluded. Eight hundred sixty-two papers were excluded, and 108 papers were included in this step. Fourth, the publications' contents were assessed for their relevancy, and papers irrelevant to this study (e.g., those that focused on the use of open data by companies or civil society organizations) were excluded. Sixty-five articles were removed, and 43 articles were kept in this step. Fifth, backward and forward searches were carried out upon the included papers using the Scopus by retrieving the publications' references and retrieving articles citing the publications, respectively. The backward and forward search aims to identify more relevant papers as the researcher may not immediately discover them in the primary search (Webster & Watson, 2002). The additional papers retrieved from the backward and forward search were not found in the first search instance because they do not contain the predetermined keywords. Sixth, the first four stages of the inclusion/exclusion process were repeated upon the publications retrieved from the backward and forward search. In the end, 52 publications were selected for the SLR (see Appendix A for a complete overview).

3.2. Theoretical background

3.2.1. Citizen engagement with OGD

Citizen engagement has been widely and traditionally studied in the PA domain (e.g., Arnstein, 1969). In the OGD domain, as depicted in Table 3.3, scholars have no consensus over the definition of citizen engagement. However, based on these definitions, we can conjecture that citizen engagement with OGD has three elements. First, citizen engagement refers to *converting OGD* into artifacts (Susha, Grönlund, et al., 2015; Zuiderwijk, Janssen, et al., 2015) such as facts, information, interface, service, and new data (Davies, 2010). Second, such engagement requires *collaboration*. Collaboration can occur between citizens and other OGD users (Sayogo, Pardo, & Cook, 2014), such as OGD providers (Zuiderwijk, Janssen, et al., 2015), or among citizens themselves. Third, the primary purpose of such engagement is to *create value,* such as helping personal decision-making, improving government processes, or providing community service (Susha, Grönlund, et al., 2015; Zuiderwijk, Janssen, et al., 2015).

Table 3.3. Definitions of citizen engagement in the OGD domain.

Reference	**Definition**
Sayogo et al. (2014)	Interactions among different OGD users (e.g., users, governments, non-profit organizations, businesses) facilitated with collaborative infrastructures (e.g., tools, methods, systems)
Susha, Grönlund, et al. (2015)	Activities carried out to convert data to other objects (i.e., fact, information, data, interface, and service) involving a different level of complexity and creating different value (e.g., individual, societal)
Zuiderwijk, Janssen, et al. (2015)	Interactions between open data providers and users for realizing the understanding of open data and creating value (e.g., improving government processes, services, and decision-making)

In the PA domain, researchers typically use the citizen engagement, *citizen involvement*, and *citizen participation* terms interchangeably to refer to the same process through which citizens express their opinions on the public policy decisions (Cogan & Sharpe, 1986; Rowe & Frewer, 2005). At the same time, *user engagement*, *user involvement*, and *user participation* are clearly defined in the IS domain. Barki and Hartwick (1989) define user involvement as “the importance and personal relevance that users attach either to a particular system or to IS in general, depending on the users' focus" (p. 59-60). Therefore, involvement can be associated with "a subjective psychological state of the individual" (Barki & Hartwick, 1989, p. 59). Barki and Hartwick (1989) define user participation as "the behaviors and activities that the target users or their representatives perform in the systems development process" (p. 59). Kappelman and McLean (1992) define user engagement as "the total set of user relationships toward information systems and their development [which] includes both user participation (the behavior) and user involvement (the attitude)" (p. 2). Based on these conceptualizations, it can be conjectured that citizen engagement with OGD, as defined in Table 3.3, is related to citizens participating in activities to convert OGD and their psychological state of importance and relevance to participate in such activities.

Building on the previous discussions, citizen engagement with OGD is defined as *citizens' collaborative activities to convert OGD into valuable artifacts that are important and relevant to them and society*.

3.2.2. Profiles of citizens who engage with OGD

The profiles of citizens who engage with OGD are crucial for understanding the OGD engagement (Johnson & Robinson, 2014, p. 355). A citizen's profile can be described by her or his age, awareness of OGD provision, education level, experience with OGD, gender, occupation, and resources (see Table 3.4 for a complete overview). Previous research showed that particular profile elements could indicate whether or not a citizen is likely to engage with OGD. For example, the older citizens are, the less they are likely to engage with OGD (Wijnhoven, Ehrenhard, & Kuhn, 2015), whereas male citizens are more likely to engage with OGD (Saxena & Janssen, 2017). Another example shows that citizens with higher educational qualifications are more likely to engage with OGD (Wang, Richards, & Chen, 2019). At the same time, the lack of particular elements may inhibit citizens from engaging with OGD. For instance, citizens who lack financial, educational, and infrastructural resources are less likely to engage with OGD (Hjalmarsson et al., 2014; Khayyat & Bannister, 2017; Martin, 2014; Ruijer et al., 2017; Wijnhoven et al., 2015). Another example shows that a lack of awareness in the OGD availability and provision has led to no citizen engagement (Canares, 2014).

Previous research also found that citizens with particular occupations, such as students, specialists, and human resource workers, are more likely to engage with OGD (Wang et al., 2019). Furthermore, citizens with different occupations are driven by different motivations to engage with OGD (Purwanto et al., 2019; Smith & Sandberg, 2018). For example, citizens, who are entrepreneurs, are likely to be driven by economic factors to engage with OGD to earn money

(Smith & Sandberg, 2018). Another example concerns developers who are likely to be caused by the desire to have fun and a feeling of joy in exploring OGD (Purwanto et al., 2019).

Table 3.4. The profiles of citizens who participate in OGD engagement.

Element	Findings Related to Engagement
Age	Older citizens are less willing to engage (Wijnhoven et al., 2015)
Awareness of OGD Provision	Lack of interest (Osagie et al., 2017), low awareness of the availability of OGD provision (Canares, 2014), lack of demand (Martin, 2014), little data literacy (Hivon & Titah, 2017) are likely to inhibit citizens from engaging with OGD
Education level	Citizens with a higher educational level are more willing to engage with OGD (Wang et al., 2019)
Experience with OGD	Citizens who have previous engagement experience will likely to engage with OGD again (Hutter, Füller, & Koch, 2011; Purwanto, Zuiderwijk, & Janssen, 2019); lack of experience will likely inhibit citizens from engaging with OGD (Veeckman & van der Graaf, 2015; Zuiderwijk, Janssen, & Susha, 2016)
Gender	Male citizens are more willing to engage with OGD (Saxena & Janssen, 2017)
Occupation	Students, specialists, and human resource workers are more willing to engage with OGD (Wang et al., 2019)
Resources	Lack of time (Hjalmarsson et al., 2014; Khayyat & Bannister, 2017; Ruijer et al., 2017; Smith & Sandberg, 2018), lack of resources (financial, educational, and infrastructural) (Hjalmarsson et al., 2014; Khayyat & Bannister, 2017; Martin, 2014; Ruijer et al., 2017; Wijnhoven et al., 2015), lack of knowledge (Martin, 2014; Ruijer et al., 2017; Wijnhoven et al., 2015), lack of skills (Ruijer et al., 2017) are likely to inhibit citizens from engaging with OGD

Generally, the citizens' occupations require the possession of capabilities to enable them to carry out their works. For instance, citizens who work as data scientists should have programming, statistical analysis, database manipulation, and data visualization capabilities. Although particular capability may only be observed during an OGD engagement, it can contribute to the way citizens engage with OGD. For example, a citizen working as a programmer can develop applications, write codes and at the same time analyze data sets. Academia is an occupation resembling the profiles of citizens who engage with OGD. Most open data studies' respondents were from academia including faculty members, researchers, students, and teachers (e.g., Beno, Figl, Umbrich, & Polleres, 2017; Charalabidis, Loukis, & Alexopoulos, 2014; Martin, 2014; Saxena & Janssen, 2017; Zuiderwijk, Janssen, et al., 2015). Other occupations typically related to citizen engagement with OGD include developers (e.g., Hivon & Titah, 2017; Juell-Skielse et al., 2014) and professionals such as managers or experts (e.g., Benitez-Paez, Degbelo, Trilles, & Huerta, 2018; Wang, Richards, & Chen, 2018).

3.2.3. Types of OGD citizen engagement

In the PA domain, researchers differentiate citizen engagement initiatives that are self-organized (*citizen-led*) and *government-led* (Edelenbos, van Meerkerk, & Schenk, 2018). Government-led citizen engagement has been widely known as the conventional type of engagement (cf. Arnstein, 1969; Cunningham, 1972; Day, 1987; Roberts, 2004). The core feature of this engagement is that governments typically determine when and under which conditions citizens can engage and the extent to which their suggestions are adopted (King & Cruickshank, 2012). Citizen-led engagement is not conceptually new; in her seminal work, “A Ladder of Citizen Participation,” Arnstein (1969) places citizen control as the ideal type of engagement. Nevertheless, there has been a surprising lack of literature on how it works in practice.

The distinction between government-led and citizen-led also applies to OGD engagement. Either type of engagement is heavily contingent upon the model in which governments operate in OGD provision (Sieber & Johnson, 2015). Government-led engagement is a typical example of the *government as an open data activist* model in which governments provide the OGD infrastructure and actively promote its use to citizens (Sieber & Johnson, 2015). Open data hackathons or innovation contests are typical examples of government-led OGD engagement (Purwanto, Zuiderwijk, & Janssen, 2018b). Open data hackathons or innovation contests are usually offline competitions funded by governmental organizations in centralized locations bringing citizens with various backgrounds (Concilio et al., 2017; Hartmann et al., 2016). These citizens work intensively together in small groups for a short period (e.g., 12 hours, 24 hours, two days) to produce artifacts using OGD. Generally, each group has to present their final idea or prototype, or analysis at the contest's end. Winning teams judged by juries would typically earn the competed prize (e.g., money, investment, support).

Citizen-led OGD engagement is likely a consequence of the *government as a platform* model. In this model, governments only provide OGD infrastructures such as portals offering access to data and tools for working on data (Sieber & Johnson, 2015). The government plays passive roles and assumes that citizens will eventually engage with OGD and create value from it (Linders, 2012). Occasionally, government-led engagement can be inefficient due to disproportionate power distribution between citizens and governments (Hivon & Titah, 2017), while citizen-led engagement can be successful (Porwol, Ojo, & Breslin, 2013). Citizen-led engagement is a kind of citizens’ reactions to government-led processes or structures utilizing the states’ instruments to obtain citizens’ objectives (Edelenbos et al., 2018).

Table 3.5 illustrates the overview of government-led and citizen-led OGD engagement and examples of such engagement. Generally, government-led engagement materializes in government-sponsored online participation (Hutter et al., 2011), open data hackathons (Juell-Skielse et al., 2014), or fellowship (Maruyama, Douglas, & Robertson, 2013). At the same time, citizen-led engagement may be identified based on its outcomes, such as application developed for election (dos Santos Brito, dos Santos Neto, da Silva Costa, Garcia, & de Lemos Meira, 2014), defense contract data analysis (Whitmore, 2014), and humanitarian mapping crowdsource (Dittus, Quattrone, & Capra, 2016).

Table 3.5. The overview of the OGD engagement type and its examples/outcomes.

Type of Engagement	Example / Outcomes	Source(s)
Government-led	Hackathons / innovation contests	Juell-Skielse et al. (2014), Kuk and Davies (2011), Hivon and Titah (2017), Gama (2017), de Deus Ferreira and Farias (2018), Choi and Tausczik (2017), Hjalmarsson et al. (2014), Purwanto et al. (2019),
	Online ideation	Hutter et al. (2011), Schmidthuber, Piller, Bogers, and Hilgers (2019), Wijnhoven et al. (2015)
	Fellowship	Maruyama et al. (2013)
	Promotion	Hellberg and Hedström (2015)
Citizen-led	Application development	dos Santos Brito et al. (2014), Purwanto, Zuiderwijk, and Janssen (2018a), Smith, Ofe, and Sandberg (2016), Rudmark et al. (2012)
	Service	Smith and Sandberg (2018)
	Service design	Jarke (2019)
	Mapping	Dittus et al. (2016)
	Data analysis	Whitmore (2014)

3.2.4. Intention to engage with OGD

The theories used by the reviewed studies to investigate OGD citizen engagement are remarkably diverse (see Table 3.6 for examples from the SLR). These studies applied particular theories or theoretical models from other fields such as IS to develop a research framework/model, test hypotheses, or reflect upon their results. The researcher suggests that the current studies of citizen engagement with OGD do not focus on theory development and that the field is exploratory to some extent. From the reviewed papers, we can see that some studies develop research models and test hypotheses by combining more than one theory or theoretical model. Regardless of the reviewed studies' different theories and theoretical models, they are essentially rooted in the intention-based theory, namely the theory of planned behavior (TPB). TPB is the underlying theory of almost all theories and theoretical models frequently used in the open data literature to understand the factors influencing citizen engagement with OGD. For example, the technology acceptance model, IS success model, and unified theory of acceptance and use of technology are built on and extend TPB. Nevertheless, the integration of different theories or consistent application of particular intention-based theories is lacking.

Table 3.6. The overview of intention-based theories or theoretical models used by the reviewed studies.

Theory or theoretical model	Source(s)
Technology Acceptance Model	Charalabidis et al. (2014), Jurisch et al. (2015), Weerakkody, Kapoor, Balta, Irani, and Dwivedi (2017), Fitriani, Hidayanto, Sandhyaduhita, Purwandari, and Kosandi (2019), Wang et al. (2018), Wirtz, Weyerer, and Rösch (2018), Wirtz, Weyerer, and Rösch (2019)
Unified Theory of Acceptance and Use of Technology	Jurisch et al. (2015), Zuiderwijk, Janssen, et al. (2015), Zuiderwijk and Cligge (2016), Saxena and Janssen (2017), Talukder, Shen, Talukder, and Bao (2019)
Information System Success Model	Charalabidis et al. (2014), Fitriani et al. (2019), Talukder et al. (2019)
Diffusion of Innovations	Jurisch et al. (2015), Weerakkody, Irani, Kapoor, Sivarajah, and Dwivedi (2017), Khurshid, Zakaria, Rashid, and Shafique (2018)
Theory of Planned Behavior	de Deus Ferreira and Farias (2018), Fitriani et al. (2019)

TPB is an intention-based theory that has been extensively applied and used to predict and explain individual behavior within the information systems (IS) field (Al-Lozi & Papazafeiropoulou, 2012). Ajzen (1991) develops TPB as an extension of Ajzen and Fishbein's (1980) theory of reasoned action (TRA). TRA postulates that the precursor of any behavior is an individual intention. The strength of the individual’s intentions to carry out a particular behavior correlates with the likelihood of the behavior.

The individual’s intention to perform particular behaviors is the core of TPB. Ajzen (1991) assumed that intention captures the motivational factors that influence behavior. Intentions are indicators of people’s willingness, effort, or readiness to perform the behavior (Fishbein & Ajzen, 2010, p. 39). Readiness to engage in a behavior incorporates *willingness*, *behavioral expectation*, and *trying* concepts measured using responses that reflect the strength of the intention. TRA posits that the antecedents of intentions are the strength of individual attitudes towards the particular behavior and subjective norms concerning the social pressure of whether to perform the behavior or not. TRA claims that individuals’ beliefs on the outcomes of a particular behavior and their evaluation of these outcomes determine attitudes towards the behavior. At the same time, an individual’s normative beliefs and motivation to conform to the social norms determine subjective norms. Moreover, TRA postulates that all other factors influence the behavior indirectly via the attitude or subjective norms. Researchers refer to these factors as external variables such as the IS user's characteristics, political influences, or types of IS development implementation (Davis, 1989).

Ajzen (1991) develops TPB that extends TRA by adding the perceived behavioral control (PBC) construct. PBC concerns the individual’s perceived ease or difficulty performing the behavior (Orbell, Hodgkins, & Sheeran, 1997). Ajzen (1991) suggests the PBC’s compatibility with Bandura's (1982) perceived self-efficacy concept. Bandura (1982) defines self-efficacy as “judgements of how well one can execute courses of action required to deal with prospective situation” (p. 122). The individual’s confidence and ability to act strongly influence her or his behavior.

Davis (1989) proposed the technology acceptance model (TAM) built on TPB. TAM is among the widely utilized theories for studying IS adoption or acceptance (Venkatesh, Morris, Davis, & Davis, 2003). The critical element of TAM, similar to TPB, is the individual’s behavioral intention that leads to IS use. TAM postulates that perceived usefulness (PU) and perceived ease of use (PEOU) influence this individual’s intention via the individual’s attitude toward using the IS. PU relates to the extent to which the individual believes that using an IS will help her or him perform her or his job better (Davis, 1989). At the same time, Davis (1989) defines PEOU as how the individual perceives that the IS will be easy to use. However, TAM differs from TPB as it omits the subjective norms variable.

Venkatesh et al. (2003) formulated a unified model integrating elements of eight competing models of technology acceptance to improve its predictive power: the unified theory of acceptance and use of technology (UTAUT). UTAUT integrates the dimensions of TRA, TAM, motivational model, TPB, a model of PC utilization, innovation diffusion theory, and social cognitive theory. Like TAM, UTAUT is also heavily based on the individual’s behavioral intention to accept and use technology. UTAUT postulates that the IS behavioral intention is directly determined by four factors: performance expectancy, effort expectancy, social influence, and facilitating conditions. UTAUT also posits that gender, age, experience, and voluntariness of use moderate the effects of these four constructs. Venkatesh et al. (2003) define performance expectancy as “the degree to which an individual believes that using the system will help him or her to attain gains in job performance” (p. 447). Effort expectancy is defined as “the degree of ease associated with the use of the system”

(Venkatesh et al., 2003), while social influence refers to the extent to which "an individual perceives that important others believe he or she should use the new system" (Venkatesh et al., 2003). At the same time, Venkatesh et al. (2003) define facilitating conditions as to how an individual believes that organizational and technical infrastructures are provided to facilitate her or him using the system.

Although built on information theories, DeLone and McLean's (2003) IS success model also has an essential intention element. DeLone and McLean (2003) introduce this dimension as an alternative for *use* construct in their original model (DeLone & McLean, 1992). In the updated model, DeLone and McLean (2003) postulated that an individual's intention to use an IS is determined by three IS quality variables: system quality, information quality, and service quality. System quality assesses the quality of the system's performance from the engineering-oriented perspective, while information quality evaluates the quality of the information created by the system (DeLone & McLean, 1992). Service quality assesses the quality of the IS help/support function (DeLone & McLean, 2003). The theory posits that these factors will singularly or jointly influence the individual's intention to use the IS.

In the context of this research, the discussed theories and theoretical models offer insights into the determinant of OGD engagement: individual citizens' intention to engage with OGD. Therefore, the citizens' intention to engage with OGD is the central construct of this research. This study investigates the factors that drive citizens to engage with OGD and the factors that inhibit them from engaging with OGD using the intention-based theoretical framework.

3.3. Factors driving individual citizens to engage with OGD

This section reports the drivers of citizen engagement with OGD, i.e., the factors that drive individual citizens to engage with OGD, according to the results of the SLR. When collecting the existing evidence concerning these drivers from the reviewed papers, we noticeably see that most papers did not focus on analyzing and reporting the drivers. As a result, empirical data indicating the factors driving individual citizens to engage with OGD and citizens' *demand*, *need*, *interests*, the *purpose* of using OGD, and *motivation* for participating in OGD engagement, were extracted.

In the end, different factors were found. Some factors can be clearly defined, such as the OGD portal's user-friendliness and OGD relevance, while other factors are similar, such as the perception of relative advantage and usefulness. Subsequently, six categories of factors were developed to reflect better the relevant definition of the identified factors and group factors that share similar, relevant and particular characteristics. The factors classification includes *intrinsic motivations*, *extrinsic motivations*, *economic factors*, *social factors*, *technical factors*, and *political factors*. Table 3.7 provides a detailed summary of the drivers of citizen engagement with OGD.

Table 3.7. The overview of factors driving individual citizens to engage with OGD.

Category	Drivers	Source(s)
Intrinsic motivations	Fun and enjoyment	Rudmark et al. (2012), Juell-Skielse et al. (2014), Wijnhoven et al. (2015), Khayyat and Bannister (2017), de Deus Ferreira and Farias (2018), Smith and Sandberg (2018), Wirtz et al. (2018), Schmidthuber et al. (2019)
	Altruistic motivations	Maruyama et al. (2013), Wijnhoven et al. (2015), Choi and Tausczik (2017), Gama (2017), Hivon and Titah (2017), Khayyat and Bannister (2017), Jarke (2019)
	Intellectual challenge	Kuk and Davies (2011), Rudmark et al. (2012), Juell-Skielse et al. (2014), Khayyat and Bannister (2017), Smith and Sandberg (2018), Wirtz et al. (2018)
	Learning new things	Kuk and Davies (2011), Gama (2017), de Deus Ferreira and Farias (2018), Jarke (2019)
Extrinsic motivations	Perceived benefit	Kuk and Davies (2011), Jurisch et al. (2015), Zuiderwijk, Janssen, et al. (2015), Toots, McBride, Kalvet, and Krimmer (2017), Weerakkody, Irani, et al. (2017), Weerakkody, Kapoor, et al. (2017), Smith and Sandberg (2018), Wirtz et al. (2018), Wirtz et al. (2019)
Economic factors	Economic motives	Kuk and Davies (2011), Khayyat and Bannister (2017), de Deus Ferreira and Farias (2018), Smith and Sandberg (2018)
Social factors	Social influence	Zuiderwijk, Janssen, et al. (2015), Choi and Tausczik (2017), Saxena and Janssen (2017), Weerakkody, Kapoor, et al. (2017)
	Broadening social networks	Hutter et al. (2011), Hellberg and Hedström (2015), Gama (2017), Jarke (2019)
Technical factors	Perceived ease of use	Jurisch et al. (2015), Zuiderwijk, Janssen, et al. (2015), Weerakkody, Kapoor, et al. (2017), Wirtz et al. (2018), Wirtz et al. (2019)
	System quality	Charalabidis et al. (2014), Smith et al. (2016), Osagie et al. (2017), Purwanto et al. (2018a)
	Data quality	Toots et al. (2017), Talukder et al. (2019)
	Service quality	Wijnhoven et al. (2015), Zuiderwijk, Susha, Charalabidis, Parycek, and Janssen (2015), Smith et al. (2016), Osagie et al. (2017), Wang et al. (2018), Talukder et al. (2019)
Political factors	Trust	Fitriani et al. (2019)
	Improvement expectancy	Hutter et al. (2011), Kuk and Davies (2011), Cranefield, Robertson, and Oliver (2014), Wijnhoven et al. (2015)
	Political interests	Hutter et al. (2011), Cranefield et al. (2014), Jurisch et al. (2015), Khayyat and Bannister (2017), Purwanto et al. (2018a), Wang et al. (2019), Wirtz et al. (2019)

Intrinsic motivations were used to classify different motivations to engage with OGD driven by the citizen's inherent interest or enjoyment, while *extrinsic motivations* group motivations caused by external rewards (Deci, 2004). Two groups of factors that wholly and partly overlap with extrinsic motivations were defined: economic and social factors. Factors related to the citizen's motivation to create economic value were grouped into *economic factors* and the motivation to comply with societal and community values and beliefs as *social factors*. *Technical factors* were established to classify the citizen's technical evaluation upon the OGD, the systems that provide access to OGD, and the supports provided to OGD users. Finally, factors concerning the citizen's political evaluation toward the government that provides the OGD were grouped into *political factors*. Researchers rarely integrate these drivers for analyzing and evaluating factors influencing citizen engagement with OGD; researchers typically study them separately.

3.3.1. Intrinsic motivations

Several factors associated with intrinsic motivations that drive citizens to engage with OGD were found. First, *fun and enjoyment* is primary driver of citizen engagement with OGD in different contexts. Having fun or enjoying doing an activity is the central idea of intrinsic motivations (Deci, 2004), where individuals do some activities for the enjoyment derived from doing them (Csikszentmihalyi, 1975). Feeling fun and enjoying the topic of particular open data (e.g., public transport data) drove citizens to participate in open data hackathons in Brazil (de Deus Ferreira & Farias, 2018), Ireland (Khayyat & Bannister, 2017), and Sweden (Juell-Skielse et al., 2014; Rudmark et al., 2012). In the context of participation in open government platforms, having fun influenced some citizens' decision to contribute to collaborative democracy projects (Wijnhoven et al., 2015) and sharing, commenting, and evaluating ideas (Schmidthuber et al., 2019). Engaging with OGD is particularly fun because it enables hobbyists with a programming background to play with the data (Smith & Sandberg, 2018).

Second, intrinsic motivations related to *altruistic motivations* can also influence citizens' intentions to engage with OGD. Altruistically motivated behavior is performed intentionally voluntarily to benefit others without expecting any external reward (Bar-Tal, 1986). The altruistic motivation was found manifesting in different contexts of the reviewed studies. The motivation can be associated with the obligation to the local community. Participants of an OGD-based co-creation initiative suggested that the Irish tradition of "meitheal" (i.e., working together, neighbors helping each other) influences them to participate in the initiative (Khayyat & Bannister, 2017). Another example concerns older adults who wanted to do something more meaningful for their local community by participating in digital public services co-creation in Bremen (Jarke, 2019). Benefiting society by solving a city's problems (Gama, 2017; Hivon & Titah, 2017) is another example of this motivation. It can also be related to the obligation to the country. A sense of civic duty motivated German citizens to participate in open government projects (Wijnhoven et al., 2015) while giving back her country motivated an ex-US marine to join an open data fellowship program (Maruyama et al., 2013).

Third, some citizens felt that *intellectual challenges* to solving particular problems influence their engagement with OGD. For example, creating new digital public services motivated some citizens to participate in an open transportation data hackathon to solve everyday commuting issues in Sweden (Juell-Skielse et al., 2014; Rudmark et al., 2012). Another example concerns challenges of creating new services from different domains for city inhabitants in the greater Dublin area (Khayyat & Bannister, 2017) or taking challenges from purely technical problems (Kuk & Davies, 2011; Smith & Sandberg, 2018).

Fourth, the motivation to *learn new things* that influence citizens to engage with OGD was also identified. Notably, citizens participated in some open data hackathons held in Australia, New Zealand, Brazil, and the UK to learn new things (de Deus Ferreira & Farias, 2018; Gama, 2017; Kuk & Davies, 2011). Surprisingly, the older adults also participated in digital public services co-creation in Bremen because they were motivated to learn new things (Jarke, 2019).

3.3.2. Extrinsic motivations

One primary factor associated with extrinsic motivations that drive citizens to engage with OGD was identified: perceived benefits. In this research stage, the *perceived benefits* term was used to refer to the degree to which citizens perceive that engaging with OGD will bring benefits or advantages to them. This factor's definition shares similarities with relative advantage (Rogers, 1983), usefulness (Davis, 1989), and performance expectancy (Venkatesh et al., 2003). Among the benefits that motivate citizens to engage with OGD are helping them make better day-to-day decisions (Jurisch et al., 2015; Weerakkody, Kapoor, et al., 2017) and the ease of obtaining public information (Wirtz et al., 2019). Benefits such as fulfilling performance or job expectancy also motivate citizens employed by companies (Smith & Sandberg, 2018) or working as social science researchers (Zuiderwijk, Janssen, et al., 2015). In addition, future career prospects can also motivate young citizens, particularly students or those in their early careers (Kuk & Davies, 2011).

3.3.3. Economic factors

The definition of economic factors closely explicates extrinsic motivations because the economic value created from engaging with OGD is the external reward pursued by citizens. However, the economic factors category was mainly created to group economic motives differentiated from the other rewards exemplified in the previous section. Hackers and developers are typically motivated to win the open data hackathon prizes (de Deus Ferreira & Farias, 2018; Kuk & Davies, 2011), which are usually a sum of money, or the funding/investment for further development of their winning prototypes (Khayyat & Bannister, 2017). At the same time, citizens working as entrepreneurs are motivated to develop new services by engaging with OGD to earn money (Smith & Sandberg, 2018).

3.3.4. Social factors

Two primary factors associated with social factors were found: social influence/approval and broadening social networks. Built on UTAUT (Venkatesh et al., 2003), *social influence* is defined as an individual's perception that significant others think that she or he should engage with OGD. Previous research provided empirical evidence on the effect of social influence on the acceptance and usage of OGD in different contexts (Saxena & Janssen, 2017; Weerakkody, Kapoor, et al., 2017; Zuiderwijk, Janssen, et al., 2015). For example, supervisors and peers influence social science researchers to engage with OGD to create scientific articles (Zuiderwijk, Janssen, et al., 2015). Another example concerns the inclusion of OGD usage for public policy by lecturers in a master's degree program (Gascó-Hernández, Martin, Reggi, Pyo, & Luna-Reyes, 2018).

Broadening social networks was also an important factor that drives citizens to engage with OGD in different contexts. For example, some citizens were motivated to participate in open data hackathons because they wanted to make contacts and meet new people (Gama, 2017; Hellberg & Hedström, 2015). Another example concerns some citizens who participated in different open government political platforms to meet like-minded others (Hutter et al., 2011). At the same time, some older citizens participated in an OGD-based digital public service creation because they wanted to socialize with the others (Jarke, 2019).

3.3.5. Technical factors

Drivers related to technical factors were categorized in the following groups: perceived ease of use, system quality, data quality, and service quality. The perceived ease of use term was adopted from TAM (Davis, 1989) and other technology-related factors were grouped based on IS success model's quality factors (DeLone & McLean, 2003). *Perceived ease of use* refers to the degree of ease associated with engaging with OGD. The significant challenge of typical engagement with OGD is the ability to utilize data and find patterns and trends in a massive bulk of data (Zurada & Karwowski, 2011). Complex and sophisticated data require all types of OGD users' capabilities and knowledge levels, such as mastering particular statistical techniques (Janssen et al., 2012). Previous research provided empirical evidence that perceived ease of use influences citizens' intention to engage with OGD in different contexts (Fitriani et al., 2019; Jurisch et al., 2015; Weerakkody, Kapoor, et al., 2017; Wirtz et al., 2018, 2019).

DeLone and McLean's (2003) three quality factors were adopted to group technical factors with similar characteristics: system quality, data quality, and service quality. OGD and the tools for accessing, viewing, analyzing, or visualizing OGD are typically published on a website or portal. These systems, technologies, platforms, and functionalities are critical ingredients to OGD engagement (Charalabidis et al., 2014). Citizens can use such a system to search for data sets, download and visualize them (Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012), or even develop applications on top of the OGD (Janssen et al., 2012). *System quality* can be defined as a group of drivers related to the citizens' evaluation of the OGD system's performance from a user-oriented perspective. The system quality drivers can be categorized into three criteria: having the required functionalities/features, user-friendliness, and availability. The first criterion refers to having the required functionalities/features for data processing (Charalabidis et al., 2014), user-level feedback (Osagie et al., 2017; Talukder et al., 2019), knowledge sharing (Smith et al., 2016), and interaction with other users (Osagie et al., 2017). The second criterion concerns the system's user-friendliness (Smith et al., 2016; Talukder et al., 2019) related to its simplicity, consistency, intuitiveness (Osagie et al., 2017). The third criterion relates to the availability of the OGD system (Talukder et al., 2019), which is often evaluated against its response time (Charalabidis et al., 2014).

Data quality can be defined as a group of drivers related to the citizens' evaluation of the OGD quality from a user-oriented perspective, i.e., suitable for use by data consumers (Wang & Strong, 1996, p. 6). Consequently, data quality is personal because its attributes are positively associated with data users' preferences (Wang et al., 2018). The researcher prefers using the data quality term to DeLone and McLean's (2003) information quality because OGD is typically published as raw data that needs to be converted to information. Data quality is generally associated with technical characteristics of data, while information quality concerns non-technical issues (Madnick, Wang, Lee, & Zhu, 2009). However, data quality can be regarded as representing both technical and non-technical characteristics of data. The following criteria related to data quality were identified. The OGD should be *relevant* (Talukder et al., 2019; Toots et al., 2017), *complete* (Talukder et al., 2019), and *reliable* (Talukder et al., 2019). At the same time, the OGD should be published *timely* (Talukder et al., 2019).

Parasuraman, Zeithaml, and Berry's (1985) definition of *service quality* was adopted into the context of this study as a group of drivers related to the citizens' evaluation of the expected OGD services based on the citizens' perceived service performance. Since the mid-1980s, organizations applying IS generally provided support or services for their end-users to enable them to use the IS productively and the information it produces effectively (DeLone & McLean, 2003). In the OGD context, services can be provided in the form of IT-mediated tools such as online guides for OGD users (Janssen et al., 2012) or features such as data quality rating and user comments (Zuiderwijk et al., 2012). Open data hackathons typically provide services taking the form of civil servants offering support or help to participants to use and analyze OGD (Purwanto et al., 2018b). Factors related to service quality were grouped into the following two categories based on the SLR results. First, the documentation or designated persons that assist OGD users (Osagie et al., 2017; Smith et al., 2016; Talukder et al., 2019; Wang et al., 2018) and examples and success stories of OGD use (Zuiderwijk, Susha, et al., 2015) should be available. Second, the ease of influencing the OGD provision (Smith et al., 2016), including the proper follow-up of citizen feedback (Wijnhoven et al., 2015).

3.3.6. Political factors

Three political factors that influence citizens' intention to engage with OGD were found: trust, improvement expectancy, and political interest. *Trust* can be defined as "the confidence a person has in his or her favorable expectations of what other people will do, based, in many cases, on previous interactions" (Gefen, 2000, p. 726). The creation and increase of citizens' trust are the anticipated benefits of opening up government data (Janssen et al., 2012, p. 261). Citizens' trust in government and technology positively influences their trust in the OGD website, which affects their intention to continually use the OGD website in the Indonesian context (Fitriani et al., 2019). At the same time, some citizens engaged with OGD because they have interests in relevant political issues. Citizens are expecting their government to stimulate the creation of public good (Cranefield et al., 2014), increase transparency (Cranefield et al., 2014; Khayyat & Bannister, 2017; Wirtz et al., 2019), and strengthen anti-corruption (Wang et al., 2019). Citizens are also expecting improvements over government performance, such as general efficiencies in government operation (Cranefield et al., 2014) or the current political situation (Hutter et al., 2011).

3.4. Factors inhibiting individual citizens from engaging with OGD

This section reports the inhibitors of citizen engagement with OGD, i.e., the factors that inhibit individual citizens from engaging with OGD. Similar to the previous section, most of the reviewed papers did not precisely analyze and report the inhibitors when collecting the existing evidence of inhibiting factors. Subsequently, empirical evidence indicating the *challenges*, *difficulties*, *problems*, *impediments*, and *barriers* that citizens felt and experienced before or during the OGD engagement, were also extracted. Table 3.8 provides a detailed summary of the inhibitors of citizen engagement with OGD.

Table 3.8. The overview of factors inhibiting individual citizens from engaging with OGD.

Category	Inhibitors	Source(s)
Technical factors	Task complexity	Whitmore (2014), Wijnhoven et al. (2015), Zuiderwijk, Janssen, et al. (2015), Dittus et al. (2016), Khayyat and Bannister (2017), Ruijer et al. (2017), Saxena and Janssen (2017), Smith and Sandberg (2018)
	System quality	Zuiderwijk et al. (2012), Cranefield et al. (2014), Martin (2014), dos Santos Brito et al. (2014), de Kool and Bekkers (2016), Ojo et al. (2016), Smith et al. (2016), Zuiderwijk et al. (2016), Beno et al. (2017), Ruijer et al. (2017), Benitez-Paez et al. (2018), Smith and Sandberg (2018), Wang et al. (2018)
	Data quality	Zuiderwijk et al. (2012), Cranefield et al. (2014), Martin (2014), dos Santos Brito et al. (2014), Whitmore (2014), Ojo et al. (2016), Smith et al. (2016), Khayyat and Bannister (2017), Osagie et al. (2017), Ruijer et al. (2017), Benitez-Paez et al. (2018), Smith and Sandberg (2018), Crusoe, Simonofski, Clarinval, and Gebka (2019)
	Service quality	Zuiderwijk et al. (2012), Ojo et al. (2016), Hivon and Titah (2017), Smith and Sandberg (2018)
Political factors	Lack of trust	Ruijer et al. (2017)
	Political participation	Wijnhoven et al. (2015)

The categorization reported in the previous section was used to classify various inhibitors found in the reviewed studies into two groups: 1) technical factors and 2) political factors. Moreover, inhibitors derived from intrinsic motivations, extrinsic motivations, economic factors, and social factors were not found in the literature. This finding suggests that although engaging with OGD offers external rewards to citizens and adds value to the economy and society are generally accepted, OGD does not intrinsically demotivate citizens.

3.4.1. Technical factors

Most of the inhibitors related to technical factors include task complexity and data quality problems. At the same time, other technical inhibitors such as system quality problems and service quality problems were found. *Complexity* in handling data typically inhibits citizens from engaging with OGD (Dittus et al., 2016; Khayyat & Bannister, 2017; Ruijer et al., 2017; Saxena & Janssen, 2017; Smith & Sandberg, 2018; Zuiderwijk, Janssen, et al., 2015). Task complexity substantially prevents those who do not have the required skills and knowledge to use OGD (Janssen et al., 2012). Sometimes, OGD is too complicated to handle (Whitmore, 2014; Wijnhoven et al., 2015) and may cause burnout (Dittus et al., 2016).

Scholars have identified that *data quality* issues inhibited citizens from engaging with OGD. This phenomenon was first reported by Zuiderwijk et al. (2012) in 2012. Nevertheless, researchers still find these problems in today's OGD engagement initiatives (e.g., Crusoe et al., 2019). Scholars have revealed that low data quality is a recurring inhibitor found in much open data research (Beno et al., 2017; Martin, 2014; Ojo et al., 2016). Inhibitors related to data quality were grouped into eight issues: timeliness, interoperability, data format, completeness, accessibility, metadata, availability, and accuracy.

Inhibitors related to *timeliness* are associated with the uncertainty of data publication sustainability (Cranefield et al., 2014; Khayyat & Bannister, 2017; Martin, 2014; Smith et al., 2016). Some data sets were not published regularly (Martin, 2014), or removed from the portal (Ojo et al., 2016; Zuiderwijk et al., 2012), or lagging (Khayyat & Bannister, 2017; Ruijer et al., 2017). At the same time, some already published data sets were not updated (Benitez-Paez et al., 2018; Ojo et al., 2016; Zuiderwijk et al., 2012). *Interoperability* related inhibitors relate to the inability of OGD infrastructures to interoperate (Zuiderwijk et al., 2012) and combining OGD (Crusoe et al., 2019) because of lack of standards (Beno et al., 2017; dos Santos Brito et al., 2014; Khayyat & Bannister, 2017; Ojo et al., 2016). The *data format* can also inhibit citizens from engaging with OGD when it is not user-friendly (Ojo et al., 2016) or machine-readable (Beno et al., 2017; Ruijer et al., 2017). OGD is too complex to handle (Whitmore, 2014) because it involves a layered request-based structure and format (Smith et al., 2016). OGD is also sometimes *incomplete* (Beno et al., 2017; Osagie et al., 2017; Ruijer et al., 2017; Whitmore, 2014; Zuiderwijk et al., 2012) with limited relevant data sets available (Smith & Sandberg, 2018; Whitmore, 2014). Relevant data are also occasionally not published (Crusoe et al., 2019; Smith & Sandberg, 2018; Whitmore, 2014; Zuiderwijk et al., 2012); even if the data are made available, they are not free to use (Zuiderwijk et al., 2012). Citizens can explore OGD when *relevant metadata* is provided (Zuiderwijk et al., 2012).

Limited, inconsistent, and incomplete metadata inhibit citizens from engaging with OGD (Beno et al., 2017; Martin, 2014). Lastly, OGD that *lacks accuracy* also inhibits citizens from engaging with it (Osagie et al., 2017; Whitmore, 2014; Zuiderwijk et al., 2012).

The following inhibitors were related to the characteristics of the system that provides access to OGD and, therefore, they were identified as the *system quality* problem group: lack of documentation, lack of functionality, lack of user-friendliness, lack of integration, responsiveness problem. *Lack of proper documentation* typically relates to the availability of information/data about the data set (metadata) (Beno et al., 2017; Ruijer et al., 2017) and about the APIs (e.g., how to access, examples of API call outputs) (Beno et al., 2017; Smith & Sandberg, 2018). Even if relevant documentation is provided on the OGD system, they are typically fragmented (Smith & Sandberg, 2018), and examples of intelligent use of OGD are unavailable (Ojo et al., 2016). *Lack of functionality* concerns the unavailability of features needed by OGD users to search and give feedback (Zuiderwijk et al., 2012) and for viewing, mapping, and visualizing multiple data (Ojo et al., 2016). *Lack of user-friendliness* typically relates to the system's interface that is not user-friendly (Martin, 2014; Ojo et al., 2016; Zuiderwijk et al., 2016). The *lack of integration* refers to data platform silos that force OGD users to access different portals to download and use relevant data sets (Benitez-Paez et al., 2018; dos Santos Brito et al., 2014). Lastly, *responsiveness problems* inhibit citizens from engaging with OGD when the system is unavailable or slow at responding to the user's request (Smith et al., 2016).

All too often, open data researchers overlook the roles of services or support provided for assisting OGD users in stimulating citizens' intention to engage with OGD (Purwanto, Zuiderwijk, & Janssen, 2020c). The following inhibitors related to *service quality* were found: the non-existence of support, communication difficulty, and feedback mechanism. *Lack of support* is one of the main inhibitors of OGD engagement (Ojo et al., 2016; Smith & Sandberg, 2018), and generally, governmental organizations rarely provide help or training for the use of OGD (Zuiderwijk et al., 2012). Sometimes, OGD users experience difficulty communicating and interacting with the civil servant representing the data owner (Hivon & Titah, 2017) or obtaining insights into the OGD providers' activities (Smith & Sandberg, 2018). OGD providers also typically did not provide feedback mechanisms, and in turn, OGD users are experiencing difficulties when trying to request follow up from the providers (Smith & Sandberg, 2018).

3.4.2. Political factors

Inhibitors related to political factors concern lack of trust and less political interests. Citizens demand data that can be trusted to make better decisions; yet, there is a lack of trust in using the data (Ruijer et al., 2017). Typically, citizens disappointed with their government performance become less interested in politics (Wijnhoven et al., 2015). As a result, they become less likely to engage with OGD.

3.5. The theoretical framework of OGD citizen engagement

The second purpose of the literature review is to develop a theoretical framework of OGD citizen engagement based on the SLR results. The theoretical framework can be used to analyze factors that drive an individual citizen to engage with (drivers) or inhibit a citizen from engaging with OGD (inhibitors). Figure 3.3 illustrates the theoretical framework. The driving factors grouped in intrinsic, extrinsic, economic, social, technical, and political factors have a positive relationship with citizen engagement with OGD. In contrast, the inhibiting factors classified in technical and political factors negatively affect citizen engagement with OGD. The fact that some factors are identified both as drivers and inhibitors simultaneously shows that they are the opposite side of the same coin. For instance, the perceived ease of OGD use contradicts the complexity in dealing with OGD. When OGD is easy to handle, the citizen's perception of ease of OGD uses increases while task complexity decreases. On the contrary, when the OGD becomes too complicated to handle, the perception of ease of OGD use decreases and task complexity increases. Subsequently, citizens may be inhibited from engaging with OGD.

In the framework, the researcher postulates that the citizen's profiles moderate the relationships between the driving and inhibiting factors and OGD engagement. The researcher also posits that the citizen's profiles described with her or his age, awareness of OGD provision, education level, gender, occupation, and resources can influence the strength of the relationships. For instance, a citizen having an occupation as a senior programmer may not be affected by the complicated OGD. The programmer may have the necessary capabilities to handle such OGD, and as a result, although task complexity increases, her or his perception of the ease of use may not decrease. Another example concerns a data science student who is highly motivated and has the required capabilities but is inhibited from engaging with OGD because she or he lacks resources such as money to participate in an OGD engagement.

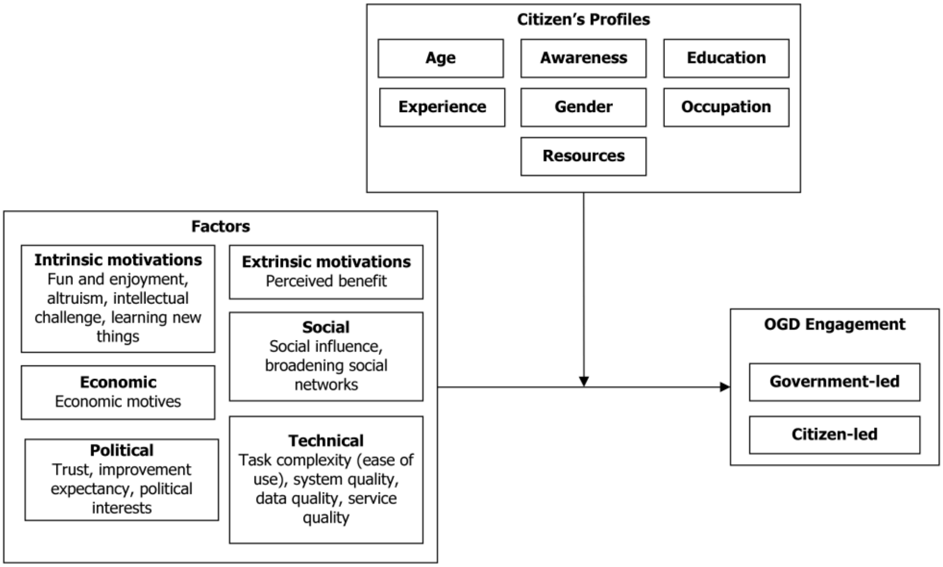


Figure 3.3. The theoretical framework of OGD citizen engagement

The researcher also postulates that the type of OGD engagement with which citizens engage is determined by different driving and inhibiting factors and the citizen's profiles. Different types of OGD citizen engagement may attract citizens with different profiles. For instance, citizens who participate in citizen-led OGD engagement initiatives are more likely to be activists. Older adult activists may engage in OGD-based service design to improve their neighborhood (Jarke, 2019), while humanitarian activists create maps vital in a crisis (Dittus et al., 2016). Such citizen profiles aiming to advance transparency and accountability agendas are also present in different fields, including the election counting processes (Purwanto et al., 2018a), war spending (Whitmore, 2014), and electoral candidacy (dos Santos Brito et al., 2014). In contrast, hobbyists and employees are more likely to participate in government-led OGD engagement than activists. For instance, open transportation data hackathons were commonly attended by citizens having a hacking hobby (hackers) (Gama, 2017; Juell-Skielse et al., 2014; Kuk & Davies, 2011). Another example concerns employees of different companies engaged in different fields related to agricultural businesses who attended Dutch open agriculture hackathons (Purwanto et al., 2019).

3.6. Conclusion and answer to the first research question

The second chapter of this dissertation has put forward the research phases to investigate citizen engagement with OGD. This chapter presented the outcome of the first research phase and provided insights into understanding OGD citizen engagement and factors that influence it. Furthermore, this chapter answers the first research question (RQ1): *what drivers and inhibitors for citizen engagement with OGD have been identified in previous research?* Based on the literature synthesis, there are two types of factors associated with citizen engagement: 1) factors that directly influence OGD engagement and 2) citizen's profiles that moderate the relationships between factors and OGD engagement. Finally, a conceptual model was proposed to describe the relationships between the influencing factors, citizen profiles, and OGD engagement (see Figure 3.3).

From the SLR, it can be concluded that various factors influence either government-led or citizen-led OGD engagement. The degree of influence is contingent upon the citizen's profiles. Based on the synthesis, the influencing factors can be categorized into six types: 1) intrinsic motivations, 2) extrinsic motivations, 3) economic factors, 4) social factors, 5) technical factors, and 6) political factors. Furthermore, the citizen's profiles are described by the citizen's age, awareness of OGD provision, education level, experience, gender, occupation, and resources in OGD engagement. In the following chapter, the theoretical framework proposed in this chapter will guide a multiple case study and analyze the reasons behind the citizen engagement initiatives in real-life cases.

4. Multiple-Case Study of OGD Citizen Engagement

The previous chapter presented the theoretical framework describing the factors influencing citizen engagement with OGD derived from the literature. Based on the synthesized model, intrinsic, extrinsic, economic, social, technical, and political influence government-led and citizen-led OGD engagement. Furthermore, these factors' strength depends on citizens' profiles: age, awareness of OGD provision, education level, experience, gender, occupation, and resources in OGD engagement. The synthesis calls for further investigation to explore whether factors synthesized in the conceptual model exist in practice and to identify whether certain factors are missing in the literature because prior research on this field is lacking.

This chapter addresses the second research question: *why do citizens engage with OGD in existing government-led and citizen-led OGD initiatives?* A multiple case study approach explores the synthesized factors and identifies emerging factors missing from the conceptual model. This chapter first describes the case study approach used in the research, including the case design and selection criteria. It then describes two cases representing different, complementary types of OGD engagement: government-led (i.e., the Hack de Valse Start hackathons) and citizen-led (i.e., the Kawal Pemilu) and presents analysis within each case. Finally, the results of a cross-case analysis of the two cases are discussed, and the studied factors are concluded. The findings of this chapter serve as the outcome of the second research phase and provide a foundation for developing a survey in the following research phase. We have published parts of this chapter in Purwanto, Janssen, and Zuiderwijk (2017), Purwanto et al. (2018a), Purwanto et al. (2018b), Purwanto et al. (2019), and Purwanto, Zuiderwijk, and Janssen (2020b).

4.1. Case study approach

In this section, the case study approach used in this research stage is described. This research follows Yin's (2014) guideline, similar to those of Eisenhardt's (1989). The following steps from the approach were applied: 1) case study design, 2) preparing for data collection, 3) collecting evidence, 4) analyzing evidence, and 5) reporting case studies. Section 4.1 and Section 4.2 describe the first four steps, while Section 4.3 and 4.4 present the last step. First, Section 4.1.1 describes the case study design. The design defines the multiple cases with embedded units of analysis and explores ways to strengthen the validity of the approach. Second, Section 4.1.2 describes the case selection criteria. Third, Section 4.1.3 describes the case study protocol, which defines the approaches used to collect and analyze data, the instrument used for evidence collection, and the approaches used for data analysis.

Finally, Section 4.2 presents the case study setup, including the overview of the selected cases.

We can define a case study as “an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident” (Yin, 2014, p. 16). A case study is a method in which the researcher does not explicitly control variables and utilizes qualitative data collection and analysis (Cavaye, 1996, p. 229). Therefore, case research is valuable when: 1) the phenomenon understudy is broad and complex, 2) an in-depth investigation is needed, and 3) the phenomenon which researchers cannot study outside the context in which it occurs (Benbasat, Goldstein, & Mead, 1987; Dubé & Paré, 2003). The case study method has been valuable and well-established in the IS field (Walsham, 1995). It is the most preferred method to study e-government-related topics (Danziger & Andersen, 2002). OGD is among the top domains of research interest in e-government (Scholl, 2013), and scholars have conjectured it as the second generation of e-government (Charalabidis, Loukis, Alexopoulos, & Lachana, 2019). More importantly, researchers can conduct case study research with a pragmatist paradigm (Teddlie & Tashakkori, 2009). Researchers can also use case research for developing theory (Eisenhardt, 1989). Therefore, the case study approach is appropriate for this research phase.

A case study approach is appropriate for this research phase of exploration and identification of factors. First, this research phase aims to investigate the reasons and motivations behind citizen engagement with OGD. As a result, this phase focuses only on the driving and inhibiting factors of OGD engagement. Second, the boundaries between OGD engagement and its context are not clearly evident. The literature demonstrates that engaging with OGD involves complex and dynamic activities and different situations that may affect the activities for which actors in an OGD ecosystem, including citizens as OGD users, depend upon each other. Thus, it is crucial to investigate OGD engagement in the context in which it takes place. Third, citizens' experiences who engage with OGD are important because they contribute to creating artifacts such as apps or infographics powered by OGD. Since this research phase aims to explore factors identified in the theoretical framework (see Chapter 3) in real OGD engagement cases and identify factors missing from the framework, citizens should be involved. Researchers can understand citizens’ feelings or experiences through words, which can only be investigated using a qualitative inquiry such as interviews, a common technique used in case study research. Fourth, research and theory in OGD citizen engagement are at an early and developing stage. Researchers can use case study research to build theories by either deductively or inductively analyzing qualitative data (Eisenhardt, 1989). Therefore, in this research, the case study is deemed a valuable approach that can be used to generate theories about OGD citizen engagement.

4.1.1. Multiple case study design

Researchers typically categorize case studies as exploratory, descriptive, explanatory, or improving (Runeson & Höst, 2009). A researcher conducting an exploratory case study can discover what is happening, seek new insights, and generate ideas and hypotheses for further research. A descriptive case study portrays a situation or phenomenon, while explanatory case studies seek an explanation of a situation or a problem. An improving case study tries to improve a particular aspect of the studied phenomenon. The case study in this research phase combines a descriptive and exploratory study, i.e., describing how citizens engage with OGD and exploring a priori (driving and inhibiting) factors that influence OGD citizen engagement. Although the research question asked in this research phase, i.e., RQ2, is concerned with explaining why citizens engage with OGD, the purpose of this case study is not to generalize causal relationships between the factors and engagement to the population. Instead, following Flyvbjerg (2006), the case study approach in this research can be used to test propositions, namely, the theoretical framework presented in Section 3.5. The results can only be generalized to the context of the research.

A case study must be well-designed to survive four standard tests of empirical social research to achieve high-quality research: 1) construct validity, 2) external validity, 3) reliability, and 4) internal validity (Yin, 2014). *Construct validity* concerns “the degree to which a data collection procedure (e.g., instrument, interview procedure, observational strategy) truly captures the intended construct that is being studied” (Teddlie & Tashakkori, 2009, p. 286). A case study researcher is encouraged to identify precise operational measures (e.g., interview questions) that represent the theories she or he is investigating (Bryman, 2012; Yin, 2014) and what is investigated under the research questions (Runeson & Höst, 2009). According to Yin (2014), utilizing multiple sources of evidence and creating a chain of evidence can improve construct validity. The case study's construct validity can be enhanced by developing a data collection instrument (see Section 4.1.4) based on the framework proposed in the literature review (see Chapter 3). Section 4.1.4 outlines multiple sources of evidence used to collect data and the chain of evidence established during the case study research.

External validity can be defined as “the degree to which findings can be generalized across social settings” (Bryman, 2012, p. 390). In MMR, pragmatist researchers define the qualitative study's external validity as *inference transferability* (Teddlie & Tashakkori, 2009, p. 32). External validity is related to case studies' incapability to provide generalizable inferences (Dubé & Paré, 2003), representing a typical problem for qualitative researchers due to small samples (LeCompte & Goetz, 1982). Lee (1989) responded constructively to this problem by suggesting that theories' generalizability should be examined and confirmed in various situations. Therefore, researchers should investigate additional cases and collect sufficient data for the case study replication. This research improves the case study’s external validity by investigating multiple cases. The results of one case study are examined in the context of the other case. Researchers can also enhance external validity by supplying adequate information about the case study design to enable replication (Lee, 1989). The provision of sufficient study design information enables the conduct of additional case studies for evaluating the results of this research and examining to which degree they can be generalized. The multiple-case study design selected in this research involves two embedded cases (see Section 4.1.3). The cases focus on specific types of OGD and engagement in a particular context. The influencing factors of OGD engagement that will be drawn out are critical in the context of these cases.

Reliability concerns establishing that the case study can be replicated or that the operations of the study (e.g., data collection procedures) can be reiterated with similar findings (LeCompte & Goetz, 1982; Yin, 2014), regardless of the researchers (Runeson & Höst, 2009). Hypothetically, another researcher employing the same methods or conducting the same study will obtain similar results (LeCompte & Goetz, 1982; Runeson & Höst, 2009). Although replicating a particular case study's observations may be impracticable, it is viable to evaluate the same theory in various sets of initial conditions (Lee, 1989). Therefore, other researchers can still replicate the findings from prior case studies. Optimizing the case reliability requires using a case study protocol, which the researcher followed when carrying out the study (Yin, 2014). In this research, the case study methodology is defined, and then the case study protocol to enhance the study's reliability is developed. The study's reliability can also be optimized by generating a case study database (Yin, 2014). The database ensures that the entire records of all stages of the research process (e.g., data collection instruments, selection of research participants, interview transcripts, data analysis decisions) are kept in an accessible manner (Bryman, 2012). These records are made available on the 4TU Research Data platform to enable future researcher audits the merits of this research. The list of links to the records are described in Appendix B.

Internal validity refers to demonstrating a causal relationship by which particular conditions are deemed to result in other conditions, which can be differentiated from spurious associations (Yin, 2014). Thus, internal validity shows the extent to which researchers' observations and the theoretical concepts they develop have a good match (Bryman, 2012). However, internal validity applies only to explanatory or causal research instead of descriptive or exploratory studies (Yin, 2014). Therefore, internal validity is of less concern for this case study research, and it is not discussed in this research phase because the investigation does not concern inferring causal relationships. Although the research question "*why do citizens engage with OGD*" indicates an explanatory study, the subject of this research is digitally literate citizens, humans with varying degrees of different conditions and situations. These variances prevent the researcher from drawing generalization such as explaining that particular citizens' conditions will cause them to engage with OGD.

4.1.2. Case study selection

A fundamental pre-requisite for conducting case studies is the selection of cases (Eisenhardt, 1989). *A case* is related to the *unit of analysis* under investigation (Yin, 2014) and can be individuals, groups, or an entire organization (Benbasat et al., 1987). A case can also be a wide variety of events or entities beyond a single individual, such as communities, decisions, programs, organizational change, and specific events (Yin, 2014). In this study, *the case* is defined as OGD engagement events. At the same time, *the unit of analysis* is defined as individual citizens who engage with OGD collectively as a group, either initiated by the government (government-led) or by the citizens themselves (citizen-led). See Section 1.1 for discussion about the role of these individuals in an OGD ecosystem.

Cases are typically selected either because of their substantive significance based on statistical sampling or their theoretical relevance based on theoretical sampling (Dubé & Paré, 2003; Eisenhardt, 1989). Selecting cases based on *statistical sampling* requires clear statistical evidence of the variables' distribution within a particular population. This sampling applies to research aiming at testing theories in the field. Cases can also be selected using *theoretical sampling* when they "replicate previous cases or extend an emergent theory, or … fill theoretical categories and provide examples of polar types" (Eisenhardt, 1989, p. 537). The cases' selection criteria are based on theoretical sampling because the study aims to build a theory of OGD citizen engagement. Although researchers have widely studied the socio-technical conditions of OGD utilization (Hossain et al., 2016; Safarov et al., 2017), the theory of OGD citizen engagement seems non-existent. This situation shows the need for theory building, and therefore, theoretical sampling is considered appropriate for selecting cases in this research.

The selection of cases in this research is based on the theoretical framework proposed in Section 3.5, whereby two different types of OGD citizen engagement, i.e., government-led and citizen-led, hypothetically attract different profiles of citizens. The literature indicates that citizens may also have various reasons to engage with OGD when it is government-led compared to citizen-led. For example, government-led OGD engagement such as a hackathon commonly offers rewards to its participants in prize money or the possibility of being funded after the competition (e.g., Ayele, Juell-Skielse, Hjalmarsson, & Johannesson, 2015). At the same time, a citizen-led engagement usually does not offer incentives to citizens participating in it (Dittus et al., 2016). Likely, the reasons citizens engage with OGD in a government-led initiative would differ from that of citizen-led. Therefore, the cases are selected based on the theoretical sampling of these types of engagement.

Although a single-case study such as a critical, extreme, common, revelatory, or longitudinal case can be justified, researchers prefer multiple-case designs over single-case designs (Yin, 2014). A multiple-case study design is more robust than a single-case study design because its evidence is frequently deemed more convincing (Herriott & Firestone, 1983). Furthermore, a multiple-case study design enables comparing cases from one or more settings, i.e., cross-case analysis (Yin, 2014). Researchers can use a cross-case analysis to compare similarities and differences in events, activities, and processes of the units of analysis (Khan & VanWynsberghe, 2008). Typically, researchers distinguish a *holistic* case design involving a single-unit analysis, where a researcher investigates the case as a whole, and an *embedded* case where a researcher studies multiple units of analysis within a case (Yin, 2014). Whether a case study research is holistic or embedded depends on the context defined by and research goals formulated by the researcher (Runeson & Höst, 2009).

In a multiple-case study design, cases must be cautiously selected based on a *literal replication* in which similar findings are expected or a *theoretical replication* in which contradictory but anticipatable results are expected (Yin, 2014). Yin (2014) indicates that the number of cases that will meet the replication requirements is six to ten cases. Two to three cases are sufficient for a literal replication purpose and an additional four to six cases for theoretical replication. At the same time, the researchers select more than one case within each setting, and they expect these cases to generate similar results (literal replication). Given that, in previous discussions, literature indicated the difference in citizens' reasons to engage with OGD between government-led and citizen-led, multiple cases included in this research should represent these two different settings. Multiple units of analysis that refer to citizens who engage collectively with OGD are studied (see Figure 4.1) within each case, representing a particular setting (either government- or citizen-led). Therefore, in this research, a multiple embedded case study design is chosen.

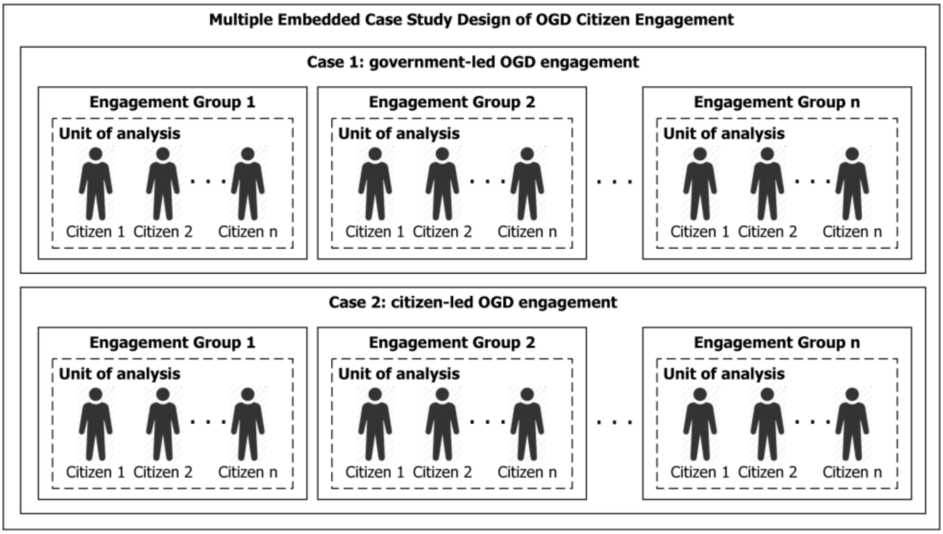


Figure 4.1. The selected multiple embedded case study design for this research.

Benbasat et al. (1987) recommend that case study researchers detail the case selection criteria that explicitly define the cases' characteristics. The selection criteria in this research, derived from the scope of this research discussed in Chapter 1 (see Section 1.3), are defined as follows:

1) *The cases involve OGD that has already been made available to the public*. This research investigates OGD as defined in Chapter 1 and citizen engagement with OGD described in Chapter 3. More specifically, in this research, OGD refers to data made publicly available on the internet by governmental organizations. At the same time, citizen engagement with OGD relates to converting OGD into valuable artifacts. These definitions mean that the selected cases ought to involve OGD that has been made available to the public to allow an OGD engagement to happen. Although this research's primary focus is on citizen engagement with OGD and not

on the provision of open data by governments, the availability of OGD remains an essential condition for enabling OGD engagement.

2) *The cases involve either government-led OGD engagement or citizen-led OGD engagement*. Investigating citizen engagement with OGD is central to this research. As proposed in the theoretical framework (see Section 3.5), different types of engagement (i.e., government-led and citizen-led) are influenced by various factors over which citizens' profiles influence the factors' strength. Therefore, the selected cases involve these two types of engagement: citizen engagement led and organized by governmental organizations (government-led), and citizen engagement led and independently organized by citizens.
3) *The cases involve actual outcomes of either government-led OGD engagement or citizen-led OGD engagement*. This research focuses on the citizens' activities of converting OGD into important and relevant artifacts as defined in Section 3.2.1. These artifacts can take the form of applications or visualization (e.g., infographics) built on top of OGD. Therefore, the selected cases should involve OGD engagement that generates verifiable actual outcomes.
4) *The cases involve OGD that is provided to the public by Dutch and Indonesian governmental organizations*. In the context of OGD provision, compared to the Netherlands, Indonesia can be regarded as a latecomer in releasing government data (Nugroho, Zuiderwijk, Janssen, & Jong, 2015). Drivers and barriers of open data-driven co-creation appear to be indistinguishable in countries with advanced open data ecosystems and those that are latecomers in adopting open data (Toots et al., 2017). However, these assumptions are thus far theoretical, and cultural influence is not the focus of this research. Involving OGD from different countries may lead to cultural influences on OGD engagement. According to Hofstede Insights (2018b) and Hofstede Insights (2018a), the Netherlands and Indonesia are culturally different in three dimensions: *power distance*, *individualism*, and *masculinity*. The Netherlands' scores are low in power distance (38) and masculinity (14) and high in individualism (80) (Hofstede Insights, 2018b). In contrast, Indonesia scores low in individualism (14) and high in power distance (78) and masculinity (46) (Hofstede Insights, 2018a).

   Hofstede, Hofstede, and Minkov (2010) describe *power distance* as "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally" (p. 61). The Netherlands' score on this dimension characterizes the Dutch culture as an independent culture where hierarchy is only convenient (Hofstede Insights, 2018b). In this culture, citizens have equal rights, superiors are accessible, leaders have a more coaching role, and management facilitates and empowers (Hofstede Insights, 2018b). In contrast, the Indonesian culture is dependent on hierarchy, no equal rights between power holders and non-power holders, inaccessible superiors, directive leaders, and controlling and delegating management (Hofstede Insights, 2018a).

Individualism concerns “societies in which the ties between individuals are loose: everyone is expected to look after him- or herself and his or her immediate family” (Hofstede et al., 2010, p. 92). On the other hand, *collectivism* refers to “societies in which people from birth onward are integrated into strong, cohesive in-groups, which throughout people’s lifetime continue to protect them in exchange for unquestioning loyalty” (Hofstede et al., 2010, p. 92). The Netherlands’ score on this dimension shows Dutch as an individualist society while Indonesian is a collectivist society. Individualist society expects individuals to take care of themselves or their immediate families. On the contrary, society and the in-groups to which individuals belong expect them to conform to their collectivist society.

Masculinity concerns people’s motivation, whether having a desire to be the best (masculine) or attachment to what you do (feminine) (Hofstede et al., 2010). In this dimension, the Indonesian culture appears to be more masculine than the Dutch. Keeping the balance between life and work and ensuring no one is left behind are essential in feminine countries such as the Netherlands, while status and noticeable symbols of success are important in Indonesia. Cases from these countries were intentionally selected because the study sought different contradictory settings to represent theoretical replication better, as discussed in the previous section. For practical reasons, since the researcher carrying out this research was based in the Netherlands, the government-led OGD engagement case study was conducted in the same country. At the same time, Indonesia was chosen for the citizen-led OGD engagement case study because the researcher had access to detailed data. Although the researcher did not speak Dutch, English is widespread and omnipresent in Dutch society (Van Essen, 1997), and he had access to Dutch-speaking social science researchers who can help translate and interpret Dutch-specific terms. Therefore, English was used in the government-led OGD engagement cases. Regarding the citizen-led OGD engagement case, the researcher spoke the Indonesian language, which provided easy access to information from the case.

5) *The cases involve different types of OGD from the education inspection and election domains*. Cases involving different OGD will allow the

researcher to investigate whether the results from one case can be applied or contradicted in another context. Suppose researchers can draw common inferences from different cases in different contexts; the research findings' external generalizability can improve (Yin, 2014). The first case centered on open education inspection data, while the second one focused on open election results. The studied cases focus on education inspection data and election data because the governmental organizations in the Netherlands and Indonesia have already published them. Besides, education inspection and election data are essential and relevant for identifying and solving different societal problems. The education inspection data, combined with other data such as school data, social and economic data of parents who have school-age children, and poverty data, provide meaningful information. Citizens can obtain insights into the number of drop-out students from specific socio-economic backgrounds, the distribution of government subsidies for schools from different performance levels, and the school advice given to students from diverse backgrounds. Insights into these issues can help governments generate interventions in education policy. At the same time, citizens can use open election data to obtain insights into the overview of votes gathered by competing candidates, the number of suspicious results, and the discrepancy between the actual results witnessed by voters and the official published results. These insights can help the government prevent corruption related to election processes and correct erroneous election results.

6) *The cases involve citizen engagement with the open education data and open election data published by governmental organizations of the Netherlands and Indonesia to solve societal problems*. This study focuses on citizens' collaborative activities to convert OGD into valuable artifacts such as visualizations and applications, which are important and relevant to them and society, to identify societal problems and contribute to them. Therefore, OGD engagement cases that focus on achieving commercial purposes and business and government engagement with OGD are beyond this research's scope.
7) *The cases involve digitally literate or technologically skilled citizens engaged with OGD in groups without holding government positions*. This research focuses on the citizens, as OGD direct users, who do not have official positions within the government. This criterion is formulated to position this research within the open data literature dominated by companies and governmental organizations' studies of OGD use. From the beginning of open data research (e.g., Davies, 2010; Gurstein, 2011;

Hivon & Titah, 2017), it is known that having skills related to data analysis is an essential requirement for citizens to use OGD.

8) *The cases involve people and organizations with access to the data and are willing to cooperate and share information needed to carry out this research*. Therefore, this criterion is primarily related to the willingness of the data users (i.e., citizens or a group of citizens), data providers (i.e., governmental organizations that provide OGD), and the officials or employees of data providers involved in the cases to provide the data needed for this study.

Based on the selection criteria mentioned above, two case studies were selected. The first case study is the Hack de Valse Start. In this case, citizens engage with the Dutch open education inspection data. The second case study is the Kawal Pemilu, a citizen engagement with the Indonesian open election data. The following section describes the applied case study protocol in this research concerning the approaches used to collect and analyze data, the instrument used for evidence collection, and the approaches used for data analysis.

4.1.3. Case study protocol

Following Yin's (2014) guide, this section elaborates on the case study protocol by describing the data collection procedures, instruments, and guide for reporting the case study. A case study protocol is crucial to enhance the reliability of case study research (Yin, 2014) and direct the researcher in collecting data (Eisenhardt, 1989; Yin, 2014). A case study protocol is a continuously updated document modified when the case study plans are changed (Runeson & Höst, 2009).

Data collection procedures

Data collection procedures refer to the detailed explanation of the researchers' procedures to collect data in each case study (Pervan & Maimbo, 2005). For example, procedures for protecting human subjects and identifying sources of data (Yin, 2014). These procedures should be applied to ensure the uniformity of the data collection process and facilitate case comparison analysis (Pervan & Maimbo, 2005; Yin, 2014).

In this research, the primary data sources were citizens' opinions who engaged with open education data in the first case study and open election data in the second case study. In addition, data from various sources, such as website pages, social media posts, news articles, government regulations, were also examined. Table 4.2 describes the sources of data analyzed during the case study research. During this research phase, the researcher stored the collected data in a local computer as the primary database and cloud storage as the backup database.

Table 4.2. Overview of the data sources of the case studies.

Information sources	Case 1: Government-led engagement (Hack de Valse Start)	Case 2: Citizen-led engagement (Kawal Pemilu)
Semi-structured interviews	9 (involving six hackathon participants, one hackathon organizer, and two open education data providers)	15 (involving 12 citizens and three open election data providers that were also election organizers)
Documents	Two government regulations, one government report, two presentation files created by team members	Eight government regulations, one government report, two documents created by contributors
Web pages	22 from the hackathon website, three from open education data providers	20 from the Kawal Pemilu website, three from open election data providers, ten news articles
Data sets	13 education and socio-economic status-related data sets	17 presidential election-related data sets
Participant-observations	Participation in one of the teams competing in the Hack de Valse Start hackathon	Participation via Facebook membership-only group during the election result digitization

Semi-structured interviews

Semi-structured interviews with key persons who had substantial knowledge of the open education and election data provision and citizen engagement with the data were carried out. Interviews were carried out mainly with citizens engaged with open education data in the Hack de Valse Start hackathon and independently engaged with open election data in the Kawal Pemilu initiative. Interviewees from the Hack de Valse Start case were selected because they were accessible to the researcher, and their respective teams had won the challenge competition.

The Kawal Pemilu volunteers interviewed, some of whom were the initiative's principal founders, were selected because they had engaged with open election data and had considerable knowledge of the case. Nevertheless, locating some of the interviewee candidates was difficult because the Kawal Pemilu volunteers did not reveal their identities to the public for safety reasons (Graft et al., 2016). Recruiting respondents through random surveys would not yield an accurate and relevant sample because of the anonymity of the volunteers. Therefore, a *snowball sampling technique* was employed to locate some of the interviewee candidates. This technique has been extensively applied in qualitative social science research (Biernacki & Waldorf, 1981). Snowball sampling refers to building a sample of respondents who are recommended or suggested by those already interviewed. It enables access to formerly hidden populations (Atkinson & Flint, 2001), such as the Kawal Pemilu volunteers. The researcher asked one of the Kawal Pemilu coordinators to identify other volunteers who had knowledge about the initiative and created an initial list of respondent candidates. Based on the list, the researcher conducted initial interviews and gathered more potential candidates for the study. In total, twenty interviewees were identified. Similar information was gradually collated when interviewing these volunteers, and no more new information was discovered at a certain point. This point appeared to indicate that data saturation had been reached. Therefore, the researcher stopped finding out more interviewees.

Although the case study did not focus on the publication of OGD, interviews also took place with individuals who worked for organizations providing open education data and open election data. The researcher interviewed these individuals because they were knowledgeable regarding the usability of the OGD provided by their organizations and regularly communicated with OGD users. In the Hack de Valse Start case, two persons from the governmental organization providing the open education inspection data (i.e., the Dutch Inspectorate of Education) and one person providing support during the hackathon were interviewed. One of the two interviewees from the Dutch Inspectorate of Education was the inspector general, while another interviewee was an information manager responsible for opening up and evaluating the use of inspection data. In the Kawal Pemilu case, the researcher interviewed three persons from the governmental organization providing open election data and organizing the election at different administrative levels. Two interviewees were commissioners of the municipal-level Election Commission. One was a commissioner of the provincial-level Election Commission.

The interviews for the two cases were conducted from October 2017 until May 2018 through different methods. The researcher recorded all interview sessions as agreed by the respondents. In both cases, interview sessions took an average of sixty minutes to complete. In the Hack de Valse Start case, four interviews were conducted through face-to-face meetings, while five interviews took place online, using Skype and WhatsApp. The researcher held online interviews to reduce travel time; the Hack de Valse Start case respondents lived in different cities in the Netherlands. Table 4.3 shows the nationalities and residences of the respondents. Typically, an interview session needed sixty minutes to complete. In the Kawal Pemilu case, nearly all of the interviews took place online using WhatsApp and Google Hangout, and the researcher conducted only one interview through a face-to-face meeting.

Table 4.3. Overview of the data sources of the case studies.

Respondent ID	Role(s)	Nationality	Resided in	Interview Freq.
Case 1 – Hack de Valse Start				
C1-01	Hackathon participant	Dutch	The Netherlands	2
C1-02	Hackathon participant	Romanian	The Netherlands	2
C1-03	Hackathon participant	Russian	The Netherlands	1
C1-04	Hackathon participant	Dutch	The Netherlands	1
C1-05	Hackathon participant	Dutch	The Netherlands	1
C1-06	Hackathon participant	Dutch	The Netherlands	2
C1-07	Hackathon organizer	Dutch	The Netherlands	1
C1-08	Open data provider	Dutch	The Netherlands	1
C1-09	Open data provider	Dutch	The Netherlands	1
Case 2 – Kawal Pemilu				
C2-01	Verifier	Indonesian	Indonesia	1
C2-02	Contributor	Indonesian	Indonesia	1
C2-03	Contributor	Indonesian	Indonesia	1
C2-04	Contributor	Indonesian	Indonesia	1
C2-05	Contributor	Indonesian	Australia	2
C2-06	Contributor	Indonesian	Indonesia	1
C2-07	Contributor	Indonesian	Indonesia	2
C2-08	Developer	Indonesian	The Netherlands	1
C2-09	Contributor	Indonesian	Indonesia	1
C2-10	Developer	Indonesian	Singapore	1
C2-11	Contributor	Indonesian	The Netherlands	1
C2-12	Contributor	Indonesian	Singapore	1
C2-13	Open data provider and election organizer	Indonesian	Indonesia	1
C2-14	Open data provider and election organizer	Indonesian	Indonesia	1
C2-15	Open data provider and election organizer	Indonesian	Indonesia	1

In the Hack de Valse Start case, the researcher developed three interview instruments in English containing questions to ask the respondents based on their roles in the case. At the same time, the researcher used the Indonesian language to develop interview instruments for the Kawal Pemilu case. These instruments are explained in detail in the following subsection and were made available online through the 4TU Center for Research Data (see Table B.1 of Appendix B). The researcher transcribed all recorded interviews were in their original language of interview. The interviews were transcribed in English in the Hack de Valse Start case, while the Kawal Pemilu case was Indonesian.

Documents

Documents examined in the case study research include government regulations and reports and documents created by citizens involved in the two case studies. Government regulations analyzed in the Hack de Valse Start case include the Dutch Education Supervision Act and the Organization and Mandate Decision of the Ministry of Education. Eight government regulations studied in the Kawal Pemilu case concerned the Indonesian General Election Acts and the General Election Commission's (KPU) particular rules on the presidential election organization. The researcher also examined two government reports in both cases, namely the Background Information about the Dutch Inspectorate of Education and the KPU's 2014 Presidential Election Results Report. In addition, the researcher analyzed two files created by team members containing the results of education data analysis used for giving a final presentation in the hackathon from the Hack de Valse Start case. Two documents created by citizens who engage with election data, the design of the Kawal Pemilu data entry crowdsourcing system and the usage report of kawalpemilu.org, were also studied.

Web pages

In both cases, web pages identified and cataloged were related to the open education and election data providers, hackathon organizers, citizen groups engaging with election data (i.e., kawalpemilu.org), and other parties, including the press covering news about the election. Web pages analyzed from the OGD provider's websites can describe the providers' brief history, functions carried out and published data sets. The researcher also examined web pages from the hackathon organizer of the Hack de Valse Start case, particularly those related to the pre hackathon information and the competition's conclusion. In the Kawal Pemilu case, the researcher studied the citizen group's web pages to understand how the election data digitization took place, the digitization results at all governmental administrative levels, and how other citizens can report anomalous election results. Also, selected official posts made on social media platforms that asked citizens' participation to contribute to the digitization of election results and published comments from public figures about the performance of Kawal Pemilu are analyzed. The researcher also examined news articles related to the impacts of presidential election campaigns, the works of Kawal Pemilu, and the comments of various election stakeholders on the election process and the Kawal Pemilu.

Data sets

In the Hack de Valse Start case, the researcher identified and cataloged data sets related to education and socio-economic status. Hackathon team members needed these data sets to analyze the problems challenged in the education data hackathon and visualize them to understand the relationships between many variables and provide suggestions about the possible key variables of the problems. The studied data sets include, for example, the number of students per primary school based on their age, the average score of National Final Test per primary school, and the number of schools (teacher) advice given to students grouped by type of higher education level per primary school. In the Kawal Pemilu case, the Election Commission's identified data sets concerned the scanned election results in JPEG format, geographical data in JSON format, and election results from different administrative levels in HTML. The researcher downloaded and examined different types of sampled data sets in their original formats, including Comma Separated Value (CSV), Excel, JSON, and HTML, in the local storage and backed them up in the cloud storage.

Participant-observations

The researcher observed the activities carried out by citizens who engaged with the open education data and open election data using different methods. In both cases, the researcher performed participant observations: joining one of the Hack de Valse Start hackathon teams and contributing to the Kawal Pemilu volunteer groups. Gaining actual access to these cases provides an exceptional opportunity to understand the OGD engagement from an insider's perspective because post-factum comprehension is non-trivial (Yin, 2014).

In the Hack de Valse Start case, the researcher did not explain the purpose of participation in the hackathon until the event concluded. The researcher used a *concealed observation* approach; the team members in which the researcher joined were not informed about the researcher's goal to minimize respondent bias. When respondents are aware that they are being observed, they may behave differently or unnaturally (Sekaran & Bougie, 2016). During the hackathon, the researcher took notes on the activities carried out by team members, their behaviors, events they were involved in, and nonverbal cues shown in their interpersonal relationships. However, the concealed observation approach suffers from ethical drawbacks because it may infringe the privacy and informed consent principles (Sekaran & Bougie, 2016). Afterward, the researcher informed group members that he participated in the hackathon for research purposes, particularly to identify potential cases and respondents. The researcher asked all team members for their agreement to be interviewed. Before the interview, the researcher explained to the team members that they could still withdraw from the study whenever they wanted despite the agreement. The researcher also explained that the team members' identities would be anonymized in the following research publications based on data collected during the hackathon and interviews. In this way, the study's likeliness of ethical violation due to concealed participant observation approach was reduced.

Kawal Pemilu ran entirely on the Facebook platform. Therefore, the researcher carried out observation online through the platform. Although the researcher could not physically observe activities carried out by contributors and verifiers, he could observe their interactions online in the Kawal Pemilu's closed Facebook group posts. The researcher examined these posts and comments and downloaded them as PDF files.

Data collection questions

Data collection questions concern the specific questions that the case study researcher must keep in mind in collecting data and the potential sources of evidence for addressing each question. The researcher developed questions in English for the Hack de Valse Start case and questions in Indonesian for the Kawal Pemilu case. The researcher made the questions available online on the data platform of the 4TU Center for Research Data. Table B.1 of Appendix B depicts the list of these questions. Questions designed for a particular role are not relevant for other roles. Therefore, the researcher developed three groups of questions for each case based on the roles of actors involved in the OGD engagement. In the Hack de Valse Start case, the researcher identified three roles: 1) hackathon participants, 2) hackathon organizers, and 3) OGD providers, whereas in the Kawal Pemilu case, 1) developers, 2) volunteers including contributors and verifiers, and 3) OGD providers. The researcher developed the interview questions to ask citizens who engage with OGD built on the conceptual model described in Section 3.5. The researcher also pilot tested the questions involving six academic researchers specializing in open data and information sharing fields: four for English and two for Indonesian interviews.

Questions for the open education data and open election data providers in the Hack de Valse Start and Kawal Pemilu cases were formulated to understand how they published election data. For example, what business processes they defined in the open data chain, what challenges they experienced when providing OGD, how barriers were tackled, and what institutional plan they would execute for future OGD provision. The questions asked to the Hack de Valse Start hackathon organizers aimed to understand how they supported the publication of OGD in events such as the hackathon. For example, what challenges they encountered when supporting OGD and how they dealt with the challenges. The questions asked to the Hack de Valse Start hackathon participants and Kawal Pemilu volunteers aimed to understand the themes related to the participants' profiles and the factors that influence them to engage with OGD. For instance, what their roles were, how they carried out their activities, what motivated them, and their challenges during the engagement. These themes were derived from the conceptual model proposed in the first research phase, and the interview questions were subsequently developed based on topics related to the subthemes. The researcher made the interview questions available online on the 4TU Research Data repository, and the reader can find a complete list of links to the questions in Appendix B.

The following three topics derived from the conceptual model (see Chapter 3) were excluded in the questions: voluntariness, altruism (intrinsic motivation), task complexity, and system quality (technical factors). The rationale for this is as follows.

- *Altruism* is among the determinants of citizens' participation in online service reporting (Schmidthuber, Hilgers, Gegenhuber, & Etzelstorfer, 2017) and citizen-sourcing and hackathons in the public sector (de Deus Ferreira & Farias, 2018). In this research, altruism can be defined as one's desire to enhance others' welfare (Hars & Ou, 2002) (see Chapter 3 for more detailed discussions about altruism). Based on this definition, altruism can be described as a pro-social behavior (Wijnhoven et al., 2015). Therefore, the question concerning altruism was reformulated to resemble "benefitting society," categorized in social factors.
- *Task complexity* was excluded from the interview topics because it is irrelevant in the studied cases. Open data literature commonly assumes that engaging individually with OGD involves high-complexity tasks (Janssen et al., 2012; Whitmore, 2014). However, research has shown that, for high-complexity tasks, working in groups is more effective than working individually (Kirschner, Paas, & Kirschner, 2011). In groups, individuals having different skills can complement each other. Since the studied cases involved engaging with OGD in groups, it was assumed that task complexity becomes irrelevant.
- *System quality* is related to the system's performance providing access to OGD, such as having the required functionalities/features, being user-friendly (e.g., simple, consistent, intuitive), and being available when accessed. System quality was excluded from the interview topics because, in the studied cases, the OGD system is not necessarily accessed and used by citizens. In the Hack de Valse Start case, the hackathon organizer provided most of the data sets needed to solve the challenges. The

hackathon participants did not have to directly access the data from each of the OGD provider's open data portals. In the Kawal Pemilu case, the volunteers did not need to access the OGD system either because the governmental organizations provided the data sets through APIs. Moreover, only citizens with a developer role accessed the APIs, and the contributors were not required to do so. Therefore, system quality was deemed as an irrelevant topic.

Finally, some other questions were not included, including awareness and resources, because the researcher could explore them using more general questions. Table 4.4 shows the final list of topics discussed in the case studies.

Table 4.4. An overview of the topics covered in the interview.

Theme	Subtheme	Topics
General	Awareness, experience	Perception and expectation regarding OGD
		Perception about citizen engagement with OGD, including the current level of engagement and ways to stimulate or increase engagement
		Perception of the respondent's engagement with OGD
	Resources	Involvement in the engagement including the respondent's roles, motivations, and activities carried out during the engagement
		Challenges experienced during the engagement and ways to address the challenges
Profiles	Age	Age of the respondent
	Gender	Gender of the respondent
	Education	Highest educational level achieved by the respondent
	Capabilities	Respondent's occupation
Extrinsic motivations	Relative advantage	The extent to which relative advantage influences the respondents to engage with OGD
	Performance expectancy	The extent to which performance expectancy influences the respondents to engage with OGD
	Career	The extent to which career concerns influence the respondents to engage with OGD
Intrinsic motivations	Fun and enjoyment	The extent to which fun and enjoyment influence the respondents to engage with OGD
	Intellectual challenge	The extent to which intellectual challenge influences the respondents to engage with OGD
	Learning new things	The extent to which learning new things influences the respondents to engage with OGD
Social	Social influence	The extent to which social relationship influences the respondents to engage with OGD
	Benefitting society	The extent to which benefitting society influences the respondents to engage with OGD

Theme	Subtheme	Topics
Social	Broadening social networks	The extent to which broadening social networks influences the respondents to engage with OGD
Technical	Data quality	The extent to which data quality (i.e., accuracy, completeness, format, currency, understandability, interoperability) influences the respondents to engage with OGD
	Service quality	The extent to which service quality (i.e., reliability, assurance, responsiveness) influences the respondents to engage with OGD
Economic	Monetary rewards	The extent to which monetary rewards influence the respondents to engage with OGD
	Financial gain	The extent to which financial gain influences the respondents to engage with OGD
Political	Trust	The extent to which trust influences the respondents to engage with OGD
	Need for change	The extent to which the need for change influences the respondents to engage with OGD
	Political participation	The extent to which political participation influences the respondents to engage with OGD
Other factors	Factors beyond those asked earlier	The extent to which other factors influence the respondents to engage with OGD

Guide for the case study report

The theoretical framework developed in Chapter 3 (see Section 3.5) guided the case study report. The framework describes the factors driving an individual citizen to engage with or inhibit citizens from engaging with OGD. The framework also shows the type of OGD engagement investigated in the case study (i.e., government-led, citizen-led). It indicates the factors classified in the following groups that need to be examined: extrinsic motivations, intrinsic motivations, social factors, technical factors, economic factors, and political factors. It also points out citizens' profiles related to their gender, age, educational background, resources, capabilities, awareness, competency, and experience. For each case, the researcher first investigated the clusters of factors and citizens' profiles from Chapter 3. The researcher then compared the findings from both cases and drew conclusions based on the comparative analysis. This research employs Computer-Assisted Qualitative Data Analysis (CAQDA) software, namely Atlas.ti, for analyzing the qualitative data collected from the multiple-case study method through interviews and documents. Atlas.ti enables researchers to develop codes, identify them in the collected data, visualize the codes and their categories, and analyze the code patterns from the qualitative data (Friese, 2012).

4.2. Case study setup

4.2.1. Overview of the cases

This subsection briefly reports the overview of the two selected case studies and the multiple units of analysis of the two cases. Table 4.5 provides a brief overview of the selected cases.

Table 4.5. Overview of the selected cases.

Characteristics	Case 1: Government-led OGD engagement (Hack de Valse Start)	Case 2: Citizen-led OGD engagement (Kawal Pemilu)
Domain	Education	Election
Involved data	Education inspection data (OI), school data (DUO), demographic data (CBS), poverty data (GA)	Election results (KPUD, KPUP, KPU)
Government level	Municipal, national	Municipal, provincial, national
Type of engagement	Government-led (a hackathon)	Citizen-led (a bottom-up initiative)
Country	The Netherlands	Indonesia
Involved actors	Hackathon organizers, OGD provider organizations, hackathon participants	OGD provider organizations, election organizations, citizens participated in election result digitization

Case 1: Government-led OGD engagement (Hack de Valse Start)

The first case study is concerned with the government-led OGD engagement in an open education data hackathon, namely, Hack de Valse Start. The Hack de Valse Start case study can be described as a hybrid approach combining holistic and embedded designs. Figure 4.2 illustrates the boundaries of this case study. The figure displays a holistic view that involves beyond the limits of individual organizations by encompassing the organization providing open education data and the organization managing the open education data hackathon, groups of citizens engaging in the hackathon, and the outcomes of the open education data engagement. It also shows an embedded view of cases illustrated with the open education data engagement groups comprising more than one citizen. The citizens participating in the hackathon in these groups are the units of analysis in the studied case.

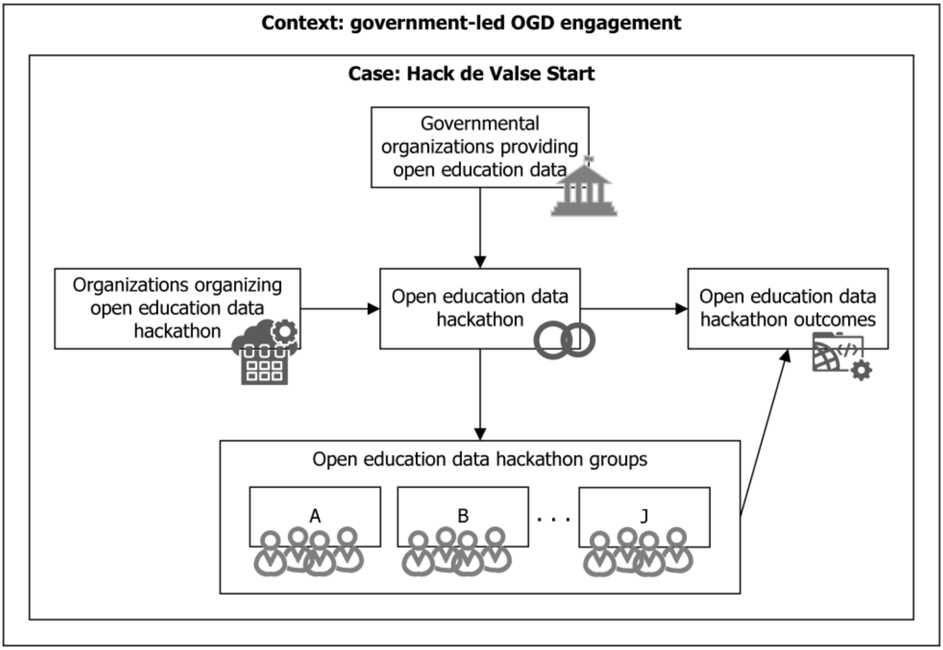


Figure 4.2. The units of analysis of the Hack de Valse Start case.

The citizen engagement with open education data manifests in the Hack de Valse Start hackathon organized by the Dutch Ministry of Education, Culture, and Science, Municipality of Amsterdam, Open State Foundation, and Young Creators (Open State Foundation, 2018a). The hackathon aimed to look at inequality in education in a new way. A recent report published by the Dutch Inspectorate of Education (Dutch: *Onderwijs Inspectie*; OI) showed that inequality of opportunity in education has increased as gaps between students with lower and higher educated parents grew. As a result, many children with low-educated parents did not receive the education they could afford, and their talent remains underutilized. The Hack de Valse Start hackathon's primary aim was to use open data to understand how municipalities and school boards spot and deal with education inequality.

Different groups of actors were involved in the hackathon. First, governmental organizations that provided the open education data. These organizations include the Ministry of Education, Culture, and Science (Dutch: *Ministerie van Onderwijs, Cultuur, en Wetenschap*; OCW) and Municipality of Amsterdam (Dutch: *Gemeente Amsterdam*; GA). OCW provided education inspection data primarily through OI, which is part of the Secretary-General of OCW. OI, founded in 1801, was one of the oldest state inspectorates of education (Onderwijs Inspectie, 2015). The Education Supervision Act (Dutch: *de Wet op* *het onderwijstoezicht*; WOT) mandated OI's legal foundation, which became effective in September 2002 (Staatscourant, 2005). Based on WOT, OI has four primary roles:

1) Stimulating schools and educational institutions to preserve and increase their education quality,
2) Assessing the quality of the individual educational institutions and the Dutch education system and its developments,
3) Communicating with all target groups and stakeholders in an accessible way and
4) Reporting the inspection results to the public.

OI's school reports published only to the schools and OCW have been made publicly available since 1998 (Onderwijs Inspectie, 2015). These reports, including an actual list of weak and very weak schools, are available on OI's website[1] for public consultation (de Kool & Bekkers, 2015). In these reports, the results of the overall assessment performed by OI on the strength and weakness of primary schools that are deemed at risk are labeled as "normal," "weak," or "very weak" (van Twist, van der Steen, Kleiboer, Scherpenisse, & Theisens, 2013). These *weak* or *very weak* schools will then receive more intense follow-up inspections. Open education inspection data provided by the Onderwijs Inspectie (2019) for the Hack de Valse Start hackathon, including:

1) Standards and quality indicators used for the OI assessments,
2) Supervisory and arrangements and final judgments made by the OI inspectors,
3) Sampling files created and used by OI to oversee the education system,
4) Profiles (the names and locations) of excellent schools that have special qualities and unique excellence,
5) Profiles of very weak schools that achieve insufficient final educational results which also show inadequate quality in crucial parts of the educational learning process,
6) A national database of school suspensions and removals per sector and type of education, the duration of suspensions and the frequency in which specific reasons are stated, and
7) School weighting of primary education that OI uses for assessing the learning outcomes of schools.

In the open education data hackathon, *Gemeente Amsterdam* (Municipality of Amsterdam) provided various data related to Amsterdam City, including 1) key figures for secondary school students in Amsterdam, 2) socio-economic status

[1] www.onderwijsinspectie.nl

per 4-digit postcode area of Amsterdam, 3) more than 500 variables classified according to a various theme, and 4) poverty and poverty schemes in Amsterdam's boroughs. *Gemeente Amsterdam* (Municipality of Amsterdam) has started to publish open data in 2011[2] and has been active in sponsoring many hackathon events that utilized various themes of Amsterdam's open data.

The second group of actors involved in the hackathon concerns organizations that organized the hackathon's events, i.e., the Open State Foundation and Young Creators. The Open State Foundation primarily organized the hackathon's main events, such as providing open data related to the hackathon's themes, technical and non-technical support, and organizing presentations from governments and participants. The Open State Foundation developed a portal that compiled and provided a catalog of additional open data relevant to the hackathon and provided access and preview to the data sets (Open State Foundation, 2018a). The portal encompasses public data opened by the Education Executive Agency (Dutch: *Dienst Uitvoering Onderwijs*; DUO), Central Bureau of Statistics (Dutch: *Centraal Bureau voor de Statistiek*; CBS), and Amsterdam City. Examples of data provided by the Education Executive Agency include data on the number of students per primary school based on their age[3], the average score of Central Final Test (Dutch: *Centrale Eindtoets*; Cito-toets) per primary school[4], and the number of schools (teachers) advice given to student grouped by type of higher education level per primary school[5]. Data provided by Statistics Netherlands include a link to the vital statistical figures of districts (Dutch: *wijken*) and neighborhoods (Dutch: *buurten*) grouped by the municipality (Dutch: *gemeente*) and year[6], and a compilation of social economy situation (e.g., number of family members, household income) of students grouped by neighborhoods[7]. The Open State Foundation helps governmental organizations disclose public information as open data and stimulates its reuse (Open State Foundation, 2018b). It works together with governments, civil society organizations, journalists, media organizations, and knowledge institutes and aims to strengthen the people's right to information to make more informed choices and exert their influence. In the hackathon, Young Creators organized the side events that involved sport and gaming activities for participants. Young Creators is a youth community that focuses on connecting its members to the Dutch start-up and business communities (Young Creators, 2018). It aims to be where young talents find

[2] https://data.amsterdam.nl
[3] https://data.openstate.eu/dataset/hack-de-valse-start/resource/77fe311e-bf63-46e2-98cc-775fd8e5007e
[4] https://data.openstate.eu/dataset/hack-de-valse-start/resource/43659254-97c3-44a1-9d8f-4d3ab64d6871
[5] https://data.openstate.eu/dataset/hack-de-valse-start/resource/4fee7993-7279-4034-a20c-3abd278bd0e9
[6] https://data.openstate.eu/dataset/hack-de-valse-start/resource/d57974a5-bd42-48fa-bdee-9bdacf324e56
[7] https://data.openstate.eu/dataset/hack-de-valse-start/resource/5fd19e5d-ed23-4b64-a9c5-b58ce9503fb2

inspiration and are encouraged to realize their ambitions (Open State Foundation, 2018b).

The third group of actors involved in the hackathon concerns citizens participating in the event. Ten groups participated in the hackathon; they were labeled A through J. The names of the groups are intentionally anonymized to prevent the tracing of their members' identity back to the Hack de Valse Start website and to protect their members' privacy as regulated in the European Union law on data protection and privacy, i.e., the General Data Protection Regulation (GDPR). Two groups, A and B, were selected and investigated in this research because they won the hackathon. Also, the researcher had made himself known to the team members during the hackathon. These citizens were also accessible to the research inquiry.

The hackathon started at eight in the morning and ended at eight in the evening (twelve hours) in the Calvijn College in Amsterdam on the 3rd March 2018. Anyone interested in participating in the hackathon could register for free. Some of the participants formed teams during the event. The hackathon was opened with a presentation by the Dutch Education Inspector regarding problems in education inequality, followed by a session in which participants were allowed to form teams. In this session, the participants introduced themselves and explained their background and specialization related to the hackathon. In addition, some participants actively recruited other participants to join their groups. During this session, ten groups were formed, including the two teams understudy—the first team (team A) comprised five citizens, and the second team (team B) was composed of two citizens. Only team A and B members were selected as the primary respondents of this case study because both teams met the case selection criteria, including the third criterion, i.e., generating outcomes based on the open education inspection data (see Section 4.1.2). Other teams could not deliver the required outcomes on time during the hackathon. These teams were also chosen because the researcher had convenient access to them. Subsection 4.1.4 describes a brief overview of these citizens. Subsection 4.2.1 provides a more detailed overview of the analysis of the cases. The hacking competition started soon after the hackathon participants formed teams. Teams had to decide which challenge(s) they attempted to address, out of the following challenges:

1) Primary Education Challenge. This challenge concerns the time spent in primary school during which students have little influence on their lives, and others usually make significant decisions during the period. Teams that addressed this challenge had to look at parents, teachers, and institutes to identify possible answers to questions that impact education

inequality, such as how often parents move, given school advice, and the primary school's profile.

2) Secondary Education Challenge. This challenge is concerned with secondary school students. Students gain more control over their lives, although external factors can influence this ability. In this challenge, teams had to obtain insights into how students' choices, such as taking a part-time job, impact students' opportunities and how inequality played a role.
3) Data Visualization Challenge. This challenge relates to creating a visualization of available data in education and inequality of opportunity to provide information and insight needed by people who influence policy.

Although teams were free to use any open data available on the internet to address the challenges, they typically relied on the hackathon organizer's open data portal. Given this portal, teams could download data relevant to the challenges, learn about the contents of the downloaded data sets, combine or integrate different data sets, and create visualization and analysis. Team members had to develop interpersonal relationships to build consensus and understanding toward each other's roles, enabling the further allocation of tasks and activities to achieve the hackathon's ultimate goals.

Team A was composed of five members, including the researcher, and led by a journalist of a crowdfunded news website who specifically wrote about education. The other three members of the team were from different backgrounds. One of them was a data scientist who worked for a travel aggregator company. Another one was a municipality employee who worked as a researcher in the cultural education unit. Lastly, one participant was a workshop organizer promoting open data use through "maker" arts, a contemporary social movement representing a technology-based *do-it-yourself* intersecting with hacker culture. Team B comprises two members: a Ph.D. candidate from the macroeconomy field and a User Experience (UX) engineer working for a bank. Not only did they vary in their capabilities, but the members also came from different countries.

In Team A, the journalist contributed primarily to brainstorming and exploring ideas about education inequality topics that the team would address. The journalist also indicated various data sets provided by the hackathon organizers that the team could use to support the ideas. Since the data sets were mainly in Dutch, the municipality researcher helped translate the attributes of the data sets into English and interpret the content of the data sets. She also raised discussions about several issues about discriminative advice to students from particular backgrounds. The researcher searched and downloaded relevant data sets suggested by the journalist and municipality employee and distributed them to the data scientist. The data scientist developed visualizations involving different variables based on the data sets that illustrate the variables' relationships. He also utilized regression analysis techniques to obtain insights into these relationships. The workshop organizer provided a data-sharing platform to distribute data sets, notes, codes, and presentation files and prepared and designed the team's final presentation. In Team B, the tasks were allocated based on the members' capabilities. The Ph.D. candidate developed visualizations that depict the relationships between different variables, and the UX engineer designed a mockup application. The hackathon organizer assigned one of its personnel to provide technical support for data use-related problems during the hacking. Both teams have used the opportunity to ask for additional data sets needed to draw meaningful inferences about the challenges.

At the end of the hackathon, the organizer held a *pitch* session to present the hacking outcomes and convince the juries that the team's insights or ideas could address education inequality challenges. A jury consultation followed the session to decide three teams that won the competition. Team A presented two histograms depicting the school advice (grouped based on the higher education level advised by teachers) and the Cito-toets[8] scores' distribution, respectively. The team also presented three data plots highlighting the correlations between school advice: 1) the Cito-toets scores, 2) parent backgrounds and 3) household income. These outcomes were assumed to provide insights on what the team might attribute to the inequality of education. Team B presented three histograms depicting the average Cito-toets scores, the percentage of the given school advice, and the school's distance, categorized by groups of the school's ideology. The team also presented a mockup application for primary school students to give direct feedback to their teacher about what they find difficult about the homework. The outcomes of these teams have been made available in the 4TU Research Data platform. Based on the jury consultation, Team B won the first prize worth €3,000, while Team A won the second prize, €1,500.

Case 2: Citizen-led OGD engagement (Kawal Pemilu)

The second case study concerns the citizen-led OGD engagement with open election data, subsequently labeled as Kawal Pemilu. Figure 4.3 presents the boundaries of this case study. Like the Hack de Valse Start case, this case can also be characterized as a hybrid approach that combines holistic and

[8] In April and May, Dutch primary school students in 8th grade are required to take the Central Final Test (Dutch: *Centrale Eindtoets*) that is commonly referred to as the Cito-toets, aiming to evaluate what has been learned in math and language in eight years. A student's Cito-toets score and school's advice are used as indicators whether a choice of secondary education will be successful. See https://www.centraleeindtoetspo.nl/ and https://www.cito.nl/onderwijs/primair-onderwijs/centrale-eindtoets for more information about Cito-toets.

embedded design. The involvement of actors beyond the limits of an individual organization indicates the holistic design of the case. The design encompasses the organization providing open election data and managing the election, groups of citizens engaging with the open election data, and the open election data engagement outcomes. It also shows an embedded view of cases within the open election data engagement group involving more than one citizen. The citizens participating in the Kawal Pemilu initiative are the units of analysis in the studied case.

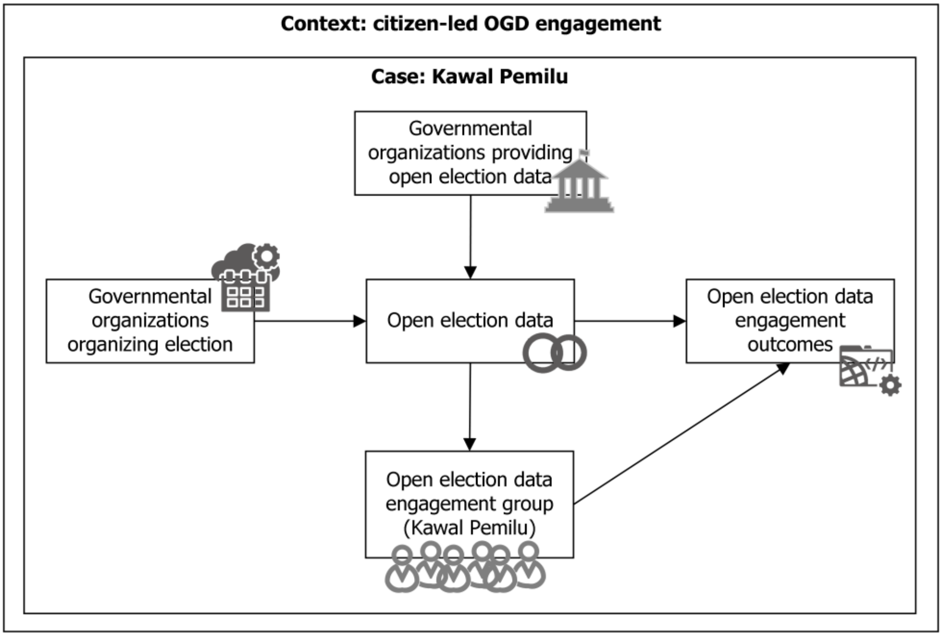


Figure 4.3. The units of analysis of the Kawal Pemilu case.

The case of citizen engagement with open election data manifests in Kawal Pemilu (In English: *guard the election*; KP), one of many citizen-led initiatives that sprang up in the Indonesian presidential election due to the opening of election results (Brajawidagda & Chatfield, 2014). Since the 2014 presidential election, the General Election Commission (In Indonesia: *Komisi Pemilihan Umum*; KPU) has scanned these results and made the scanned results available on the internet. As a result, there was a surge of citizen-led voting applications built on open election data. The reasons why Kawal Pemilu was selected in this research were twofold. First, international practitioners from different fields have mentioned Kawal Pemilu in their practical reports as an important example of participatory election (ACE Electoral Knowledge Network, 2014) and open data impact (Young & Verhulst, 2016). Second, the researcher had been involved in the initiative, and therefore, the KP members were accessible to the research inquiry. KP aimed to digitize the election results, make them readily available for the public to see the real count of the aggregated votes, and provide a way to report back to KPU any anomalous results for verification and correction (Kawal Pemilu, 2014).

Different groups of actors were involved in open election data engagement. First, governmental organizations that provided open election data and organized the Indonesian election, i.e., KPU. The history of KPU can be traced back to the early years of Indonesian independence when its first president, Soekarno, formed the Forming Board of Central National Committee Structure (Republik Indonesia, 1946). After undergoing different organizations in the Reformation Order, the government transformed KPU into a new election organizer following the Soeharto regime's fall. The government established KPU as a professional, capable, and accountable election organizer with integrity that aims to increase the quality of the general election organization that can ensure the exercise of the people’s political rights (Republik Indonesia, 2011).

KPU organizes three types of elections every five years at the national level: The House of Representatives, the members of the Senate, and the presidential and vice-presidential elections. The election organization is hierarchical from the national level to the voting booth (Republik Indonesia, 2011). At the provincial and municipal levels, the Provincial Election Committee (Indonesian: *Komisi Pemilihan Umum Provinsi*; KPUP) and the Regency Election Committee (Indonesian: *Komisi Pemilihan Umum Daerah*; KPUD) organized the elections, respectively. At the district and village levels, organizations of the election include the District Election Committee (Indonesian: *Panitia Pemilihan Kecamatan*; PPK) and the Village Voting Committee (Indonesian: *Panitia Pemungutan Suara*; PPS), respectively. Finally, at the voting booth, the Voting Booth Organizer Group (Indonesian: *Kelompok Penyelenggara Pemungutan Suara*; KPPS) organized the voting.

Elections and vote counting in Indonesia are conducted manually, involving complex hierarchical processes throughout vast geographical areas. This complexity increases the election results’ susceptibility to fraud and requires longer to count (Brajawidagda & Chatfield, 2014). Figure 4.4 illustrates the tabulation's hierarchical process and structure based on Komisi Pemilihan Umum (2014b). The counting processes start at the polling station as follows. First, the Voting Booth Organizer Group (KPPS) read the paper ballots depicting the voter’s choice. Then, one KPPS member tallies the result using a form labeled C1 Plano, followed by another KPPS member who writes the results on a form named C1. The C1 form reports each candidate's vote results and is declared valid when the chairman and two KPPS members have signed it. Next, the KPPS members send a copy of the C1 to the Village Voting Committee (PPS) and another one to the Regency Election Committee (KPUD). At the village level, the village voting committee aggregated the vote results written in C1 forms collected from voting stations, which is then written in the D1 forms, followed by the aggregation of vote results reported in D1 forms at the district level (PPK), which manifests in the DA1 forms. Parallel with these activities, the Voting Booth Organizer Group (KPPS), the Village Voting Committee (PPS), the District Election Committee (PPK), and the candidates' witnesses at the village level held a recapitulation meeting to verify results recorded in C1 forms. The meeting's outcomes became inputs in a higher-level meeting between the Voting Booth Organizer Group (KPPS), the Village Voting Committee (PPS), the District Election Committee (PPK), the Regency Election Committee (KPUD), and witnesses at the district level. At the municipal level, the Regency Election Committee (KPUD) aggregated the vote results written in DA1 forms and reported them in DB1 forms. The Provincial Election Committee (KPUP) aggregated DB1 forms and wrote them in the DC1 forms at the provincial level. Again, the meeting participants conducted verification on the forms. Any errors spotted on the C1 forms were corrected, and these findings would lead to re-scanning and re-uploading the error-containing forms. KPU aggregated vote results written in DC1 forms at the national level and reported them in DD1 forms. The witnesses of the candidates observe the overall vote counting processes.

In 2014, KPU made the election results publicly available online and accessible and usable through Application Programming Interface (API) services. KPU also instituted new vote-counting processes at the regency/municipality and provincial levels in its internal memo number 1395/KPU/VII/2014 (Komisi Pemilihan Umum, 2014a) for opening election results data. The method of opening vote results data began at KPUD with collecting copied C1 forms from PPK. KPU required KPUD to scan the C1 forms gathered from polling stations as Joint Photographic Experts Group (JPEG) image files and upload the files to the KPU website through an internal application named SITUNG. KPU also obliged KPUDs to create Excel versions of the DA1 and DB1 forms and upload them. Once the forms are uploaded in their entirety, they will be visible to the KPU presidential and vice-presidential election portal visitors. Also, KPU instructed KPUPs to create Excel versions of the DC1 forms and upload them using a similar mechanism. KPU also published geographical data to locate where C1, DA1, DB1, and DC1 forms were collected. KPU developed API instances as a protocol for accessing JavaScript Object Notation (JSON)-

formatted geographical data[9] , C1 images[10] , and Hypertext Markup Language (HTML) formatted DA1[11], DB1[12], and DC1[13] forms. Voters can also view these images and forms using a website developed by KPU for the presidential and vice-presidential election that utilizes the APIs. However, navigating through the website to display a C1 form is discouraging because visitors have to select hierarchical filters by choosing a province, a regency/municipality, a district, a village, and a polling station.

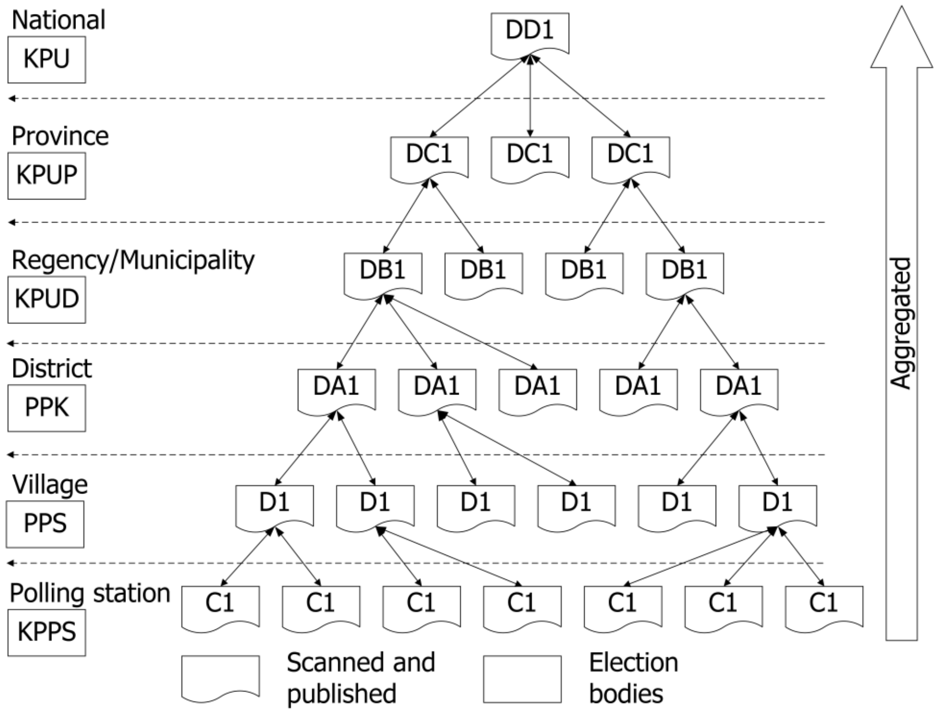


Figure 4.4. The researcher's interpretation of the vote tabulation, based on Komisi Pemilihan Umum (2014b).

The second group of actors involved in the open election data engagement concerns the open election data engagement group, i.e., Kawal Pemilu. There were only two pairs of the president and vice president candidates who competed in the 2014 presidential election. The first pair was Prabowo Subianto and Hatta Rajasa while the second pair was Joko Widodo and Jusuf Kalla (henceforth "Jokowi"). Severe competitions involving fierce debates that

[9] http://tps.kpu.go.id/pilpres2014.php
[10] https://pilpres2014.kpu.go.id/c1.php
[11] https://pilpres2014.kpu.go.id/da1.php
[12] https://pilpres2014.kpu.go.id/db1.php
[13] https://pilpres2014.kpu.go.id/dc1.php

regularly resulted in a confrontation between the camps' supporters on platforms such as Facebook and Twitter have led to highly polarized social conversations (Lim, 2017). Both camps proclaimed their victory over the election right after the voting day, while KPU counted the votes and made no conclusion because they had not finished collecting the entire election results. The claims were made solely based on different survey organizations' quick count results that were factually inconclusive (Lim, 2014). As a result, these claims worsened the situation and confused the public because the conclusion would only be announced when all results had been counted (Graft et al., 2016).

The problems discussed above and the release of election data have resulted in many citizen-led initiatives for counting and digitizing the election results, including Kawal Pemilu (Brajawidagda & Chatfield, 2014). Initially, an Indonesian citizen living in Singapore who worked as a data scientist founded Kawal Pemilu. Later, four other Indonesians in different countries (i.e., California, Sidney, the Netherlands, and Germany) developed the application. Calculating the votes is time-critical, and therefore, the application and its support systems should be ready to use as soon as KPU releases the open election data. As a result, the initiators employed the agile development technique to prevail over this limitation. On 12 July 2014, three days following the voting day, the developers released the application, and the volunteers started using it. However, they confronted a challenge: recruiting citizens to volunteer for digitizing 478,829 C1 forms. The initiator used the Multi-Level Marketing recruitment method to tackle this problem. He enlisted ten 'downline' friends who then enlisted another ten, and so on, and added them all to a private Facebook group. Later on, the founder recruited 700 Indonesian volunteers worldwide three days after the voting date for crowdsourcing the verification and digitization of election results in C1 images.

Nevertheless, until recently, the volunteers have not disclosed their identities to the public. The volunteers used the application's back-end site to key in the vote results to a particular database to enrich the opened data and examine the C1's validity. At the same time, the application's front-end site provides public access to the C1 images and vote results entered at the national level. Using the front-end site, citizens could also drill down the election results into the lower hierarchy levels. The front-end site also enabled citizens to view the current digitization results and report anomalous C1 forms and their digitization results. Eventually, Kawal Pemilu's results only departed 0.01 percent from the KPU's final tally.

Within the Kawal Pemilu initiative, three different types of volunteers that have different roles existed. Activities carried out by these groups were coordinated and facilitated through a Facebook secret group. First, the initiator or developer group comprised one volunteer who founded the initiative and four volunteers who developed the Kawal Pemilu application for digitization and its underlying systems. They were the core team that had the idea behind the initiative and realized the idea. The main idea of these developers is to design a crowdsourcing platform with the following requirements. Firstly, the volunteers should digitize the election results in a closed environment, but the digitization results and errors should be opened to the public. Secondly, the system should have verification mechanisms for tackling data quality problems. Thirdly, the system's interface should be intuitive and easy to operate to accelerate the digitization progress. Fourthly, the system should also employ a rewarding strategy to drive volunteers to digitize election results continually. Whoever completes keying in the data for a particular village can see their names on the leader board. One developer programmed the closed crowdsourcing platform, while two developers programmed the public view-only website[14]. Another developer created scripts for scraping the KPU's website to collect the DA1, DB1, and DC1 data and programmed a mirror version of the Kawal Pemilu's public-facing website. In the end, the developers worked with several volunteers to test the system, identify bugs, and fix them to improve the system's performance.

Second, most of the volunteers constituted the contributor group. They were the team that digitized the election results by keying in the numbers of votes received by each candidate into the Kawal Pemilu application. They also flagged any election results that contained errors (anomalous results). A group coordinator leads each group. Coordinators allocated the digitization tasks to volunteers and kept track of the regions where the volunteers had not entirely digitized the C1 scanned forms. Contributors then selected a particular voting booth under the assigned region and displayed its C1. They keyed in each candidate's election results based on the form. Occasionally, contributors ran into problems arising from blurry, skewed, or vertically flipped C1 scan results that make it difficult to read. When the C1 form is not readable, the contributor can mark it as an error. On the other hand, contributors can intentionally enter the wrong results into the application to increase or decrease a candidate's votes. Since the keyed-in results were automatically aggregated, the public could indirectly supervise the contributors' works by reporting errors found on the Kawal Pemilu website.

[14] https://kawalpemilu.org

The third group of actors involved in the open election data engagement is the verifier group that comprised eight teams in which ten to twelve volunteers constituted each team. They were the teams that verified the number of votes put by the inputter group into the Kawal Pemilu application and corrected any input errors. Two levels of verification were applied to ensure the accuracy of the digitization results. Another group further re-examined a verifier team's results to improve data reliability. They also compiled the unresolved anomalous election results. They reported these results to one of the verifiers, the Kawal Pemilu liaison officer to KPU.

The onus was on the contributors to digitize the entire C1 forms before KPU announced the election winners. In the end, they have succeeded in digitizing slightly more than 464,000 forms, nearly 97% of the total C1 forms. The Kawal Pemilu initiators claimed that their digitization outcomes (i.e., the election results) differed only 0.14% from KPU's official results. The verifiers have discovered moderately more than 10,000 C1 forms containing errors during digitization. KPU's staff have investigated most reports made by Kawal Pemilu's liaison concerning these errors, corrected the mistakes, and re-uploaded the revised C1 forms to the KPU's website. Among the error-containing C1 forms, 4,000 errors were mistakes found by contributors and have been corrected through the verification mechanisms. However, the remaining 6,000 errors have not been resolved by KPU. Finally, KPU officially certified Jokowi as the presidential election victor. KPU's commissioners appreciated the Kawal Pemilu initiative and perceived that the initiative had enhanced KPU's legitimacy and credibility in managing and organizing the elections. Although the victor's rival did not accept the defeat and challenged the election results to the Constitutional Court, the Court dismissed the lawsuit and reaffirmed KPU's decision as valid.

4.3. Within-case analysis

By examining the citizens' responses in the two case studies concerning the seven categories of the theoretical framework proposed in Section 3.5, this study explored the factors influencing citizen engagement with OGD. The following two sections explain the qualitative data based on context types, i.e., government-led OGD engagement and citizen-led OGD engagement. Each section reports and discusses the findings related to citizens' profiles, intrinsic motivations, extrinsic motivations, economic factors, social factors, technical factors, and political factors. The engaged citizens' responses concerning new factors missing from the literature were also examined. The final subsection shows the results of the examination.

4.3.1. Government-led OGD engagement

Background of the citizens

Table 4.6 depicts the team members' profiles, representing citizens understudy, respondents who participated in the hackathon. Four respondents are male, while two are female. Five respondents are in the age range 21 – 30, and only one respondent was above 50 years old. Based on the educational background, all of the respondents are highly educated, of which three of them have a bachelor's degree, and the other three have a master's degree. In addition, the respondents have diversified research-related occupations.

Table 4.6. The profiles of citizens engaged in the open education data hackathon.

Respondent ID	Age	Gender	Education	Occupation	Experience
C1-01	30	Male	Master	Journalist	First hackathon but a regular user of the data of Statistic Netherlands
C1-02	26	Female	Master	Workshop designer	First hackathon but has been involved in some open data workshops
C1-03	26	Male	Bachelor	Data scientist	Second hackathon
C1-04	58	Female	Bachelor	Local government researcher	Eight years of involvement in opening up government data
C1-05	26	Male	Master	Ph.D. candidate	First hackathon
C1-06	24	Male	Bachelor	UX engineer	Not his first hackathon, but he has used open data

For three respondents, the Hack de Valse Start Hackathon was their first hackathon. However, two of them had worked with open data before: one of them regularly used data from Statistics Netherlands, and one was involved in organizing open data workshops. Two respondents had participated in more than one hackathon, and one of them had regularly used open data. One respondent is a local government employee who was involved in opening up education data. Concerning awareness about OGD, all respondents, except the Ph.D. candidate, knew that governmental organizations are providing open data. However, according to the workshop designer and UX engineer, more needed data should be opened to the public. Moreover, usability is one of the main issues for the journalist, the data scientist, and the researcher because interpreting and using the data is complex.

Table 4.7 provides an overview of factors derived from the literature examined in the qualitative data collected in the Hack de Valse Start case. Some factors derived from the conceptual model proposed in Section 3.5 played a role in the case. For instance, fun and enjoyment, getting to know new people, creating benefits for society, data understandability, and the possibility of political change. Other factors derived from the model did not contribute to the case: status and reputation, influence from a close social relationship, and government responsiveness.

Table 4.7. The overview of factors, derived from the literature and examined in the Hack de Valse Start case.

Factors	Play a role in the case?	Evident in citizens interviewed
Intrinsic Motivations		
Fun and enjoyment	Yes	C1-01, C1-02, C1-03, C1-05, C1-04, C1-06
Intellectual challenge	Yes	C1-01, C1-02, C1-03, C1-05, C1-04, C1-06
Status and reputation	No	-
Extrinsic Motivations		
Learning and skills development	No	-
Getting to know new people	Yes	C1-02, C1-03, C1-05, C1-04, C1-06
Future career concerns	Yes	C1-02, C1-05, C1-04, C1-06
Economic Factors		
Financial benefit	Yes	C1-01, C1-06
Social Factors		
Influence from close social relationships	No	-
Influence from wider social relationships	Yes	C1-02, C1-06
Create benefit for society	Yes	C1-01, C1-02, C1-03, C1-05, C1-04, C1-06
Technical Factors		
Data quality: accuracy	Yes	C1-02, C1-05
Data quality: completeness	Yes	C1-02, C1-03, C1-06
Data quality: format	Yes	C1-02, C1-03, C1-04, C1-06
Data quality: currency	Yes	C1-02, C1-04
Data quality: understandability	Yes	C1-01, C1-02, C1-03, C1-05, C1-04, C1-06
Data quality: interoperability	Yes	C1-01, C1-03, C1-05, C1-04, C1-06
Service quality: reliability	Yes	C1-03, C1-05, C1-06, C1-02
Service quality: assurance	Yes	C1-01, C1-02, C1-05, C1-06
Service quality: responsiveness	Yes	C1-01, C1-02
Political Factors		
Trust in Open Government Data	Yes	C1-02, C1-05, C1-06
Government responsiveness	No	-
Interests in politics	Yes	C1-01, C1-02, C1-06
Possibility of political change	Yes	C1-01, C1-02, C1-03, C1-05, C1-06
Involvement in political activities	Yes	C1-01, C1-02, C1-06

Findings related to the intrinsic motivations

The results show that *fun*, *enjoyment*, and *intellectual challenge* influence respondents to engage with open education data in the government-led engagement initiative. In contrast, *status and reputation* were not found stimulating to the respondents.

Although the respondents expressed differing views about what makes the open education data hackathon fun and enjoyable, all agreed that it is crucial to have fun and enjoy working with data. For instance, two respondents working as a journalist and a workshop designer commented that the feeling of hacking *together* had triggered the fun, which the local government researcher also felt. Another respondent, the data scientist from an online travel agency, mentioned that the source of his fun and enjoyment in the hackathon was when he could discover important facts while working with data. The UX designer commented:

> *I think [...] the fun is [when] you work together [with] people and do some crazy things. [...] You'll [not only] learn new skills, but you [will also] find, most of the time, new connections because you work together. You'll find new things.*

One respondent, the workshops designer, stated that the hackathon's side event was held after lunch, which led to fun, stimulating social relationships, and better collaboration among team members. In contrast, the Ph.D. candidate said that the fun was working with the data and using it to overcome the challenges.

All respondents agreed that being intellectually challenged with educational issues contested in the hackathon had kept them engaged with open education data. The data scientist felt that the issues challenged him. He felt that his data analytics skills were under examination, and he wanted to pass the exam. Other respondents, the journalist, and the local government researcher perceived that the issues had challenged them to do something greater than they usually did use the data. Another respondent, the UX designer, mentioned that many data sets provided in the hackathon challenged him, while another respondent, the Ph.D. candidate, commented:

> *There were some [...] intellectual challenges [in] thinking about what are actual causes and differences [...] between correlation and causation [...] that we have to control to [...] come to the core relationship that we were interested in, in order to answer the questions that we were thinking about.*

However, the workshop designer felt that multiple socio-technical barriers obstructed her participation in the intellectual discussions about the data sets and how to use them to answer the hackathon challenges. She faced difficulty understanding the data sets because they were all in Dutch, and she did not have the required technical skills to analyze them. She also mentioned a social barrier: she did not know them well enough to cooperate intensely.

Concerning status and reputation, all respondents mentioned that this factor is not essential. The journalist claimed that his socio-economic status did not drive him to engage with the open education data. Instead, the desire to help others with lower social and economic status has driven him to engage with the data sets. Other respondents, the data scientist, the local government researcher, and the Ph.D. candidate, even said that no one is concerned with status and reputation because it is irrelevant in solving the hackathon challenges.

Findings related to the extrinsic motivations

Regarding extrinsic motivations, *getting to know new people* and *future career concerns* were found to play a role in stimulating the respondent engagement with open education data in the hackathon. Getting to know new people and future career concerns were important motivating factors of equally three respondents, while no respondent was motivated by *learning and skills development*. However, it is interesting that most respondents indicated that they might learn new skills or improve their existing skills during their participation in the hackathon, yet they barely did.

Most of the respondents considered meeting and getting to know new people who have similar interests in social issues such as inequality in education as a motivation to engage with open education data in the Hack de Valse Start hackathon. The local government researcher, the Ph.D. candidate, and the UX designer had similar opinions that by knowing new people from different backgrounds in the hackathon, they can learn new or different ways or approaches to work with data. Another respondent, the data scientist, wanted to meet new people. After all, he needed teammates to work with data because his colleagues did not participate in the hackathon. At the same time, another respondent, the workshop designer, commented that her primary motivation was to broaden social networking:

> *My motivation to come to the hackathon was to meet new people to see what brings them, their motivations to come to the hackathon, and ideally also to keep in touch with some of them. Actually, there aren't many platforms where you can meet with others interested in social issues and understand them through data. So, this hackathon is one of the rare*

occasions you can do outside of an academic environment, even though the format is not ideal.

Concerning future careers, some of the respondents were motivated to engage in the hackathon because the engagement may impact their job assignment or prospective employment. For example, the workshop designer organized an open data workshop that deals with social topics. Organizing such a workshop required her to use similar approaches employed by the hackathon for producing different outcomes of social interventions. This assignment had motivated her to participate in the hackathon to learn how to organize such events, facilitate participants, and stimulate participant engagement with datasets. Other respondents, the local government researcher, and the UX designer viewed their engagement as helpful to their future career because they won the competition, and their colleagues appreciated this winning. Another respondent, the Ph.D. candidate, also thought that his experience in solving societal challenges using data could add value to his future career if he decided to work outside academia.

Respondents stated that they were barely motivated to engage with the open education data to learn and develop new skills. They had a similar opinion that learning new skills or improving existing skills in the hackathon could attract participants. However, they admitted that they did not learn new skills nor enhance their existing skills either. The data scientist commented that a hackathon held in a limited time, i.e., one day, would not allow participants to acquire new skills. Another respondent, the journalist, wanted to find out whether he needs to learn how to code and analyze data as he stated:

I didn't want to improve my coding skills. I just wanted to see if there is a need to improve my coding skills because I didn't know anything yet. So, the question was, do I need to learn this and not to really get better at it, because for me, it was just looking at: should I learn it?

Findings related to the economic factors

The results show that most of the respondents were not motivated by economic factors. However, the journalist and the UX designer expressed their interest in winning the prize money offered in the hackathon. Although in the beginning, these two respondents were not focused on the money, later on, approaching the final presentation, they thought about the possibility of winning the challenge competition. The UX designer felt fine to add competition and a prize for the winners in hackathons. The prize money did not influence the data scientist to engage with open education data because the amount was not attractive. He considered it a little bonus if he could win the competition.

Similarly, the prize money did not influence the other respondents. While the local government researcher did not explain why she was not affected by the prize, the workshop designer and the Ph.D. candidate stated they were interested in other things beyond money.

Findings related to social factors

The motivation to *create benefits for society* appeared to be an important factor influencing the engagement with open education data. In addition, two respondents noted that the *influence from wider social relationships* (e.g., neighbors, communities, society) had motivated them to engage with the open education data. In contrast, none of the respondents stated that their *close social relationships* (e.g., family, friends, colleagues, supervisors) had influenced them to engage with the open education data.

The journalist, the workshop designer, and the local government researcher claimed that their motivation for creating societal benefits conforms with their occupation. The journalist felt sympathy with others and wanted to help them uncover inequality in education and its possible causes. The workshop designer wanted to transpose the techniques of engaging with data to a community that does not necessarily have a data science background. The researcher corroborated the connection between her motivation and occupation as she worked with similar data and inequality issues in specific neighborhoods, and she wanted to improve the situation. She highlighted the case of discrimination against students from particular background categories. In the Netherlands, the type of secondary education that a primary school student can go to is determined by her or his final exam score and teacher's advice. The researcher said that some teachers discriminate against students based on their backgrounds. The teachers recommended continuing at a lower level of secondary school to these students while giving a higher level of recommendation to other students from different backgrounds, although they all have a similar level of the final exam score. Another respondent, the data scientist, believed that the hackathon's findings and outcomes could contribute to societal benefits. Interestingly, the Ph.D. candidate boldly associated himself with those who had less access to better education as follows:

> *I care about society as a whole. So, I'm trying to improve things like education, and significantly, to help the weaker people in the community, create a society where everyone has more or less equal possibilities—particularly, [I tried] to help those with fewer opportunities to make something of their lives.*

Regarding the influence of wide social relationships, most respondents stated that this factor did not play a role. The data scientist and Ph. D. candidate acknowledged communities and society's values. However, the values did not influence them, particularly in open education data engagement. On the contrary, the workshop designer and the UX designer were closely related to the communities they worked with, the open data interpreter and IT communities. Their communities encouraged them to engage with open data and share their experiences with their peers.

All respondents stated that their close social relationships did not influence them, although their friends or colleagues also worked in data-related domains or education fields. For example, according to the data scientist, the local government researcher, and the UX designer, their colleagues were not interested in participating in a hackathon. The journalist had an appointment with his girlfriend, a teacher, to participate in the hackathon, but she canceled it. At the same time, the workshop designer's friends were critical and opposed the hackathon idea. She commented:

> *I happen to have very critical friends, and they were not very enthusiastic about hackathons. In a way, this feeling of wanting to guard themselves against a hackathon was transposed to me. I was quite reticent to consider the idea of a hackathon, and I'm not sure that I would have joined had there not been other factors, such as being contacted by the Open State Foundation. Being in touch with the people who organized it, having some sort of outside connection other than just knowing about the event from a web page motivated me to attend.*

Findings related to the technical factors

Regarding technical factors, different characteristics of the open education data quality and the quality of service provided in the hackathon appeared to influence the respondents. The researcher also found that the level of influence varies. The results show that data *understandability* and *interoperability* are the most important factors, while service reliability, data format, service assurance, and data completeness are of medium influence. At the same time, service *responsiveness*, data *accuracy,* and data *currency* are the least important factors.

All respondents mentioned that the open education data's understandability is important to enable them to do something with it. For the Ph.D. candidate and the UX designer, the data sets were well understandable. Likely, they can easily understand the data set's contents because both respondents are Dutch and regularly work with data. Similarly, the local government researcher also thought open education data were easy to grasp. Nevertheless, she experienced difficulty in comprehending non-educational data sets.

On the contrary, the journalist felt it challenging to discern the education data sets because he rarely worked with raw data. Other respondents, the workshop designer, and the data scientist mentioned that they needed help from the Dutch respondents to explain the meaning of the data sets. The workshop designer commented:

> *Once you knew what the tests were to get into the Dutch education system or what their symbol stood for, how they relate to the teachers' advice. It was quite easy, I think, for me. I meant that once you knew. So, once we knew after, the other two participants explained how the Dutch system works and what we should be looking out for. And at what point the teacher advised because this was also our focus point during the hackathon—those teachers' advice related to one particular exam. So, once that information was clear, it became easier to read. But indeed, without these two people in our group, again, I'm not sure I would have been able to get much very far.*

Except for the workshop designer, most respondents stated that data interoperability strongly influenced their engagement with open education data. She said that interoperability did not affect her because she did not work with the data, particularly combining data sets like another respondent, the data scientist. All respondents agreed that the ability to combine data sets is central to open data engagement. However, the Ph.D. candidate and the UX designer experienced difficulty combining data and needed considerable time to do so because they detected errors in the data sets. In contrast, another respondent, the data scientist, did not encounter difficulty. He commented:

> *[Interoperability] is the issue, specifically with this hackathon. The original formats, the open education data that we had, already implied that you have a really wide collection of separated [...] CSV files. That way, you had to be able to link them together simply [...] to do pretty much anything.*

The respondents mentioned service reliability as one factor influencing most respondents to engage with the open education data. This factor concerns the availability of a complete guide for accessing and using the data. All respondents agreed that the hackathon organizer had provided fairly complete data in a well-structured portal. However, the journalist needs additional data that deal explicitly with the education inequality topic. Another respondent, the workshop designer, thought that the hackathon organizer assumed that all participants already understood how to use the data. In practice, she could not comprehend the data. The local government researcher corroborated this statement by commenting that more explanation is needed because she had questions that were left unanswered. Another respondent, the UX designer, complained that although guides describing the data sets were available, the descriptions were not detailed. As a result, according to the data scientist, this situation complicated his engagement with open education data.

Four respondents stated that their engagement was influenced by the open education data format, mainly in CSV. The workshop designer was impressed because the provided data sets were well-formatted and -ordered. Similarly, the local government researcher and the UX designer stated that the data set's format was familiar to them. Another respondent, the data scientist, also agreed with the impression, yet he experienced difficulty handling the CSV files because the data descriptions were in Dutch. In contrast, two other respondents, the journalist and the Ph.D. candidate said that although the format was relatively known, it did not influence their engagement.

Four respondents mentioned that the service assurance factor was influential concerning specific staff assisting difficulties regarding open education data. On the contrary, other respondents, the data scientist and the local government researcher stated that the factor did not influence their engagement. The data scientist claimed that the hackathon organizer's help was lagging behind the outcomes of his work with data, while the researcher stated that the support was not adequate. Although the service assurance factor influenced other respondents, they had different experiences receiving help or support from the hackathon organizer. The workshop designer and the Ph.D. candidate felt very much helped to find the right or missing data. In contrast, other respondents, the journalist and the UX designer, asked for data sets related to the hackathon, but the data were unavailable. The journalist commented:

> *I was looking for a data set that wasn't there, but I thought we needed it, and there was someone who could help me. It took him some time to discuss if they could give the data, and it ended up that the data set wasn't available at the moment.*

Three respondents, the workshop designer, the data scientist, and the UX designer, stated that the completeness of data sets influenced their engagement. In contrast, other respondents, the journalist, the local government researcher, and the Ph.D. candidate, were not influenced by the factor. However, they had similar opinions that the data sets provided were not enough, and more data were needed to be combined and compared to support ideas from different points of view. The local government researcher pointed out that either the hackathon organizer did not provide specific non-educational data sets or that the data sets were unavailable. She thought she could approach the primary education problems using individual (or micro) data because both educational and non-educational data sets provided in the hackathon contained data aggregated at the school level.

The journalist and the workshop designer mentioned that the service responsiveness factor influenced them in engaging with open education data. This factor refers to the degree that open education data initiative staff always follow up with a user's report on data quality. However, the journalist had a similar opinion with other respondents. He, the Ph.D. candidate, and the UX designer stated that the hackathon organizer's personnel were responsive yet not resourceful enough to help them obtain specific needed data sets. The local government researcher did not consider the personnel responsive, while another respondent, the data scientist, claimed that he could analyze the data sets without the personnel's help.

Two respondents, the workshop designer and the Ph.D. candidate said that data accuracy influenced their engagement, while the factor did not influence other respondents. It is important to note that all respondents took the accuracy of data sets they engaged with for granted. Interestingly, some respondents from different backgrounds, the journalist with no data analytic skills, and the data scientist, mentioned that they could not validate the data and judge its accuracy.

The workshop designer and the local government researcher stated that the data currency factor influenced their engagement with open education data. Other respondents said that the factor did not influence their engagement. However, they all agreed that the open education data were up to date.

Findings related to the political factors

Concerning the political factors, the results indicate that the *possibility of political change* is an important factor that influences the respondent engagement with open education data. Most respondents stated that this factor influenced their engagement. Some respondents indicated that their *trust in OGD* influenced them to engage with open education data. Some respondents mentioned *interests in politics* and *involvement in political activities*, while none stated that *government responsiveness* had influenced them.

Except for the government researcher, all respondents agreed that the possibility of political change had motivated them to engage with the open education data. The researcher said that the outcome of the hackathon could not influence politicians and school management. She argued that websites built by the primary and secondary education council on top of open data about schools such as scholenopdekaart.nl are more influential. On the contrary, the data scientist believed that the hackathon's products might impact society. Another respondent who supported this belief, the UX designer, said that citizens' ability to monitor what the government is doing by engaging with open education data could stimulate changes. However, other respondents, the journalist and the Ph.D. candidate, disagreed. The Ph.D. candidate assumed that the hackathon's outcomes would not immediately change the current education policy. Although the journalist was motivated to change the policy using the hackathon outcomes, he admitted that it would not be realistic to influence politics in one day. Nevertheless, he was optimistic that in the long term, hackathons could be part of a change in education policy, as he commented:

> *I didn't go there because I thought I could make a difference and have any influence. That would be a bit naive, right, to go to a hackathon and think that you can really influence politics in one day. On the other hand, in the long run, there is no other reason than that you want to influence policy. So, this hackathon maybe was part of a process of changing education policy in the long run. But I really have a feeling that this hackathon wouldn't change anything about educational policy.*

Three respondents, the workshop designer, the Ph.D. candidate, and the UX designer, said that their trust in OGD influenced their engagement with the open education data in the hackathon. Interestingly, similar to data accuracy, the Ph.D. candidate and the UX designer took their trust in governments and their data for granted. At the same time, another respondent, the workshop designer, claimed that she had no confidence in the Romanian government, from which she came, but trusted the Dutch government. However, this factor did not influence the other respondents, including the journalist, the data scientist, and the local government researcher. While the data scientist and the local government researcher did not provide particular reasons, the journalist believed that the hackathon was merely part of the government's public relations programs.

Interests in politics had influenced some of the respondents, including the journalist, the workshop designer, and the UX designer, to engage with open education data. These respondents and another respondent, the Ph.D. candidate, claimed that they were interested in politics. The journalist claimed that the hackathon topic, inequality in education, was highly political. Other respondents, the workshop designer and the UX designer corroborated the statement. According to the UX designer, open education data enabled him to see what the government is doing with public money. At the same time, the workshop designer said that her teammates had done little using the data sets. In contrast, the data scientist and the local government researcher did not have interests in politics. The Ph.D. candidate stated that he was not motivated by political interests because he wanted to promote evidence-based policymaking.

Three respondents, including the journalist, the workshop designer, and the UX designer, stated that their involvement in political activities influenced their engagement with open education data. One respondent, the journalist, identified himself, particularly on the education inequality issues, as a non-partisan leftist. However, he was still officially registered to the Labor party. Another respondent, the workshop designer, claimed that she was involved in modern political activities by organizing open data workshops for citizens and providing a space to work with government data.

On the contrary, the factor did not influence the engagement of other respondents, the data scientist, the local government researcher, and the Ph.D. candidate. The data scientist stated that he has never been involved in public activities in the Netherlands. Another respondent, the Ph.D. candidate, was on a break from political activities, but he wanted to be involved more in the future as he commented:

> *I'm currently not really actively involved in any political activities anymore. But I would like to get more involved in the future. I think it really helps to talk to policymakers and meet people who were also interested in tackling societal challenges to see what kind of ideas are floating around to solve certain issues. Also, to look at the data that were available.*

The results show that government responsiveness did not influence respondent engagement with open education data. The workshop designer stated that the juries in the hackathon represented the governments decided which outcomes answer the education inequality challenges. According to the Ph.D. candidate, the government representatives were interested in the hackathon outcomes. Another respondent, the data scientist, said that he was not interested in building relationships or interacting with the Dutch government.

Findings related to the new factors missing from the literature

Factors found in the literature review were identified, and new factors missing from the literature were explored in the Hack de Valse Start case. Notably, the respondents were asked whether other factors influenced them to engage with open education data in the hackathon. Table 4.8 provides an overview of new factors missing from the literature but emerged in the case. Some respondents answered these questions, including the journalist, the data scientist, the local government researcher, and the Ph.D. candidate. Four factors influencing open education data engagement emerged from the interview data are: *broadening the respondent's horizon*, *interaction with the hackathon organizer*, *learning from others*, and *contributing to policymaking*.

Table 4.8. The overview of missing factors from the literature emerged in the Hack de Valse Start case.

Factors	**Evident in citizens interviewed**
Broadening my horizon	C1-01
Pre-event interaction with organizer	C1-03
Want to know what others can do with data	C1-04
Contribute to policymaking	C1-06

According to the journalist, his regular Saturday was usually spent watching Netflix and visiting a museum, but he wanted to do something new. He wanted to learn new skills from other people who have different expertise and feel new experiences. He wanted to broaden his current horizon with new skills and experiences. He was also curious about the hackathon. He knew about the Open State Foundation that organized the event and subscribed to its newsletter. However, he did not know how the organizer would manage the hackathon, and therefore, he wanted to understand more about the hackathon's organization.

Another respondent, the data scientist, said that he was impressed with his interaction with the hackathon organizer in a pre-event meeting. He was invited to a preliminary event in Amsterdam to bring together participants, introduce themselves to each other, and form a team. Although he could not speak Dutch, the data scientist attended the event and felt that the organizer helped him understand the hackathon's challenges and meet other participants.

The local government researcher has worked with data sets similar to open education data used in the hackathon for years. She used the data sets and created an analysis using the same method every year. She wondered what other people could do with the data and wanted to compare it with the regular research that she did for her employer. She was also interested in learning from other people's analyses.

The Ph.D. candidate saw the potential of open data for policymaking purposes. He also felt that education inequality issues were necessary for society to be solved. Therefore, he participated and wanted to explain how people can contribute to policymaking using open education data.

These answers did not appear as new factors missing from the SLR, and they can even be grouped into the factors that have similar characteristics proposed in the SLR. The two missing factors mentioned by the journalist and the local government researcher, i.e., *broadening the respondent's horizon* and *wanting to know what others can do with data,* represent the intrinsic motivation to learn new things. Another factor mentioned as missing by the Ph.D. candidate, i.e., *contributing to policymaking*, is a motivation to benefit society included in the social factors. The factor mentioned by the data scientist, i.e., *interaction with the hackathon organizer*, was neither a new missing factor. The pre-event meeting held by the hackathon sponsors arguably resembles the quality of services offered in the OGD engagement identified in technical factors.

4.3.2. Citizen-led OGD engagement

Background of the citizens

Table 4.9 depicts the profiles of the respondents of the Kawal Pemilu case. These respondents are citizens who were Kawal Pemilu's volunteers and have engaged with the open election data. Nine respondents were male, while three other respondents were female. Seven respondents were between 31 and 40 years old, five were between 41 and 50 years old, and two were above 50 years old. Based on the educational background, all of the respondents were highly educated. Six of them had a master's degree, four respondents had a bachelor's degree, and two respondents had a doctoral degree. The respondents' occupations were highly diversified. Four of the respondents worked in the IT domain, while most respondents worked in the social field.

Table 4.9. The profiles of citizens engaged with open election data.

Respondent ID	Age	Gender	Education	Occupation	Experience
C2-01	42	Female	Bachelor	Freelance translator	Regularly used open data
C2-02	52	Male	Master	Psychologist	The first use of open data
C2-03	40	Male	Master	Manager at company A	The first use of open data
C2-04	39	Male	Bachelor	IT consultant	Has used open data from different domains
C2-05	58	Female	Doctoral	University researcher	The first use of open data

Respondent ID	Age	Gender	Education	Occupation	Experience
C2-06	43	Male	Doctoral	Government researcher	The first use of open data
C2-07	34	Female	Master	Auditor	The first use of open data
C2-08	32	Male	Master	Software engineer	The first use of open data
C2-09	40	Male	Bachelor	Company employee	The first use of open data
C2-10	32	Male	Bachelor	Manager at company B	The first use of open data
C2-11	41	Male	Master	Entrepreneur	The first use of open data
C2-12	38	Male	Master	Software engineer	The first use of open data

Ten respondents engaged with open data for the first time in the Kawal Pemilu initiative. Only the translator and the IT consultant regularly engaged with open data from different domains. The translator utilized open data primarily for her work and activism, while the IT consultant used it to satisfy his curiosity about what the government was doing. Concerning awareness about OGD, all respondents, except the university researcher, know that governmental organizations provide open data. Moreover, according to manager B, all non-confidential government data should be opened by default. However, half of the respondents underlined the importance of opening more data. Some of these respondents emphasized different types of high-value data that the government should particularly open. For example, the psychologist and the IT consultant mentioned government financial data, while the government researcher said data related to policymaking, such as meeting minutes.

The company employee mentioned another example concerning data related to the past involving confidential data that has passed retention period, such as the United States of America Central Intelligence Agency’s data. Furthermore, the translator and manager B said that the government could improve OGD by providing structured raw data. The university researcher, the auditor, and the software engineer mentioned that more socialization is needed to increase the citizen awareness of OGD. The software engineer even highlighted the importance of incentivizing the use of OGD.

Table 4.10 provides an overview of the examined factors derived from the literature review in Chapter 3, in the Kawal Pemilu case. The same factors, excluding *system reliability*, were also examined in the Hack de Valse Start case (see Table 4.7). The results show that some factors contributed to the case. For example, some respondents mentioned fun and enjoyment, creating societal benefits, data accuracy, and government responsiveness. However, other factors derived from the model did not play a role in the case: status and reputation, influence from the close social relationship, and financial benefit.

Table 4.10. The overview of the factors, derived from the literature and examined in the Kawal Pemilu case.

Factors	**Play a role in the case?**	**Evident in citizens interviewed**
Intrinsic Motivations		
Fun and enjoyment	Yes	C2-01, C2-02, C2-06, C2-08, C2-10
Intellectual challenge	No	C2-08
Status and reputation	No	
Extrinsic Motivations		
Learning and skills development	Yes	C2-01
Getting to know new people	Yes	C2-01, C2-11, C2-12
Future career concerns	No	
Economic Factors		
Financial benefit	No	
Social Factors		
Influence from close social relationships	No	
Influence from wider social relationships	Yes	C2-01, C2-02, C2-03, C2-06, C2-09, C2-11, C2-12
Create benefit for society	Yes	C2-01, C2-02, C2-03, C2-04, C2-06, C2-07, C2-08, C2-09, C2-10, C2-11
Technical Factors		
Data quality: accuracy	Yes	C2-02, C2-03, C2-04, C2-06, C2-07, C2-08, C2-09
Data quality: completeness	No	
Data quality: format	Yes	C2-01, C2-02, C2-06
Data quality: currency	Yes	C2-02, C2-06, C2-11
Data quality: understandability	Yes	C2-08
Data quality: interoperability	No	
System quality: reliability	Yes	C2-04, C2-08, C2-10
Service quality: reliability	Yes	C2-02, C2-03, C2-04
Service quality: assurance	Yes	C2-01, C2-02
Service quality: responsiveness	Yes	C2-01, C2-06, C2-09, C2-10
Political Factors		
Trust in Open Government Data	Yes	C2-01, C2-02, C2-03, C2-04, C2-10
Government responsiveness	Yes	C2-01, C2-04, C2-06, C2-07, C2-10
Interests in politics	Yes	C2-02, C2-06, C2-07, C2-09, C2-11
Possibility of political change	Yes	C2-03, C2-06, C2-07, C2-11
Involvement in political activities	Yes	C2-01, C2-04

Findings related to the intrinsic motivations

The results show that *fun and enjoyment* are important intrinsic motivations that influence respondents to engage with open election data in the citizen-led engagement initiative. In contrast, *intellectual challenge* and *status and reputation* did not significantly impact the respondent engagement with the open election data. Only the software engineer mentioned that the election problem had challenged him intellectually to engage with election data.

Five respondents, the translator, the psychologist, the government researcher, the software engineer, and manager B, agreed that fun and enjoyment influence their engagement. However, these respondents had a different opinion about what makes the Kawal Pemilu initiative fun and enjoyable. For instance, the translator felt joy because she met new friends and quickly socialized with them. At the same time, the psychologist said that Kawal Pemilu was fun and enjoyable because its application was easy to use. The translator explained how it was fun interacting with other contributors as follows:

> *The fun was because the people involved in it were a bit crazy. One day in the group, for example, someone said, "Guys, please take a photo of your desks while doing data entry." The work tables of those who were already over forty with herbs to treat colds and ginger drinks appeared. Then some put piles of snacks. They were asked, "Is it data entry?" "Yes, data entry while snacking." The interaction with the people involved in the data entry itself was enjoyable.*

The software engineer stated that intellectual challenge influenced him to engage with the open election data. He felt challenged to find the solution in developing the application that many contributors can use at the same time to digitize the election results—solving technical challenges as part of his daily jobs at the software development company.

Findings related to the extrinsic motivations

Concerning extrinsic motivations, *getting to know new people* and *learning and skills development* influenced some respondents to engage with open election data in the Kawal Pemilu initiative. However, none of the respondents stated that future career concerns influenced their engagement. This conclusion can be drawn because the volunteers' occupations did not directly relate to the initiative.

Three respondents, the translator, the entrepreneur, and the software engineer, stated that they were motivated to engage with the open election data to get to know new people and broaden their social networks. The translator said that having new friends meant acquiring new insights and learning new things from them. The entrepreneur and the software engineer stated that they wanted to broaden their social networks by meeting new people. The entrepreneur further explained that widening his network would ease him organizing activities and initiatives with new acquaintances who share common perspectives and visions as he commented:

> *I was more into networking. It is getting to know a lot of people, understanding the ideas in their heads. If what is being thought is in line with or similar to mine, it can be potentially continued with various kinds of discussions and different other types of initiatives. For example, I got to know [xxx]. We shared many similar thoughts that intersect. In the end, we also discussed many other matters and organized some initiatives related to the open election data at a local level.*

One respondent, the translator, said that she was also motivated to learn from her engagement with the open election data. This motivation appeared to be driven by new people that she knew from the Kawal Pemilu initiative. She mentioned some contributors whom she became acquainted with during the engagement and insights that these contributors had given to her. For example, she mentioned one contributor working as a local urban planner who actively shared articles about Jakarta city planning from the *kampong* (local community) perspectives. She also mentioned another respondent, the entrepreneur living in the Netherlands, who frequently shared architecture pictures from European cities he had visited. She thought that she had learned new insights from these new acquaintances.

Findings related to the economic factors

The results show that none of the respondents were motivated by economic factors, such as receiving money or other financial rewards when engaging with the open election data. Instead, some of the respondents said that they had to finance the Kawal Pemilu initiative. For example, the translator claimed that she had to pay for the internet connection to verify the data herself. Another example concerns manager B, known as the initiator of Kawal Pemilu, who even said that he had not received any financial incentives to develop the open election data-based application. He further claimed that he and other founders had to pay all of the costs to build and maintain the application publicly accessible.

Findings related to the social factors

Concerning the social factors, the results show that *creating benefits for society* plays the most important role in influencing the respondents to engage with the open election data, followed by *wider social relationships*. Ten respondents said that their engagement was motivated to benefit the public. At the same time, seven respondents stated that their social relationships influenced them to engage with the open election data. In contrast, close social relationships' influence did not play a role in the open election data engagement. No respondent mentioned that this factor influenced their engagement.

Creating societal benefits played an important role in influencing nearly all respondents to engage with the open election data. These respondents claimed that they were motivated to contribute to benefit the public. However, these respondents had different opinions about what constitutes benefits for society. The psychologist, manager A, the government researcher, and the company employee had similar views that they have to share the results of the election data digitization with the public as widely as possible. The purpose of sharing the information is to inform the public about the election's outcome and scrutinize the election. Another respondent, the translator, wanted to ensure that the winning candidates were outputs of a clean, honest, and fraud-free election tally and minimized the possibility of social and political conflicts arising from a fraudulent election. Likewise, the auditor wanted to contribute to an honest and just election to guard democracy. Comparably, another respondent, manager B, commented that he tried to prevent further polarization among citizens:

> *So, what moved me at the earliest was seeing the danger of this nation being divided when Prabowo declared victory, and Jokowi also claimed victory. Both of them declared victory even though there were only two candidates, and the public had been highly divisive for months, even more than a year. And, we all know that it was precisely a fifty-fifty, neck-to-neck competition, a term used by one of the popular media at that time. Our nation was split in two. It was very dangerous because it can cause horizontal conflicts involving citizens. Therefore, I tried to find a solution to show who really won.*

Seven respondents said that some people or groups beyond their close social relationships influenced them to engage with the open election data. The translator said that a non-partisan community group where she actively organizes social activities heavily influenced her to engage with the open election data. Other respondents, the psychologist, the government researcher, and the software engineer, claimed that their distant friends encouraged them to involve in Kawal Pemilu. Interestingly, other respondents, manager A, the auditor, and the entrepreneur, stated that their Facebook friends whom they rarely interact with had invited them to engage with the open election data. The entrepreneur was willing to get involved because he and his Facebook friends had similar perspectives as he commented:

> *Frankly, most of my friends were on Facebook. Although I had Dutch friends, they had no interest in Indonesian politics. I actively expressed my opinions on my social media accounts. And I saw that some friends whose views are in sync with mine participated actively in voicing critical and transparent attitudes or participated in Kawal Pemilu. These friends hugely affected me to be more active.*

None of the respondents stated that their close social relationships influenced them, although their friends, colleagues, or family knew what they were doing. According to the translator and the government researcher, their close friends and family were not interested in the election, particularly politics. In a similar vein, manager B said that his wife did not know that he was involved in Kawal Pemilu. In contrast, another respondent, the software engineer, claimed that his family, mainly his wife supported his involvement in Kawal Pemilu.

Findings related to the technical factors

Regarding technical factors, different characteristics of the open election data quality, the quality of the system providing access to the open election data, and the service provided by governmental organizations publishing the open election data influenced the respondents. The results show that data *accuracy* played a moderately important role in influencing respondent engagement with open election data. Furthermore, service *responsiveness*, data *format*, data *currency*, service *reliability*, system *reliability*, service *assurance*, and data *understandability* were of little importance. At the same time, data *completeness* and *interoperability* did not play a role in the open election data engagement.

Seven respondents stated that data accuracy influenced their engagement with the open election data. They claimed that they were motivated to scrutinize election data and input accurate results into the Kawal Pemilu application. The IT consultant even checked the election result data against the actual result he had photographed at the polling station where he cast a vote. However, they also acknowledged that there were errors found during the data inspection. They further asserted that, in the end, the number of mistakes was insignificant compared to the actual election results.

Interestingly, these findings did not hinder the respondents from continuing their engagement with open election data. Instead, they became more enthusiastic about finding other errors. One of these respondents, the government researcher, specifically commented on this topic:

> *Before we started the recapitulation, we had a suspicion that there were many incorrect recaps on the forms, so we checked them. But after the checking, it turned out that there were one or two mistakes, but in my opinion, it could still be tolerated because the numbers were very few and would not affect the results. So, after the Kawal Pemilu activities, I became convinced that the data were actually accurate. Yet, those incorrect results affected society's perception. We were even more excited and enthusiastic to look for other problems if we met something like that.*

Four respondents said that service responsiveness influenced their engagement with the open election data. One of these respondents, the translator, witnessed the KPU personnel responsiveness and stated that it was helpful. Other respondents, the psychologist and the company employee corroborated the translator's statements. Both stated that in Kawal Pemilu, the translator played a role as a coordinator who collected all the incorrect election results found during digitization, compiled and reported them to another contributor who liaised with KPU to follow up. Another respondent, the company employee, claimed that he had once reported such an anomalous election result to his group coordinator and re-checked that the result was corrected and updated by KPU. Other respondents, manager B and the entrepreneur confirmed this claim. They gave an example involving results from a particular area in eastern Indonesia.

Three respondents, including the translator, the psychologist, and the government researcher, said that data format influenced their engagement with the open election data. They did not have a particular preference for the data type, yet they preferred an easy-to-understand format. They claimed that the open election data format was easily understood and, consequently, they can effortlessly contribute to Kawal Pemilu. One respondent, the government researcher, admitted that he would have been reluctant to contribute if the data format was difficult to grasp. The translator illustrated the effortless process of digitizing election results from contributors due to the easy-to-understand data format.

> *The data format was easy to understand. [...] All we had to do was look at the numbers. There were several columns. The number of ballots used, the number of votes for A, the number of votes for B, the number of ballot papers that were annulled, the number of ballot papers available, and those unused. It was actually easy. So, we just had to enter the numbers, and in*

> *the system, specific algorithms had also been created so that if the numbers didn't match, the flag would come out. Flag, it wasn't the job of data entry to correct it. It was verification's job.*

Three respondents, the IT consultant, the software engineer, and manager B, stated that system reliability influenced their engagement with open election data. The IT consultant and manager B explained that the Kawal Pemilu application heavily depended on the KPU portal that provides the open election data. The IT consultant further described that the contributors use the Kawal Pemilu application to access and display election data from the portal to digitize the election results in real-time. These three respondents agreed that the portal was remarkably reliable and survived during frequent attacks using specific hacking methods.

Three respondents, the psychologist, manager A, and the IT consultant, said that service reliability influenced their engagement with the open election data. They agreed that the supports given by the open election data provider (i.e., KPU) were substantially reliable. On the other hand, the psychologist believed that KPU did not provide a dedicated team to offer service to those interested in the election result data. However, some of its personnel corresponded with one of the coordinators of Kawal Pemilu regarding anomalous results. Another respondent, manager A, said he would report it to the coordinator who liaises with KPU personnel when he saw a questionable result.

According to three respondents, the psychologist, the government researcher, and the entrepreneur, the currency of the open election data was an important factor that influenced their engagement. Manager B labeled it a time-critical initiative because the public wanted to know the election's overall results promptly. The psychologist supported this opinion and claimed that the public demanded regular updates to the election results. He believed that if no progress were made, the public would start thinking that the vote recapitulation was experiencing severe problems. As a result, the public would begin distrusting the government. Another respondent, the government researcher, felt excited about digitizing the election results because they were up to date.

The translator and the psychologist stated that the service assurance factor influenced their engagement with the open election data. Although the psychologist maintained that KPU did not have a designated team providing support and help to open election data users, he believed that the coordinators of Kawal Pemilu contributors had communicated with the IT personnel of KPU. The translator, one of Kawal Pemilu's coordinators, supported this assumption. As a coordinator, she directly communicates with the personnel regarding anomalous results that must be checked and corrected. In her opinion, they were supportive:

> *When I was in charge of collecting the anomalous C1s [election result forms], I really knew that the KPU personnel who take care of the server must be very overloaded. Every time we send an email containing like twenty anomalous C1s, they would protest, 'Ma'am, please don't send twenty problems at a time. Five issues per email will make our checking easier.' But they corrected the C1s as fast as they can. So, even though they were busy and I did not know whether they were overwhelmed or not, at eleven o'clock in the evening or two o'clock in the morning, I sent them an email, and they would respond. Yes, they didn't respond at that time, but they definitely responded.*

Only one respondent, the software engineer, stated that data understandability influenced his engagement with the open election data. The respondent was one of the developers of the Kawal Pemilu application. He claimed that the process of tallying votes and its data structure were easy to grasp, and as a result, the understanding enabled him to advance in developing the application.

None of the respondents mentioned that data completeness and interoperability factors influenced their engagement with the open election data. One respondent, manager B, the founder of Kawal Pemilu, claimed that the election data were sufficiently complete to conclude the election outcomes. However, he admitted that data from remote areas, which were not densely populated, such as Papua's hinterland, were not entirely uploaded by KPU. Another respondent, the software engineer, believed that interoperability was not an essential issue in the open election data because all data supplied by KPU and linking one data set to another was deemed unnecessary.

Findings related to the political factors

The results show that different political factors influenced the respondent engagement with open election data. *Trust in OGD*, *government responsiveness*, and *interests in politics* played a moderately important role in influencing engagement. At the same time, the *possibility of political change* and *involvement in political activities* were of little importance.

Five respondents, the translator, the psychologist, the IT consultant, and two managers, stated that trust in OGD influenced their engagement with the open election data. Most of them trust KPU and the election data it published, yet at the same time, they are skeptical for different reasons. One respondent, the translator, trusted some of the KPU commissioners because she knew them personally and had high moral integrity. The other two respondents, the psychologist and the IT consultant trusted the KPU because, unlike its predecessor, it had published election results data.

On the other hand, one respondent, manager A, was skeptical because there exists a possibility that the election results data could be hacked. Another respondent, manager B, supported this assumption. He commented that election data were prone to manipulation:

> *We trusted, but we weren't blind. We believed that the KPU was professional, mostly because everything was opened because we mostly saw the data. But still, we took it with a grain of salt. We still didn't rule out the possibility that there may be elements in government organizations, such as the KPU, who might manipulate or participate in changing the election results. So, we kept on guard over that; we were skeptical. We downloaded all data that have been opened. If there were people inside KPU, who closed the open data for whatever reason, for example, turning off the server, we would've had backups. We trusted but maintained a critical attitude.*

Five respondents, the translator, the IT consultant, the government researcher, the auditor, and manager B, mentioned government responsiveness as an important factor that influences their engagement with the open election data. These respondents referred to KPU as part of the government. In the translator's opinion, if KPU were not responsive, Kawal Pemilu would have been a failed initiative. Other respondents, the government researcher and manager B agreed with the translator. The researcher said that the KPU was responsive toward complaints about anomalous data submitted by the Kawal Pemilu contributors and toward problems that went viral on social media platforms such as Facebook and Twitter.

Five respondents, the psychologist, the government researcher, the auditor, the company employee, and the entrepreneur, stated that their interests in politics influenced their engagement with the open election data. They claimed that they were highly interested in politics. One of the respondents, the psychologist, said that he would not have contributed to Kawal Pemilu if he had no particular political preference. Interestingly, a few respondents, including the translator and the IT consultant, stated that their political interests did not influence their engagement. Nevertheless, they had a desire related to politics. For example, the translator did not want a candidate associated with a human rights violation in the past to become a president. The IT consultant contributed to Kawal Pemilu because he wants the public to be politically literate on the government's tax money.

Four respondents, including manager A, the government researcher, the auditor, and the entrepreneur, said that the possibility of political change influenced their engagement with the open election data. They all agreed that political change was inevitable because it was the end of the previous administration, and the election would result in a new governmental regime. Two respondents, the auditor, and the entrepreneur, agreed that it was essential to transition to a cleaner government. Another respondent, the government researcher, wanted a change because he had experienced living under an authoritarian government:

> *The most significant influence is the desire for change. When I was a student, I had experienced how to live under the leadership of an authoritarian regime that is repressive. I didn't want such a rule to win the election. Therefore, I supported the Kawal Pemilu activities because there was a tendency that if one particular candidate won, he would establish an oppressive administration because he was part of the old regime. This would mean no change.*

Two respondents, the translator and the IT consultant stated that their involvement in political activities influenced their engagement with the open election data. One respondent, the translator, acknowledged that her involvement with political activists greatly affected her to engage in Kawal Pemilu. The IT consultant said that he was involved in a volunteer movement supporting one candidate. In contrast, ten other respondents claimed that they never engaged in political activities. One respondent, the government researcher, explained that the law prohibits him from participating in political activities as a civil servant.

Findings related to the new factors missing from the literature

In the Kawal Pemilu case, the factors that have been found in the literature review were identified, and new factors missing from the literature were explored. Notably, the respondents were asked whether other factors influenced them to engage with open election data in the Kawal Pemilu initiative. Nearly all respondents contributed to the question with different answers. Similar opinions were identified, and they can be categorized into relevant, similar groups of factors proposed in the literature review. Table 4.11 provides an overview of these new factors that emerged in the case. Five factors influencing open election data engagement emerged from the interview data: *personal satisfaction*, *data availability*, *novelty/new experience*, *desire to have a clean government*, and *social media sharing behavior*.

Table 4.11. The overview of missing factors from the literature but emerged in the Kawal Pemilu case.

Factors	Evident in citizens interviewed
Personal satisfaction	C2-01, C2-03, C2-04, C2-06, C2-07, C2-08, C2-10
Data availability	C2-01, C2-08, C2-12
Desire to have a non-fraudulent government	C2-01, C2-02, C2-05
Novelty/new experience	C2-01
Social media sharing behavior	C2-10

Seven respondents mentioned personal satisfaction as a factor that influenced their engagement with the open election data. However, what made them satisfied varies. For example, three respondents, the translator, the IT consultant, and manager B, were motivated to initiate and contribute to the Kawal Pemilu to derive satisfaction from preventing adversity and polarization. Another example includes three other respondents, manager A, the government researcher, and the software engineer, who satisfied their curiosity about the election outcomes because they had direct and faster results.

Three respondents, the translator, the software engineer, and another software engineer, mentioned data availability as factors that influence their engagement with the open election data. They agreed that the availability of election data was key to the success of the Kawal Pemilu initiative.

Three respondents, the translator, the psychologist, and the university researcher, stated that one of the main reasons they engaged with the open election data is their desire to have a clean government. One respondent, the freelance translator, admitted that she was active on social media. However, instead of being involved in heated discussions about who won the election, she preferred to ensure that the election was fair. She believed that a fair election was the ground for a clean government. Another respondent, the psychologist, claimed Indonesia's Corruption Perceptions Index had been the worst in the global context for decades. He believed that Kawal Pemilu had been an important example to show that a clean government is achievable.

One respondent, the translator, thought that the Kawal Pemilu initiative's novelty had influenced her to engage with the open election data. For the first time, as both a citizen and a commoner, she actively contributed to ensuring that the government cleanly and transparently tabulates the election results.

She claimed that Kawal Pemilu was the first to use open data, and such an initiative never existed before.

Another respondent, manager B, believed that social media sharing behavior influenced him and the contributors to engage with the open election data. According to him, they were motivated to share anomalous election results and inform their social networks and the public about the election outcome.

Almost all of these answers were not new factors missing from the literature review; they can even be grouped into factors with similar characteristics. *Personal satisfaction* mentioned by seven respondents concerns different factors. For example, three of these respondents said that they were motivated to gain satisfaction at preventing crises from happening. This motivation closely relates to benefitting society from social factors. Other respondents were motivated to satisfy their curiosity; this motivation represents an intrinsic motivation of enjoying the OGD topic. *Data availability* mentioned by some respondents also constitutes one of the technical factors under the data quality group, i.e., data completeness, while the *desire to have a non-fraudulent government* closely represents the social factor of benefitting society. One respondent, manager B, mentioned *social media sharing behavior* as a new missing factor. However, according to him, he wanted to share the anomalous election results the election outcome to the public. From his statement, it can be concluded that sharing information on social media is not a motivation in itself. Instead, his desire to post valuable information to the public, a manifestation of giving benefit to society, is the actual motivation. Therefore, this factor is not a new one either. Overall, only one new missing factor was identified, i.e., the *novelty* of experiencing the OGD engagement.

4.4. Cross-case analysis

This section discusses the cross-case analysis of the two case studies based on the findings reported in previous sections. Yin (1981) suggests that the case-comparison approach can produce beneficial results for cross-case analysis. This approach involves identifying similar factors across cases and explaining the differences between cases. Culture appeared to influence how the respondents carried out OGD engagement in this study. In the Hack de Valse Start case, team members proactively discussed their capabilities and openly informed them of their roles in solving the hackathon challenges. The team leaders did not direct or give instructions to other members to contribute to the solutions, and the members divided all necessary tasks among themselves according to their roles. The members also worked independently and adapted fluidly to the task workflow. The team focused on solving the hackathon challenge by creating a visualization based on the open education inspection data. On the contrary, in the Kawal Pemilu case, the volunteers had to wait for their group leaders to give orders and prioritize tasks. They liked to compare each other's works and created appointments with their peers to work together.

The volunteers were also motivated to compete in the leaderboard created by the developers to show that they had completed the digitization of the election results at the village level. Beyond the cultural influences, the type of data also influences how the respondents engage with OGD. In the Hack de Valse Start case, the metadata was written in Dutch and introduced a language barrier for the non-Dutch team members. Also, engaging with aggregated inspection data increases the difficulty of making a relatively accurate conclusion about the problems faced at the primary education level. In the Kawal Pemilu case, the government organization published election result data through APIs accessible and understandable to the developers. The volunteers did not have to understand the technical dimension of data access. This situation helped volunteers focus only on digitizing election results. Table 12 summarizes both factors identified from the conceptual model (see Section 3.5) and factors missing from the model that emerged in the cases.

Table 4.12. Comparison of factors found in both case studies.

Factors	**Identified in the engagement case**	
	Hack de Valse Start Hackathon (Government-led)	**Kawal Pemilu (Citizen-led)**
Intrinsic Motivations		
Fun and enjoyment	X	X
Intellectual challenge	X	
Status and reputation		
Extrinsic Motivations		
Learning and skills development		X
Getting to know new people	X	X
Future career concerns	X	
Economic Factors		
Financial benefit	X	
Social Factors		
Influence from close social relationships		
Influence from wider social relationships	X	X
Create benefit for society	X	X
Technical Factors		
Data quality: accuracy	X	X
Data quality: completeness	X	
Data quality: format	X	X

Factors	Identified in the engagement case	
	Hack de Valse Start Hackathon (Government-led)	**Kawal Pemilu (Citizen-led)**
Technical Factors		
Data quality: currency	X	X
Data quality: understandability	X	X
Data quality: interoperability	X	
System quality: reliability		X
Service quality: reliability	X	X
Service quality: assurance	X	X
Service quality: responsiveness	X	X
Political Factors		
Trust in Open Government Data	X	X
Government responsiveness		X
Interests in politics	X	X
Possibility of political change	X	X
Involvement in political activities	X	X
New Factors		
Novelty/new experience		X

Intrinsic motivations

The results show that *fun and enjoyment* are important intrinsic motivations that influence respondents to engage in both cases. In contrast, *status and reputation* did not play a significant role in the cases, although the literature mentioned it as an influencing factor. A possible explanation for this finding is that the respondents focused more on solving societal problems than individual benefits. Interestingly, *intellectual challenge* is identified only in the Hack de Valse Start case. This factor is not found in the Kawal Pemilu case, likely because both cases involved different task characteristics and citizens' roles carrying out the tasks. The Hack de Valse Start case involves intellectual tasks such as understanding metadata and drawing inferences from various data sets related to education and socio-economic conditions. At the same time, tasks carried out by contributors in the Kawal Pemilu case, such as inputting the number of election results and verifying numbers inputted to the application, are more personal and less intellective. Although other roles, such as application developers, are involved in intellective tasks, the developers interviewed indicate that the problems did not intellectually challenge them when engaging with the open election data. Based on their profiles, most developers work for leading internet platform companies such as Google. They have also participated and even won several times in the Olympics of Programming Competitions. Likely, developing the Kawal Pemilu application is not challenging.

Extrinsic motivations

Getting to know new people is a factor that influences citizen engagement in both cases. *Learning and skills development* are found only in the Kawal Pemilu case, while future career concerns only the Hack de Valse Start case. Different settings of the cases and citizens' profiles can explain this difference. A government-led engagement, such as a hackathon, generally provides its participants with limited time to engage with OGD. At the same time, citizen-led engagement may involve a limited but lengthy time.

Additionally, the hackathon participants interviewed indicated that if they learned new skills and developed existing skills, they must be associated with their occupation. However, these participants do not have enough time to learn and develop their specialized skills from others. In the Kawal Pemilu case, the contributors have relatively sufficient time (i.e., nearly two weeks) to interact with each other. The interactions give them ample opportunity to learn new skills or obtain new insights from their teammates.

Concerning the future career factor, the different settings of the cases and profiles of respondents interviewed might explain the difference between the two cases. The government-led engagement is organized as a scientific-like competition and provided with incentives for teams to win. Besides, most of the respondents are below their thirties and may look for a better job in the future. Therefore, winning a hackathon can be translated into an additional selling point in the participant's curriculum vitae for his or her career. On the contrary, the tasks carried out by contributors in the citizen-led engagement, inputting the election results, are rather run-of-the-mill. Most of the respondents have a steady job.

Economic factors

The results show that only respondents from the Hack de Valse Start case are influenced by *financial benefit*, while the factor does not influence respondents from the Kawal Pemilu case. Different settings of the cases can explain this contrast. The organizer holds the hackathon as a competition in which sponsors reward the winners with prize money. This financial reward appears to motivate respondents to engage with open education data. At the same time, the Kawal Pemilu case offers no financial incentives to contributors. In contrast, the initiators have to bear all the costs incurred during and after the engagement to enable public access to the election results without financial reward.

Social factors

In both cases, respondents are affected by *wider social relationships and the motivation to create benefits for society*. This finding indicates that citizens engaged with OGD in both cases, regardless of their nationality and culture, desire to improve social conditions by contributing to their community. In contrast to what has been found in the literature (e.g., Purwanto et al., 2019; Saxena & Janssen, 2017; Zuiderwijk, Janssen, et al., 2015), the *influence from close social relationships* did not play a role in both cases. These studies indicate that citizens may be encouraged by their colleagues or supervisors to engage with open data. A possible explanation may be that this research is focused more on contributing to solving social problems. At the same time, previous studies focus on using open data to support the respondents' work.

Technical factors

All technical factors were nearly found in both cases, except *data completeness* and *data interoperability* identified only in the Hack de Valse Start case and *system reliability* only in the Kawal Pemilu case. The difference in the case settings may explain the differences in the influence of data completeness and interoperability. In the Hack de Valse Start case, respondents have to draw inferences using different data sets published by various governmental organizations. Therefore, each data set's completeness ensures that the data sets are interoperable and respondents can link them. On the contrary, only one governmental organization published all data sets involved in the Kawal Pemilu case. The difference in the case settings can explain the difference in the influence of system reliability. In the Hack de Valse Start case, the hackathon organizer has provided the participants with all data sets related to the education inequality issues needed to solve the contested challenges. At the same time, the Kawal Pemilu contributors engage with data sets hosted online in the OGD provider web portal. The citizens in the Kawal Pemilu case are likely to be concerned about whether the portal or system that provides access to the open election data is available and responds to their requests timely during the engagement.

Political factors

Almost all political factors are found in both cases, while *government responsiveness* was identified only in the Kawal Pemilu case. The interaction between the citizens and personnel of governmental organizations that publish the data sets can explain this difference. In the Hack de Valse Start case, the interaction is more formal, and the time given for the hackathon team members to communicate with the government personnel is minimal. Therefore, it is also likely that the citizens involved in the case cannot assess the personnel's responsiveness and whether their responses in the communication are needed to solve the hackathon challenges. At the same time, although the Kawal Pemilu case interaction is relatively informal, the communication between the contributors and government personnel is intensive and is not limited to a particular schedule. Furthermore, the personnel is responsive toward complaints and questions submitted by the contributors; they follow up the feedback by correcting erroneous data.

New factors

Among the new missing factors mentioned by the respondents, one factor cannot be grouped into the factors identified in the theoretical framework. This new factor concerns the novelty of the OGD engagement experience derived from the Kawal Pemilu case. As a result, it seemed that the interview data had achieved saturation, and no more cases were needed. Most of the respondents mentioned new factors similar or overlapping with the factors proposed in the theoretical framework. Therefore, these factors were not deemed as new factors missing from the literature. However, among the mentioned factors, both cases share a common thread: the respondents are motivated to contribute to societal benefits.

4.5. Conclusion and answer to the second research question

The conduct of a systematic literature review and the proposal of a conceptual model for analyzing citizen engagement with OGD built on the review that Chapter 3 has discussed. In this chapter, applying the factors proposed in the conceptual model, qualitative case studies were reported. The report includes the design of multiple case studies, the case selection, the overview of two cases under investigation, the collection of qualitative data, and the establishment of chains of evidence. This chapter also discussed both the findings from within-cases and cross-case analysis based on the two case studies. Moreover, this chapter answers the second research question (RQ2): *why do citizens engage with OGD in existing government-led and citizen-led OGD initiatives?* Based on cross-case analysis results, it can be concluded that many different factors influence citizen engagement with OGD in government-led and citizen-led initiatives.

Interestingly, nearly all of these factors proposed in the theoretical framework in Chapter 3 were identified in either case. Only two factors that the literature had mentioned were not identified in the cases: *status and reputation* and *influence from close social relationships*. At the same time, only one factor missing from the literature review was identified.

Fifteen factors grouped in five categories from the proposed model (see Section 3.5), including intrinsic and extrinsic motivations, social, technical, and political factors, were influential to government-led and citizen-led OGD engagement. In contrast, two factors categorized in intrinsic motivations and the social factor category did not influence engagement types. Eight factors from five categories, i.e., intrinsic and extrinsic motivations, economic, technical, and political factors, were identified in one engagement type. Five of these factors are intrinsic and extrinsic motivations, economic and technical factors that influence government-led engagement. In contrast, three other factors from extrinsic motivations, technical and political factors were found in the citizen-led engagement. Only one new factor missing from the literature was identified in the citizen-led engagement.

Regarding citizens' profiles, the evidence collected did not support analyzing whether a particular factor has more influence on a citizen with a specific profile. Furthermore, the number of citizens who participated in the case study is substantially low, and thus, investigating their age, gender, and other background factors has no use. From an ethical perspective, studying these citizens' profiles may also be considered violating their privacy. However, in both cases, some of the respondents were politically active, and their profiles resembled that of activists. They engaged with OGD to expose societal problems and seek out solutions to these problems.

The outcomes of this research stage, i.e., the findings from the two case studies, lay the ground for developing a quantitative research model that the final research stage will assess. In summary, based on the literature review and case study results, the following propositions are formulated.

Proposition 1. Intrinsic motivations positively influence the intention to engage with OGD.

Proposition 2. Extrinsic motivations positively influence the intention to engage with OGD.

Proposition 3. Social factors positively influence the intention to engage with OGD.

Proposition 4. Technical factors positively influence the intention to engage with OGD.

Proposition 5. Political factors positively influence the intention to engage with OGD.

Proposition 6a. Citizens' age moderates the positive relationship between intrinsic motivations and the intention to engage with OGD.

Proposition 6b. Citizens' gender moderates the positive relationship between intrinsic motivations and the intention to engage with OGD.

Proposition 6c. Citizens' education level moderates the positive relationship between intrinsic motivations and the intention to engage with OGD.

Proposition 6d. Citizens' OGD experience moderates the positive relationship between intrinsic motivations and the intention to engage with OGD.

Proposition 7a. Citizens' age moderates the positive relationship between extrinsic motivations and the intention to engage with OGD.

Proposition 7b. Citizens' gender moderates the positive relationship between extrinsic motivations and the intention to engage with OGD.

Proposition 7c. Citizens' education level moderates the positive relationship between extrinsic motivations and the intention to engage with OGD.

Proposition 7d. Citizens' OGD experience moderates the positive relationship between extrinsic motivations and the intention to engage with OGD.

Proposition 8a. Citizens' age moderates the positive relationship between social factors and the intention to engage with OGD.

Proposition 8b. Citizens' gender moderates the positive relationship between social factors and the intention to engage with OGD.

Proposition 8c. Citizens' education level moderates the positive relationship between social factors and the intention to engage with OGD.

Proposition 8d. Citizens' OGD experience moderates the positive relationship between social factors and the intention to engage with OGD.

Proposition 9a. Citizens' age moderates the positive relationship between technical factors and the intention to engage with OGD.

Proposition 9b. Citizens' gender moderates the positive relationship between technical factors and the intention to engage with OGD.

Proposition 9c. Citizens' education level moderates the positive relationship between technical factors and the intention to engage with OGD.

Proposition 9d. Citizens' OGD experience moderates the positive relationship between technical factors and the intention to engage with OGD.

Proposition 10a. Citizens' age moderates the positive relationship between political factors and the intention to engage with OGD.

Proposition 10b. Citizens' gender moderates the positive relationship between political factors and the intention to engage with OGD.

Proposition 10c. Citizens' education level moderates the positive relationship between political factors and the intention to engage with OGD.

Proposition 10d. Citizens' OGD experience moderates the positive relationship between political factors and the intention to engage with OGD.

It is important to note that the economic motives found in the case study (i.e., the Hack de Valse Start case) can only be generalized to OGD engagement initiatives that offer monetary rewards. Other initiatives, especially citizen-led (and some forms of government-led), do not provide such incentives. Therefore, the researcher does not postulate that economic factors are positively related to engaging with OGD. Although generalization about the roles of citizens' profiles in the case study cannot be drawn, they are tested in the following research phase because the literature indicates their moderating roles (see Section 3.2.2). Understanding whether age, gender, education level, and experience with OGD moderate the relationships between factors and intention to engage with OGD was particularly the research interest. The new factors found in the case study results were excluded because they were bound particularly within the context of the cases.

5. Modeling OGD Citizen Engagement

In Chapter 4, a set of propositions about the factors that influence citizens to engage with OGD based on the multiple case study findings has been developed. In the propositions, intrinsic motivations, extrinsic motivations, economic factors, social factors, technical factors, and political factors were hypothesized to influence citizens' intention to engage with OGD positively. Citizens' profiles such as age, gender, education level, and OGD experience were also hypothesized to moderate the positive relationships between factors and intention to engage with OGD. These propositions can be generalized only to the particular contexts of the multiple case studies. In contrast, this study aims to develop a research model that can be generalized to a larger sample of citizens regardless of the engagement context. Therefore, in this chapter, hypotheses are formulated based on the propositions and tested using quantitative research methods. More specifically, open data users who have experience with OGD engagement were surveyed to assess the hypotheses.

This chapter addresses the third research question: *what model explains citizens' intention to engage with OGD?* The partial least squares structural equation modeling (PLS-SEM) approach is used to answer the question. PLS-SEM enables researchers to evaluate the associations between constructs (factors) and their latent variables (indicators) and the structural model hypothesizing the causation among constructs. This chapter is divided into four main sections. First, the justification for applying a PLS-SEM technique and the research model is described (Section 5.1). Reports about research instruments developed for collecting quantitative data, data collection processes, and pre-analysis processes for preparing data (Section 5.2) follow this section. Next, the results of the PLS-SEM approach concerning the descriptive analysis, measurement model assessment, structural model assessment, moderation analysis, multigroup analysis, and importance-performance map analysis are reported in Section 5.3. Then, Section 5.4 discusses the interpretations of the PLS-SEM results, followed by the answer to the research question and the conclusions of the quantitative study. We have published parts of this chapter in Purwanto et al. (2020c).

5.1. Research design

In this section, the proposed hypotheses and the research model are presented. As mentioned in the previous section, an analytical technique named PLS-SEM was employed to test the hypotheses and assess the model. This section describes the grounds for using the technique and the guideline on its application adopted from Hair et al. (2017).

5.1.1. Research model

At the end of Chapter 4 (see Section 4.5), a set of propositions about factors influencing citizen engagement with OGD has been presented. In this research phase, these propositions were formulated into testable hypotheses. The hypotheses relate intrinsic, extrinsic, social, technical, and political factors influencing citizens' intention to engage with OGD. The following ten hypotheses are tested.

H1 *Intrinsic motivations positively influence intention to engage with OGD*

H2 *Extrinsic motivations positively influence intention to engage with OGD*

H3 *Social factors positively influence intention to engage with OGD*

H4 *Technical factors positively influence intention to engage with OGD*

H5 *Political factors positively influence intention to engage with OGD*

H6a *Citizen's age moderates the positive relationship between intrinsic motivations and intention to engage with OGD*

H6b *Citizen's gender moderates the positive relationship between intrinsic motivations and intention to engage with OGD*

H6c *Citizen's education level moderates the positive relationship between intrinsic motivations and intention to engage with OGD*

H6d *Citizen's experience moderates the positive relationship between intrinsic motivations and intention to engage with OGD*

H7a *Citizen's age moderates the positive relationship between extrinsic motivations and intention to engage with OGD*

H7b *Citizen's gender moderates the positive relationship between extrinsic motivations and intention to engage with OGD*

H7c *Citizen's education level moderates the positive relationship between extrinsic motivations and intention to engage with OGD*

H7d *Citizen's experience moderates the positive relationship between extrinsic motivations and intention to engage with OGD*

H8a *Citizen's age moderates the positive relationship between social factors and intention to engage with OGD*

H8b Citizen's gender moderates the positive relationship between social factors and intention to engage with OGD

H8c Citizen's education level moderates the positive relationship between social factors and intention to engage with OGD

H8d Citizen's experience moderates the positive relationship between social factors and intention to engage with OGD

H9a Citizen's age moderates the positive relationship between technical factors and intention to engage with OGD

H9b Citizen's gender moderates the positive relationship between technical factors and intention to engage with OGD

H9c Citizen's education level moderates the positive relationship between technical factors and intention to engage with OGD

H9d Citizen's experience moderates the positive relationship between technical factors and intention to engage with OGD

H10a Citizen's age moderates the positive relationship between political factors and intention to engage with OGD

H10b Citizen's gender moderates the positive relationship between political factors and intention to engage with OGD

H10c Citizen's education level moderates the positive relationship between political factors and intention to engage with OGD

H10d Citizen's experience moderates the positive relationship between political factors and intention to engage with OGD

Economic factors such as *financial gain* and *monetary rewards* were not tested in this research phase (see Section 4.5 for a detailed explanation). Instead of testing the economic factors, an indicator of extrinsic motivations named *relative advantage* with a broader meaning of *gain* or *reward* was tested. Engaging with OGD is assumed to offer many benefits, for example, improving job performance (Zuiderwijk, Janssen, et al., 2015), which are not limited only to monetary measures.

Figure 5.1 provides a schematic representation of the research model that this research phase assesses. The model consists of five independent variables (factors), comprising twelve indicators influencing citizens' intention to engage with OGD, and one moderating variable (factor) hypothesized to affect the influence of the independent variables. The model is assessed using a larger sample of citizens in this research phase than the multiple case studies in the previous stage.

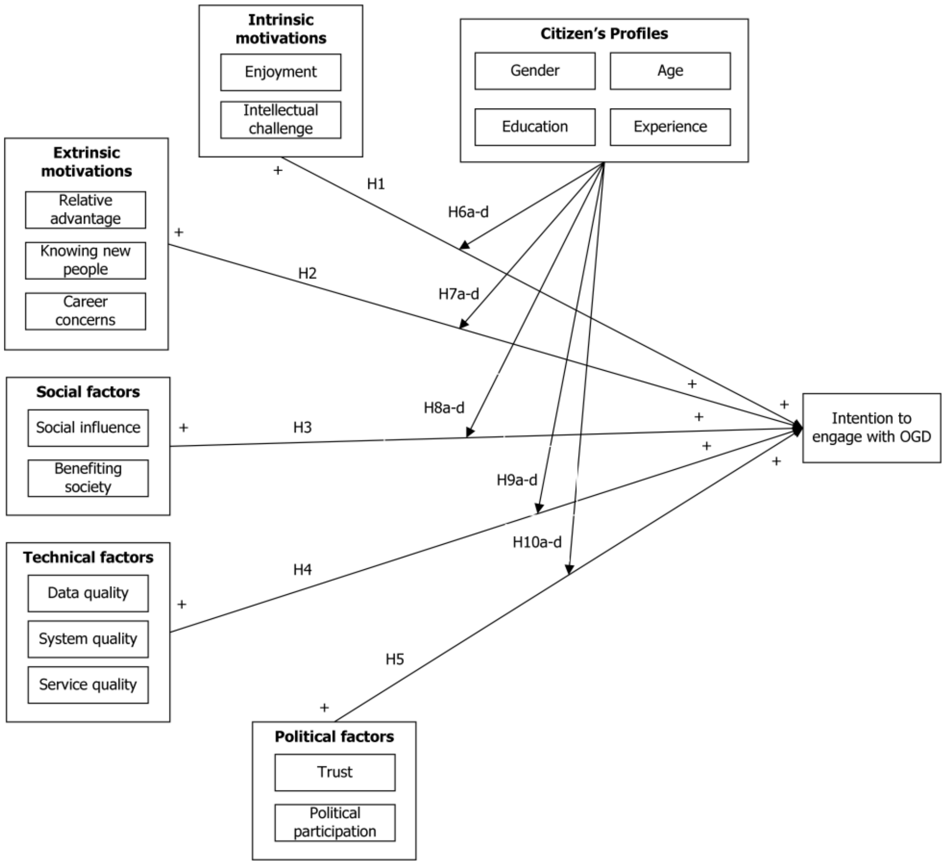


Figure 5.1. The evaluated research model of OGD citizen engagement.

5.1.2. Partial least squares structural equation modeling (PLS-SEM)

All too often, researchers face a set of interrelated research questions involving relationships between multiple dependent and independent variables. This situation also applies to this study. However, statistical techniques such as multiple regression or factor analysis can test only a single relationship (Hair, Black, et al., 2014). *Structural equation modeling* (SEM) is a set of statistical techniques extending factor analysis and multiple regression analysis for assessing hypotheses about relations between one or more independent variables and dependent variables (either continuous or discrete) (Hair, Black, et al., 2014; Hoyle, 1995; Ullman, 2006). One of its features is the ability to specify latent variable models providing estimates of *measurement model* (the relations among latent constructs and their manifest indicators) and *structural model* (the relations among constructs) (Tomarken & Waller, 2005). The research model involves six latent constructs (previously labeled *factors*): intrinsic motivations, extrinsic motivations, social factors, technical factors, political factors, intention to engage with OGD, and one observed variable named citizens' profiles. In this research phase, the model is tested to understand the relations between each construct (factor) and its indicators (variables comprising the construct) and among constructs. Therefore, SEM is deemed appropriate for testing these relationships.

SEM integrates two basic statistical approaches, *exploratory factor analysis* (EFA) and *confirmatory factor analysis* (CFA), and this integration constitutes the core strength of SEM (Hoyle, 2012). The goals of EFA are twofold (Fabrigar & Wegener, 2012; Hair, Black, et al., 2014; Ullman, 2006). First, EFA aims to uncover the model structure *unknown* to the researcher from a large set of variables. Second, EFA aims to define factors and highly interrelated sets of variables that constitute the model structure. CFA aims to test a hypothesized structure or competing theoretical models about the model structure based on a theoretically known assumption (Ullman, 2006). This assumption concerns the number of constructs, relations among the constructs, and the relationship between the constructs and measured indicators. Researchers can use SEM to perform both factor analyses (i.e., EFA's main feature) in a restrictive way because the assessed model structure is *a priori* and CFA's regression analysis simultaneously (Hoyle, 2012). SEM also allows researchers to evaluate an a priori model and compare it with alternative models that reflect competing theories or offer a more parsimonious account of the data (Hoyle, 2012; Ullman, 2006).

Currently, two types of SEM exist covariance-based SEM (CB-SEM) and PLS-SEM (also named PLS path modeling) (Hair et al., 2017). The PLS-SEM technique was chosen for the following reasons.

First, researchers consider CB-SEM a confirmatory technique because they use it primarily to confirm or reject theories (i.e., hypotheses) (Hair et al., 2017). CB-SEM technique calculates to what extent a proposed theoretical model can approximate the covariance matrix of a sample. On the other hand, PLS-SEM is exploratory in nature because researchers use it primarily for developing theories. Researchers focus on explaining the variance in the dependent variables with PLS-SEM when investigating the proposed model. Researchers have also applied PLS-SEM extensively in the IS literature. However, researchers should base their decision of selecting either CB-SEM or PLS-

SEM on the research objective (Hair, Ringle, & Sarstedt, 2011). Open data research is still at an infancy stage, as the lack of theories developed mainly for open data usage among citizens (Wirtz et al., 2019). A theoretical model that explains open data adoption by individual citizens is needed (Hossain et al., 2016). This study aims to fill this gap by developing a model that explains the factors influencing citizens' intentions to engage with OGD (see Section 1.5). Since building theory is more emphasized rather than testing strong prior theory, this study is exploratory. Therefore, it is assumed that the PLS-SEM technique is more appropriate for this study than CB-SEM.

Second, the CB-SEM technique only applies to sample data that follows a multivariate normal distribution (Hox & Bechger, 1998; Kline, 2016; Ullman, 2006). On the contrary, the PLS-SEM technique uses very soft distributional assumptions (Chin, 2010); Hair et al. (2017) even argue that it makes no distributional assumptions. This study focuses on citizens' attitudes toward particular objects (i.e., intrinsic and extrinsic motivations and social, technical, and political factors) to explain their intention to engage with OGD. Both attitudes and intention are hypothetical constructs inaccessible to direct observation and measured using self-report instruments (Ajzen, 2005; Fishbein & Ajzen, 2010). Ajzen (2005) describes an attitude as "a disposition to respond favorably or unfavorably to an object, person, institution, or event" (p. 3). Given this definition, instruments used to measure attitudes must provide responses that reflect positive or negative evaluations of the objects. Researchers typically build a self-reporting instrument for measuring attitudes on *Likert scales* (Likert, 1932). Likert scales enable researchers to measure positive or negative evaluations of an attitude object and strength of behavioral intentions by responding from, for example, "strongly disagree" to "strongly agree." Given its nature, researchers classify a Likert scale as an ordinal level of measurement (Hair, Black, et al., 2014). In ordinal measurement, "the response categories have a rank order, but the intervals between values cannot be presumed equal" (Jamieson, 2004, p. 1217). Ordinal data such as Likert-type data are discrete and cannot normally be distributed (Finney & DiStefano, 2013). Since Likert scales were used to measure the evaluation of the factors and the intention to engage with OGD, PLS-SEM was deemed the most appropriate approach for achieving the objective of this research phase.

5.2. Research approach

This section discusses the research approach taken to design the online survey instrument, disseminate the survey to collect data from samples and prepare the collected data for subsequent analyses. First, Section 5.2.1 provides a discussion and rationalization of selecting the survey research strategy and a detailed description of the research instrument development.

Next, Section 5.2.2 discusses the sampling technique adopted in this research phase and how the online survey was distributed. Finally, Section 5.2.3 offers a discussion about processes that need to be taken into account before reporting the results of the PLS-SEM technique.

5.2.1. Instrument design

A survey was chosen as the main instrument for collecting data in this research phase. A survey can be defined as "a systematic method for gathering information from (a sample of) entities for the purposes of constructing quantitative descriptors of the attributes of the larger population of which the entities are members" (Groves et al., 2009, p. 2). The efficiency of survey research on extrapolating data from samples to a population of interest has led to its growing popularity in many disciplines (Lee, Benoit-Bryan, & Johnson, 2011), including information systems (Pinsonneault & Kraemer, 1993) and public administration (Enticott, Boyne, & Walker, 2009). The survey is one of the most widely applied methods for investigating OGD use (Purwanto et al., 2020a).

A web-based, online survey was selected among different surveys such as internet, telephone, and postal/mail surveys (Dillman, Smyth, & Christian, 2014). In general, online surveys offer advantages, such as inexpensive, quick, efficient, immediate data entry, effective anticipated questions, global reach, flexibility, and technological innovations (Evans & Mathur, 2018; Sue & Ritter, 2007, p. 7). The decision to use an online survey as the data collection technique was deemed justifiable for the following contextual reasons. First, a database about the community of OGD users does not exist (Beno et al., 2017), and the community's membership is not clearly defined (Martin, 2014). Second, defining the boundaries of the potential OGD user community is challenging because the openness nature of OGD leads to use by unpredicted actors (Martin, 2014). Third, according to recent studies, engaging with OGD requires citizens' Internet competence (Wirtz et al., 2018) because they have to perform activities on the Internet (Jurisch et al., 2015). Therefore, using an internet-based survey can tackle the problems related to identifying the sample of this research and confirm the current research findings.

Despite their advantages, web surveys continue to experience low response rates (Evans & Mathur, 2018). Compared to other conventional survey methods, response rates of web surveys are lower (Kaplowitz, Lupi, Couper, & Thorp, 2012). Particularly, web surveys completed on mobile devices suffer a lower response rate than those on computers (de Bruijne & Wijnant, 2013). Pre-notifications and follow-up reminders are approaches that researchers can use to tackle issues related to low response rates (Mellahi & Harris, 2016).

In this research, the survey was designed and created using the Delft University of Technology software, namely the Collector[15]. The software has a simple design that provides respondents with many features and a simple user interface throughout the generated survey webpages. Furthermore, the software generates survey webpages that support different browsers and allows various types and formats of questions, including open-ended answers. Furthermore, it requires only basic technical knowledge, which in turn is helpful for a first-time researcher to create a web-based survey.

The survey used mainly closed-ended questions in this study and consisted of fifty-two questions grouped in seven sections (see Appendix C). The survey commenced with an introductory page that serves as a protocol. This page provided an explicit statement about the aim and relevance of this research. The researcher also published his contact information on this page. Also, the respondents were provided with a brief statement pointing out the reasons behind their selection as candidates and approximating the time needed to complete the survey. The introductory page also briefly explained the anonymity and voluntariness of the respondents and presented information about the treatment of personal and non-personal data. Table 5.1 provides an overview of the structure of the survey instrument. The following sections introduce and explain each section of the structure.

Table 5.1. The overview of the survey structure.

Section	Area
A	Experience in OGD engagement
B	Personal drivers (intrinsic and extrinsic motivations)
C	Technological factors
D	Social factors
E	Political factors
F	Behavioral intention
G	Respondent demographic information (including personal data, i.e., email address)

Section A: Experience in OGD engagement

Section A was designed to elicit information on a respondent's experience in engaging with OGD from different angles, i.e., the time, type of engagement, domain of OGD, output of engagement, and relation of purpose with solving societal problems. The definitions of *OGD* and *engagement with OGD* and their examples were provided at the beginning of the section. The first question asked in this section was whether respondents have ever engaged with OGD given the definitions and examples. It was crucial because it served as the

[15] https://tbm.collector-survey.tudelft.nl/nq.cfm?q=A6FA7521-9391-4405-B613-000D807ADBAF

primary filter for respondents to continue completing the survey or stop immediately. In addition, this research phase aims to test the hypotheses and assess the research model (see Section 5.1.1) among a larger sample of citizens who *have experience* in engaging with OGD. Therefore, the question was designed to filter out respondents who have no experience. If respondents answered “No,” the page was automatically directed to its end; otherwise, they proceeded to the following questions. The next questions were designed to capture the respondent’s typical experience in engaging with OGD, while the second question captures information on the last time respondents engaged with OGD. The third question asked respondents’ typical engagement within four categories: self-organized individual engagement, organization-sponsored individual engagement, self-organized collective engagement, or organization-sponsored collective engagement. The fourth question dealt with domains of OGD in which respondents frequently engaged, while the fifth question elicited information on the typical output of respondents’ engagement. At the end of the section, the respondents were asked to identify whether their engagement goal solved societal problems. Different scales such as nominal and Likert scales were used to measure the answers to these questions. Table 5.2 provides an overview of the questions asked in this section, including the scales and measurements used.

Section B: Personal drivers (intrinsic and extrinsic motivations)

The survey questions in Section B were designed to obtain information about a respondent’s motivations (both intrinsic and extrinsic) to engage with OGD. These questions were adapted from previous literature and modified to fit the context of this research. The questions were measured in Likert scales ranging from “strongly disagree” to “strongly agree.” Some of the questions have been used in the previous open data survey (see Chapter 3) except EXT4 related to career concerns. This question was derived from a multiple case study of open data hackathons in the UK (Kuk & Davies, 2011). It became widely known that programmers who engage in hackathons want to show the public what they can do with their programming skills and impress prospective employers. One question related to a motivation to know new people (EXT3) was derived solely from a survey of Bavarian citizens engaging in open government platforms (Hutter et al., 2011). These citizens were interested in building communities with like-minded who previously participated in such platforms. Therefore, it was decided to include these questions in the survey section.

Section C: Technological factors

The questions formulated in Section C were developed to gather information on respondents’ evaluation of the quality of the OGD (data quality), OGD systems (system quality), and OGD services (service quality) they frequently engaged.

The questions were grouped according to the three constructs generated in the systematic literature review (see Section 3.4 and 3.5). All questions were measured in Likert scales ranging from “strongly disagree” to “strongly agree.” Most of the questions were adopted from previous open data research, which mainly employed a case study approach, except OGD system availability (SYSQ1) and empathy of OGD service (SERVQ4). In the Kawal Pemilu case, the availability of the systems providing access to OGD, such as an open data portal, was an essential factor because the application developed by citizens requires a real-time display of the opened election results (see Section 4.3.2). Prioritizing the OGD users’ needs was also significant because in both studied cases, citizens felt that the supports from the OGD personnel were needed either to obtain different data sets in the hackathon (see Section 4.3.1) or correct anomalous election results data (see Section 4.3.2). One question explicitly related to data interoperability, i.e., DQ4, was added to data quality. The DeLone and McLean's (1992) information quality and Wang and Strong's (1996) data quality concepts did not include interoperability because their works focused on data and information as the output of a single information system.

In contrast, OGD is predominantly raw data and involves different data sets generated by different governmental organizations. Developers or programmers that engage with OGD need to combine these data sets to generate meaningful facts. Therefore, it is urgent to include interoperability in the OGD quality questions.

Section D: Social factors

Section D was designed to collect respondents’ evaluation of social relationships’ influence on their OGD engagement. The questions were measured in Likert scales ranging from “strongly disagree” to “strongly agree.” All questions were adopted from previous open data research employing either a survey approach, e.g., Zuiderwijk, Janssen, et al. (2015), or a case study, e.g., Hivon and Titah (2017). Three of the questions were derived from the unified theory of acceptance and use of technology (Venkatesh et al., 2003) and modified in the context of this research. One new question (SOC4) was also added. The case study showed that social values such as benefitting society (SOC4) are important factors influencing citizens in both government-led and citizen-led OGD engagement (see Section 4.4). Therefore, the question was added in this survey section.

Section E: Political factors

Survey questions in Section E were designed to elicit information about respondents’ evaluation of their trust toward OGD and political participation.

The questions were formulated to operationalize these two constructs, i.e., trust and political participation, proposed in the systematic literature review (see Section 3.4 and 3.5). All questions were measured in Likert scales ranging from "strongly disagree" to "strongly agree." Three questions on political participation were adopted from previous OGD research (Hutter et al., 2011; Jurisch et al., 2015; Wijnhoven et al., 2015). These questions were derived from political participation studies (Brady, Verba, & Schlozman, 1995; Quintelier & Vissers, 2008) and modified to fit the context of this research. Opening government data is expected to build citizens' trust (Janssen et al., 2012) and is assumed to strengthen trust in government (Cranefield et al., 2014). However, the effects of OGD on citizens' trust are merely conjectural because the literature lacks research that can provide such empirical evidence (Safarov et al., 2017). Conversely, according to e-government research (e.g., Carter & Bélanger, 2005; Teo, Srivastava, & Jiang, 2009; Warkentin, Gefen, Pavlou, & Rose, 2002), trust is found to be an essential catalyst of citizens' intention to engage with e-government applications. A recent study from Indonesia, included in the SLR (see Chapter 3), found that citizens' trust in open data websites influences their intention to continually use the OGD website (Fitriani et al., 2019). Therefore, questions related to trust from e-government literature were added and modified to fit the context of this research. However, instead of adopting trust in open data website construct from Fitriani et al. (2019), the researcher maintains that, following Zuiderwijk (2015), trust in OGD combines trust in government and trust in the OGD.

Section F: Behavioral intention

Section F was designed to obtain information about a respondent's intention to engage with OGD in the future. All questions asked in this section were measured in Likert scales ranging from "strongly disagree" to "strongly agree." Fishbein and Ajzen (2010) define behavioral intentions as "indications of a person's readiness to perform a behavior" (p. 39). The fundamental dimension that characterizes an intention is estimating the perceived probability or likelihood of a person performing a given behavior. Researchers expect that the higher this subjective probability, the more likely the behavior under question will be performed. Questions reflecting a citizen's intention to engage with OGD from Zuiderwijk, Janssen, et al. (2015) were adopted and modified to fit the context of this research.

Section G: Demographic information

Section G was designed to gather information about a respondent's demographic background such as gender, age, nationality, education level, work status, and current job. Citizens' profiles were hypothesized to moderate the relationship between different factors and intention to engage with OGD (see Section 5.1.1). Therefore, respondents' demographic information that represents their profiles was collected. However, it is essential to note that demographic-related questions were not mandatory due to privacy preservation. Respondents can opt-out from answering the demographic questions. Different measurements were used to collect demographic data. For example, categorical scales (nominal) were applied in questions related to respondents' gender, nationality, education level, and work status. An interval scale was used to measure respondents' age, while an open-ended question was employed to elicit information on respondents' current jobs.

It is anticipated that respondents might consider that the questions asked in Section B (personal drivers) through Section F (behavioral intention) as not relevant. Therefore, a "not applicable" answer was provided beyond the given five-point Likert-scale-based answers to represent the choice. Thus, six choices of answers were provided in these sections of the survey: "strongly disagree," "disagree," "neither disagree nor agree," "agree," "strongly agree," and "not applicable."

5.2.2. Data collection

Before collecting survey data, researchers had to design the sampling frame to determine the study participants. The sampling frame addresses who will provide data, how they will be selected, and the number of participants needed to answer the research questions (Creswell & Plano Clark, 2018). The third research phase aims to test the hypotheses (see Section 5.1.1) and assess the research model developed from the systematic literature review and case studies using a larger sample of citizens. More specifically, quantitative data provided by open data users who have experience with OGD engagement were sought. In the OGD literature, it is not uncommon for researchers to not determine the sampling frame in survey-based studies due to the following reasons. First, there are no central, commonly agreed databases of OGD users. Second, the *population* of open data users is unknown. Although identifying the characteristics of the population is conceptually doable, no database about the community of OGD users exists (Beno et al., 2017). Therefore, determining the individuals and their numbers in the population is barely possible. Although open data user communities exist in many parts of the world (Kuk & Davies, 2011), the membership of the open data user community cannot be clearly defined because the openness nature of OGD leads to use by unanticipated actors (Martin, 2014). Users who have experience with OGD engagement may not join any communities and thus it is difficult to locate them. On the other hand, those who are members of a community may not have engaged with OGD. Moreover, those who have engaged with OGD and joined an open data user community may not be able or willing to participate in a survey.

As discussed above, designing a sampling frame for survey-based open data research focusing on OGD users' perspectives is not feasible. Therefore, it is common for such research to apply non-probability sampling techniques instead of probability sampling, which is the standard of quantitative-oriented studies (Teddlie & Tashakkori, 2009). Researchers define probability sampling as a sampling in which members of the population have been selected using a random selection so that the probability of each unit's inclusion can be computed (Bryman, 2012; Sue & Ritter, 2007). On the contrary, non-probability sampling generates "a sample that has not been selected using a random selection method" (Bryman, 2012, p. 187). Using non-probability sampling implies that some members of the population are likely to be selected than others. The use of online non-probability samples for social science research has exploded in recent decades and brought mostly positive consequences (Bryman, 2012; Coppock & McClellan, 2019). From a practical view, probability sampling significantly incurs more costs and requires more time than the non-probability approach (Bryman, 2012). Nevertheless, the use of non-probability sampling can be justified when the purpose of the research is exploratory (Bryman, 2012; Lehdonvirta, Oksanen, Räsänen, & Blank, 2020) or modeling the relationships between variables (Baker et al., 2013). Since this research phase aims to test and assess a model that predicts the determinants of citizens' intention to engage with OGD, the use of a non-probability survey can be warranted.

Several types of non-probability approaches for online surveys exist. These approaches include *river* and *panel* sampling. River sampling refers to recruiting respondents by inviting them to participate in a survey via a link placed on a web page, email, or somewhere else where members of the target population will likely notice (Lehdonvirta et al., 2020). Panel sampling relies on commercial online panel providers that assemble a group of individuals who volunteer to participate in future surveys (Lehdonvirta et al., 2020; Sue & Ritter, 2007). In the systematic literature review phase (see Chapter 3), nearly all survey-based open data research used these techniques. Examples of research that applies convenience sampling include Wang et al. (2019) (in a Chinese context), Wirtz et al. (2018) (Germany), Saxena and Janssen (2017) (India), Martin (2014) (the UK), and Zuiderwijk, Janssen, et al. (2015) (global). In comparison, those that utilize panel sampling include Weerakkody, Irani, et al. (2017) (the UK), and Jurisch et al. (2015) (global). Lehdonvirta et al. (2020) conducted an empirical comparison between non-probability samples and benchmark data using a subpopulation of cyber-harassment victims from a comprehensive population registry. They found no statistically significant differences between the river and panel samples when assessing the characteristics of the victims. This finding suggests that different non-random sampling approaches can yield similar results in statistical analysis. However, panel sampling is substantially more costly than river sampling. For instance, in 2009, Ramo, Hall, and Prochaska (2010) had to pay \$19.24 per participant who completed their surveys using a panel provider company named SSI. Amazon's MTurk, another panel provider company, charged approximately \$.55 per respondent per survey minute in 2010 (Berinsky, Huber, & Lenz, 2012).

In non-probability sampling, sample size cannot be estimated based on the variability in the underlying population because it is impossible to determine the likelihood of any particular participant being selected for the sample (Sue & Ritter, 2007). Many scholars have suggested some rules of thumb as a guideline for researchers conducting non-probability survey research. One of these rules suggests that in multivariate research such as ours, the sample size should be no less than ten times larger than the number of examined variables (Hill, 1998). Since this study utilizes the PLS-SEM technique for assessing the research model that predicts factors influencing citizen engagement with OGD, particular PLS-SEM guidelines (i.e., Hair et al., 2017) were followed to determine the acceptable sample size. Although identification issues with small sample sizes in PLS-SEM are inexistent, establishing a minimum sample size is relevant to safeguard that the approach results have sufficient statistical power (Hair et al., 2017). Furthermore, the minimum sample size will ensure the robustness of the results of the statistical method and the generalizability of the model (Hair et al., 2017, p. 23).

Hair et al. (2017) suggest two alternative guidelines for calculating the minimum sample size in PLS-SEM. First, researchers can use Cohen's (1992) rules of thumb to compute statistical power analysis for multiple regression models, considering that the model has a certain minimum quality of outer loadings. Second, built on Cohen's (1992) analyses, Hair et al. (2017, p. 26) created a sample size recommendation table to detect minimum values of the explained variance (R^2) of 0.10, 0.25, 0.50, and 0.75. The recommendation applies for 1%, 5%, and 10% significance levels, assuming 80% statistical power and a specific maximum number of independent variables. For example, given that the number of independent variables in the research model is five, more than 45 but less than 122 responses would be needed to attain an 80% statistical power to detect R^2 values of no less than 0.15 with a 5% error probability.

Alternatively, researchers can calculate the minimum sample size using computer programs like G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), built on Cohen (1988), which can compute power analyses specific to statistical analysis setups. The program can be downloaded from www.gpower.hhu.de. It provides five different power analyses: a priori power analysis, post hoc power analysis, compromise power analysis, sensitivity analysis, and criterion analysis (Faul et al., 2007). *A priori power analysis* is a method researchers can use to control statistical power before a study is conducted (Faul et al., 2007). It helps calculate the non-probability survey research sample sizes and controls the number of minimum completed responses targeted during data collection. In a priori power analyses, researchers calculate the sample size for multiple regression as a function of the effect size (proportion of explained variance), the prespecified significance level, the required power level, and number of predictors (Cohen, 1988). Applying similar values used in the first method (i.e., Hair et al., 2017), the results of the G*Power computation suggest that at least 91 observations would be needed. The minimum sample size can be simulated at a range of power levels, for example, from 80% to 95%, using the X – Y plot provided in the program. The plot showed that if the required level is increased to 95%, 138 observations would be needed (see Appendix F). Given this number, data collection progress was evaluated daily in this research phase to ensure that the minimum sample size (91 to 138 completed responses) can be achieved.

As suggested above, a non-probability sampling approach, particularly convenience sampling, was applied in this survey research. The sampling began with creating a list of possible communication channels the open data user community used to advertise the survey, including websites, conferences, social media platforms, and mailing lists. Table 5.2 provides an overview of communication channels used for advertising the survey. Having informed that the Kawal Pemilu's founders were involved in an initiative to engage with Indonesia's 2019 open presidential election results (Purbo et al., 2019), similar to that in 2014 (Purwanto et al., 2020b), the group was added to the list. An Indonesian survey translated from the original English version was developed to target such a group (see Appendix C). Groves et al. (2009) proposed translating surveys to languages understood by people with limited English skills to address the unit nonresponse due to the language barrier. Before distributing the survey links, the survey was piloted in each language with ten open data researchers from different backgrounds. The Indonesian-speaking researchers specifically tested the Indonesian questionnaire. During the piloting, they were provided with the English version to enable assessing the translated survey. Bryman (2012) suggested that the questions asked in the translated survey into other languages should be comparable with the English ones.

Table 5.2. The channel of communications used for advertising the surveys.

Channel	**Name**		**Description**	**Version**
Conference	International Conference on Electronic Participation 2019		Printed surveys (for those interested in filling out questions manually) and links (for those who wanted to fill in later) were distributed during the conference	English
Websites	The Dutch open government data portal (overheid.nl)		Two web pages in Dutch and English that advertised the survey link were hosted	English
Social media platforms	WhatsApp	Former Kawal Pemilu 2014 group	Messages that advertised the survey link were posted, followed by posting a reminder once a week (for two weeks)	Indonesian
		dg.o 2018 PhD colloquium group		English
	Facebook	Kawal Pemilu 2019 group	An advertisement about the survey link was posted by the group's admin, followed by sending private notifications to group members through messaging feature	Indonesian
	Twitter		Tweets that advertised the survey link in Dutch, English, and Indonesian were posted using hashtags and quoting open data provider accounts from different parts of the world or prominent open data figures	English, Indonesian
	LinkedIn	Open Data Research Network group	An advertisement about the survey link was posted, followed by posting a reminder once a week (for two weeks)	English
		Open Audit Tools		
		Open Data Nederland		
Mailing lists	Regional Open Knowledge Foundation Network (OKFN) – 31 mailing lists		An email advertising the survey link was sent to 31 mailing lists, followed by posting a reminder once a week (for two weeks)	English
	eGovernment (hosted by University of Washington)		An email advertising the survey link was sent to the mailing list, followed by posting a reminder once a week (for two weeks)	
	National Institute of Computer-Assisted Reporting (hosted by University of Missouri)			
	TU Delft open data research group			
	Indonesian Education Endowment Fund scholars			

International participants from different open data user communities across countries and cultures were also targeted. Although cultural differences may influence the survey analysis results, little or no empirical evidence supports such an assumption. Instead, previous research found that drivers and barriers of open-data-driven co-creation in countries with advanced open data ecosystems and those that are latecomers in adopting open data appear indistinguishable (Toots et al., 2017). Instead of advertising the questionnaires' long internet address, two shortened links that encapsulate the URL addresses of the English and Indonesian versions of the survey were created. One of the advantages of applying a link shortener includes the ability to count the unique click rate of the advertised surveys based on the IP address of the clickers. The number of these clicks can be treated as the *unique visitors* of the advertisement site. The survey links were distributed from May to September 2019 across different communication channels mentioned in Table 5.9. Sue and Ritter (2007) suggested using metrics other than *response rate* because researchers cannot compute the sampling frame in non-probability surveys. Eysenbach (2004) proposed three metrics that can replace response rate reporting in Internet-based surveys: view rate, participation rate, and completion rate. He defines *view rate* as the ratio between unique survey page visitors and unique site visitors (Eysenbach, 2012). Secondly, he computes the *participation rate* as the ratio between those who agreed to participate and unique first survey page visitors (Eysenbach, 2012). Lastly, the *completion rate* should be the ratio between the number of respondents who finished the survey and those who agreed to participate (Eysenbach, 2012). The completion rate of a non-probability survey is the single most critical and informative metric that informs about data quality (Callegaro & DiSogra, 2008; Liu & Wronski, 2018). This rate may indicate the respondents' interest in the survey (Callegaro & DiSogra, 2008). The view rate of the surveys was low, that only 21.20% of those who clicked the survey links actually visited the first page of the surveys.

In contrast, the total participation rate was substantially high, that nearly all visitors of the first page agreed to participate in the survey by answering the first question (98.26%). However, the completion rate was somewhat acceptable that 61.64% of those who agreed to participate finished the survey by clicking a *submit* button at its' end page. Table 5.3 gives an overview about the response metrics rate of each survey advertised through manual and online communication channels.

Table 5.3. The response metrics of the survey computed based on Eysenbach's (2012) checklist.

Survey Version	**English**	**Indonesian**
Manual		
Number of surveys distributed manually	20	
Number of surveys collected manually	4	
Completion rate of manual distribution	***20%***	
Online		
Number of unique clicks to survey links (unique site visitor)	1,489	1,500
Number of unique survey visitor	282	352
Number of participating respondents	280	343
Number of respondents who finished the survey	165	219
View rate	18.94%	23.47%
Total view rate		***21.20%***
Participation rate	99.29%	97.44%
Total participation rate		***98.26%***
Completion rate	58.93%	63.85%
Total completion rate of online surveys		***61.64%***

5.2.3. Data preparation

Before data analysis using the PLS-SEM technique, the survey responses should be prepared, cleaned, and transformed into usable data sets (Sue & Ritter, 2007). In the preparation stage, researchers code each answer to the questioned variable in the questionnaire into numerical data (Groves et al., 2009). The SPSS software version 25.0 was used to prepare the collected responses, and Ringle et al.'s (2015) SmartPLS software version 3.0 was used to analyze these responses. In the end, responses to the following seven questions were coded as follows.

1. The domain of OGD engagement (DOM) (see Table E.1 in Appendix D). Apart from the choices of OGD domains provided in the questionnaire, such as agriculture, climate, and business, it is also anticipated that alternative domains might exist. Therefore, an open-ended question was provided to give respondents an answer that best fits their frequently engaged domain. These answers were interpreted and classified into the available domains. For example, some participants responded with "election," coded into "government and management." If the classification of the answers into the available domains cannot be approximated, they were coded into "others." Overall, 32 responses that contained alternative domains were found. Twenty-nine among these responses were coded into the provided domains, while the rest were grouped into "other."
2. The output of OGD engagement (OUT) (see Table E.1 in Appendix D). Like the DOM question, an open-ended question was also provided for those who cannot find the suitable output of a respondent's engagement with OGD. The answer to this question was classified into an approximate

output. For instance, a respondent answered the question with "consultancy report," which was then categorized as "article." Twenty responses addressed alternative outputs of OG engagement. 17 out of 20 responses were classified into the provided outputs and the rest into "other."

3. Respondent's nationality (NAT) (see Table E.7 in Appendix D). Although 113 choices of nationality have been provided in this question, the probability of missing nationalities still exists. Therefore, an open-ended question was provided for respondents who could not find the correct answer from the provided choices to fill their nationalities manually. In the end, four new missing nationalities (i.e., "Azerbaijan," "Burkina Faso," "Kyrgyzstan," and "Tanzania") were created; one response was grouped to "English," and another one to "Other."
4. Respondent's current job (JOB) (see Table E.7 in Appendix D). Since this question is an open-ended one, codification of its answer becomes necessary. Therefore, a codification list containing 17 groups of jobs was developed, including an "other" category for answers that cannot be specifically grouped, such as "staff." One hundred seventy responses were coded according to the list.

Coding also involves reversing the scale of negatively worded questions (Sue & Ritter, 2007). The questionnaire asked three questions related to intrinsic and extrinsic motivations, i.e., INT2, INT5, and EXT2 (see Table E.3 in Appendix D). Initially, the answers to the Likert-scaled question were coded into ordered numbers with equidistant, i.e., 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), and 5 (strongly agree). However, the scale of the answers to these three questions was reversed by replacing 1 with 5, 2 with 4, 4 with 2, and 5 with 1. One response to the respondent's age was also corrected. Instead of writing her or his age, the respondent seemed to write her or his birth year, i.e., "1974". This value was replaced with "45".

The survey collects information about the attitudes of particular respondents who have engaged with OGD. However, there is a possibility of participation from those who do not have experience with OGD engagement. Therefore, a page displayed right after the introduction and consent of the survey was devised to filter out inexperienced respondents in OGD engagement and prevent them from continuing to participate in the survey (see EXP in Table E.1 in Appendix D). As a result, among 627 participating respondents from both manual and online surveys (see Table 5.3), 471 reported that they had experience engaging with OGD.

Five sections representing the research model under assessment constituted the second part of the questionnaire. Therefore, it was highly expected that the 471 respondents would indicate their evaluation of intrinsic and extrinsic motivations, technological factors, social factors, political factors, and behavioral intention to engage with OGD. However, nonresponse units exist in that some of the participants' responses contained missing values. Missing data is inevitable in research that utilizes survey methods (Little & Rubin, 2020). Problems with missing data are a fact of life in any statistical analysis, including SEM (Allison, 2003; Hair, Black, et al., 2014; Tabachnick & Fidell, 2019). The most common way to deal with missing data is by discarding an entire case when any value in the case is missing (i.e., listwise deletion or case-wise deletion). However, dropping cases as the only way to handle missing data is not recommended (Harrington, 2009). It can reduce the sample size available for analysis (Hair, Black, et al., 2014), leading to reduced statistical power (Harrington, 2009).

Hair, Black, et al. (2014) recommended classifying missing data and the reasons for missingness through a sequence of steps to determine the effects of the missing data and provide treatments for handling it in the analysis. In doing so, social science researchers usually refer to the work of Rubin (1976). The latter have rigorously set a solid foundation for defining three plausible assumptions about missing data mechanisms (Enders, 2013). In his seminal work, Rubin (1976) theorized that each survey participant has a hidden likelihood of missing data on a variable. This inclination for missingness might or might not have a connection with the variables in a specific analysis model. The assumption of *missing completely at random* (MCAR) is met when the probability of missing data on a variable, suppose Y, is unrelated to Y or the values of other variables in the data set. The *missing at random* (MAR) assumption holds when the probability that missing data on variable Y may depend on the value of another variable X but is unrelated to the value of Y when X is held constant. Finally, the assumption of *missing not at random* (MNAR) occurs when the probability of missingness depends on Y. If researchers can determine the mechanism of missingness as MCAR or MAR, then the nonresponse can be considered ignorable (Brown, 2015; Osborne, 2013; Schafer & Graham, 2002). On the contrary, when the missingness is MNAR, the data loss mechanism is deemed nonignorable (Brown, 2015; Kline, 2016).

The Collector software, used to develop the online survey, allows researchers to specify codes representing answers for skipped questions. Specifically, these answers were coded as “8888888” for the Likert-scale-based questions and “9999999” for the open-ended ones. Therefore, responses containing these values on questions directly related to the measurement of the research model (i.e., Section B, C, D, E, and F of the survey) are considered missing data. Among 471 responses gathered from the participating respondents who asserted that they have experience with OGD, 207 responses contain the defined missing data. Hair, Black, et al. (2014) advised researchers to use Little's (1988) test and investigate the missing data pattern using tools such as SPSS's Missing Value Analysis module to determine whether the missingness mechanism is MCAR. The test was applied to the 471 responses, and the results showed that the test had a significant level of 0.876. This value indicates that the observed missingness pattern does not significantly differ from a random pattern because the *p*-value of the test is more than 0.05. Therefore, the missingness mechanism in the 207 responses was deemed as MCAR. As a result, the missing data can be ignored and handled using a casewise deletion. In the end, 264 responses that did not contain missing data were included in the subsequent data preparation stage.

In addition to the Likert scale answers ranging from "strongly disagree" to "strongly agree" applied in the five primary sections of the survey (i.e., Section B, C, D, E, and F), a "not applicable" option was added to the answers (see Section 5.2.1). This inclusion allowed respondents to evaluate whether an indicator of a construct is relevant or not. Therefore, responses that contain the "not applicable" answers do not reflect the respondent's evaluation (agreement or disagreement) toward a construct's indicator. Since missing data is related to the survey questions' missingness, the responses mentioned above cannot be classified as missing. Instead, they should be analyzed separately from the sample. Among 264 responses, 98 of them have "not applicable" answers.

In contrast, 166 responses have complete answers. The latter responses were included in the following data preparation stage. These responses represent a 26.48% rate of survey completion based on the number of participating respondents (n=627). If the number of the completed responses (n=166) is compared with the minimum requirement of sample size between 91 and 138 (see Section 5.2.2), then the responses can be deemed fit. Table 5.4 depicts the data preparation stage concerning responses that have missing values and incomplete answers.

Although the survey's third section (i.e., Section G) is also essential to elicit the profiles of the respondents, it was designed as a non-mandatory section by providing an option not to answer the questions due to GDPR compliance. As a result, only respondents who gave consent to provide privacy-related data can answer the section entirely. Subsequently, missing values on the respondents' profiles are inevitable. For example, among 166 responses, 32 of them have missing data in Section G.

Table 5.4. The preparation stage of data analysis.

Missing/incomplete data preparation stage	Number of responses	
	Absolute	Pct.
Collected responses	627	100.00%
Stage 1: Experience with OGD (based on the collected responses)		
No experience	156	24.88%
Having experience	471	75.12%
Stage 2: Missing data on the tested variables (based on respondents who have OGD engagement experience)		
Having missing data	207	43.95%
Having complete data	264	56.05%
Stage 3: Relevance of the questions on the tested variables (based on responses with complete data)		
Having "not applicable" data	98	37.12%
Not having "not applicable" data	166	63.88%

Researchers' performance relies significantly on the goodwill of research participants with little or even no incentive to disburse effort in supplying data to researchers (Christensen, Johnson, & Turner, 2015). Therefore, some scholars assumed that data collection instrument such as survey is particularly susceptible to the operation among respondents, namely "response sets" (Bryman, 2012). Christensen et al. (2015) define a response set as a form of response bias used by people (consciously or otherwise) when responding to questionnaires. This bias is closely connected to multiple-indicator measures, where respondents answer an array of related questions that use a standard unit such as the Likert scale (Bryman, 2012). Two of the most critical types of response set are *acquiescence* (also recognized as the "yeasaying" and "naysaying" effect) and *social desirability*. Acquiescence is the respondents' tendency to consistently agree or disagree with a set of questions. The collected data were examined for the possibility of acquiescence bias by averaging the responses' scores and assessing responses whose average score is integers. Two responses with an average score of 4 were identified, but an acquiescence pattern was not found. Thus, these responses have varying answers. Social desirability refers to the respondents' tendency to conform to social norms by denying socially undesirable features and claiming socially desirable ones, and the tendency to "look good" to the researcher (Nederhof, 1985). Self-administered questionnaires that offer anonymity to respondents in which researchers are not present can effectively reduce social desirability bias. Since the respondents can administer the online questionnaire independently and be anonymous when participating in the survey, the social desirability bias of this research can be greatly reduced.

An online survey is also prone to multiple responses in which some respondents may mischievously complete the questionnaire multiple times (Sue & Ritter, 2007). The survey was designed to record the respondent's Internet Protocol (IP) address and browser agent and create cookies for respondents who agreed to prevent this issue. Respondents cannot complete the questionnaire more than once if the survey found the cookies the browser created in the previous response. However, respondents can still opt out of cookies to protect their privacy. This bias was also minimized by sending a specific link of the survey version to specific communication channels (see Table 5.2). Two responses with a similar IP address originating from 167.205.22.105, owned by "Asia Pacific Network Information Centre, Indonesia, Jawa Barat, Bandung," were found. Provided that the browser agents used are distinct, it was assumed that different respondents submitted these responses.

In addition, the researcher should examine the possibility of nonresponse bias occurrence (Groves et al., 2009). Nonresponse bias generally happens when a proportion of the target respondents did not participate in the survey (Groves et al., 2009; Urbach & Ahlemann, 2010). This bias results in an unreliable representation of the selected sample, mainly if the non-responders constitute a unique group. Researchers can evaluate nonresponse bias by comparing the responses between respondents and non-responder. However, this comparison cannot be conducted because it is not possible to collect data from non-responders. Armstrong and Overton (1977) proposed an approach to assess nonresponse bias by validating that the early and late responses do not differ significantly. This technique assumes that the late respondents are more likely similar to non-respondents than the early respondents (Armstrong & Overton, 1977). The online survey was advertised in four waves of time (i.e., 28th May 2019, 8th July 2019, 20th July 2019, and 29th July 2019). Reminders were sent in the second, third, and fourth advertisements. The fourth advertisement represented the last notifications sent to the target respondents. Different codes were assigned to the response waves accordingly, and the Kruskal-Wallis H test was applied to verify whether the four groups of responses were statistically different. The test results demonstrated that the distribution of the constructs across four waves of responses did not differ significantly at the 5% level. Therefore, the collected data were deemed highly unlikely threatened by nonresponse bias.

The data were prepared in sequential steps involving data coding, excluding irrelevant responses, identifying missing values, listwise deletion of responses containing missing values, and examining response bias and nonresponse bias. The number of responses was decreased mainly due to the exclusion of missing data. In the end, 166 responses that did not contain missing values were included in the data analysis stage using PLS-SEM.

5.3. Results of the PLS-SEM

This section reports the PLS-SEM analysis results that involve evaluating measurement models and structural models using model estimation. Measurement models concern the relationships between the indicators (the question items asked in the survey) and the constructs (the factors under study such as intrinsic motivations, extrinsic motivations in Section 5.1.1). The structural model deals with the relationships between the constructs. The measurement model results and the structural model assessments are presented in Section 5.3.2 and 5.3.3, respectively. Guidelines proposed by Hair et al. (2017), Hair, Sarstedt, Ringle, and Gudergan (2018), and Hair, Risher, Sarstedt, and Ringle (2019) were used to perform the assessments and report the results. Section 5.3.1 provides insights into the respondents' profiles and the general characteristics of their evaluation of the factors and their intention toward OGD engagement. In addition, three evaluations were presented at the end of this section. Section 5.3.4 discusses the moderation effects of respondents' profiles on the relationships between factors and intentions, while Section 5.3.5 reports the multigroup analysis that compares the research model based on the types of engagement (citizen-led and government-led). Finally, Section 5.3.6 reports and discusses the importance-performance map analysis to identify constructs and indicators with relatively high importance for citizens' behavioral intention to engage with OGD and relatively low performance.

5.3.1. Descriptive analysis

This section reports and discusses the descriptive analysis of the responses to all sets of questions asked in the survey instrument. This analysis provides a simple description of the collected data and forms the basis of the quantitative study. In addition, three analyses are reported and discussed in this section: the respondents' profiles (demographic characteristics), the respondents' engagement characteristics, and the respondents' attitudes and behavior toward OGD.

Table 5.5 provides an overview of the survey respondents' characteristics. The table's last column depicts the percentage of the validated sample comprising responses with completed answers only. Responses containing missing values were excluded from the calculation. Among the respondents who filled in the demographic questions, more than three-quarters were between 22 and 50 years old (n=114, 85.07%) with an average respondent's age of 40 years, and moderately more than two-thirds were men. The majority of the respondents (n=124, 92.54%) had a minimum higher education degree (i.e., bachelor's degree). Regarding the working status, 84.33% of the respondents were employees or freelance (self-employed). Academia (i.e., teacher, lecturer, and researcher) constituted the majority of the respondents' job roles (n=28, 24.78%), while other roles were pretty represented, such as IT-related (n=14, 12.39%) and entrepreneurial jobs (n=18, 15.93%). Slightly more than half of the respondents had Indonesian nationality (n=84; 62.69%). This composition introduced the question, namely, whether the perspectives of Indonesian respondents biased the response of the whole sample. The sample was evaluated by applying the Kruskal-Wallis H test to confirm whether there are significant differences between nationality groups on the determinant and behavioral intention variables asked in the survey. The results showed no statistically significant differences in the respondents' attitudes and intentions among different nationality groups.

Table 5.5. The survey respondents' profiles (n=166).

Characteristics	**Category**	**Sample**		
		N	**%**	**Valid (%)**
Gender	Female	42	25.30%	31.34%
	Male	89	53.61%	66.42%
	Prefer not to answer	3	1.81%	2.24%
	Unknown (missing)	32	19.28%	
Age	22-30 years old	15	9.04%	11.19%
	31-40 years old	60	36.14%	44.78%
	41-50 years old	39	23.49%	29.10%
	51-60 years old	15	9.04%	11.19%
	61 years old or over	5	3.01%	3.73%
	Unknown (missing)	32	19.28%	
Education	High school diploma	1	0.60%	0.75%
	College degree	2	1.20%	1.49%
	Vocational training	1	0.60%	0.75%
	Bachelor's degree	51	30.72%	38.06%
	Master's degree	62	37.35%	46.27%
	Professional degree	6	3.61%	4.48%
	Doctorate degree	11	6.63%	8.21%
	Unknown (missing)	32	19.28%	
Working status	Employed	91	54.82%	67.91%
	Self-employed/Freelance	22	13.25%	16.42%
	Unemployed – Looking for work	3	1.81%	2.24%
	Homemaker	1	0.60%	0.75%

Characteristics	Category	Sample		
		N	%	Valid (%)
Working status	Studying	12	7.23%	8.96%
	Retired	2	1.20%	1.49%
	Other	3	1.81%	2.24%
	Unknown (missing)	32	19.28%	
Job	Data scientist	5	3.01%	4.42%
	Software engineer	6	3.61%	5.31%
	IT architect	1	0.60%	0.88%
	Researcher	14	8.43%	12.39%
	Teacher	5	3.01%	4.42%
	Lecturer	9	5.42%	7.96%
	Journalist	6	3.61%	5.31%
	Manager	11	6.63%	9.73%
	Consultant	10	6.02%	8.85%
	Entrepreneur	8	4.82%	7.08%
	Freelance	1	0.60%	0.88%
	Professional	11	6.63%	9.73%
	IT support	2	1.20%	1.77%
	Editor	2	1.20%	1.77%
	Other	22	13.25%	19.47%
	Unknown (missing)	53	31.93%	
Nationality	African	2	1.20%	1.49%
	American	6	3.61%	4.48%
	Asian – Indonesian	84	50.60%	62.69%
	Asian – non-Indonesian	5	3.01%	3.73%
	Australian	3	1.81%	2.24%
	European	33	19.88%	24.63%
	Other	1	0.60%	0.75%
	Unknown (missing)	32	19.28%	

It is important to note that a non-probability sampling approach was used for this study because the database of OGD users is non-existent, and the population of open data users is unknown. Therefore, it is just possible to describe and establish the demographic representativeness of the samples. Detailed arguments about overcoming the issue of using the non-probability sampling approach have been provided in Section 5.2.2. However, based on the collected data, it can be inferred that the OGD users are empowered citizens (Gurstein, 2011), indicated by most respondents with a higher education degree. Unlike Zuiderwijk, Janssen, et al. (2015), which mainly focus on international social science researchers, this research did not collect data from citizens with a specified occupancy. The demographic representation of the samples is relatively similar to Jurisch et al.'s (2015) international samples: the majority of respondents' age ranges from twenty to fifty years old.

Table 5.6 describes the characteristics of the survey respondents' OGD engagement experience. Only slightly more than a quarter of the respondents had more than one year of experience engaging with OGD (n=43, 25.91%), suggesting that most respondents have a relatively limited experience with OGD use. On the other hand, more than half of the respondents frequently engaged in citizen-led OGD engagement settings (n=102, 61.45%). Three types of output that the respondents commonly created in the OGD engagement include visualizations (21.09%), applications (20.10%), and maps (17.37%). At the same time, the respondents primarily engaged with OGD produced in three areas: government and management (20.00%); society and social (10.42%); and business, economy, and finance domains (9.75%).

Table 5.6. The characteristics of the respondents' OGD engagement (n=166).

Characteristics	Category	Sample	
		N	%
Experience	Less than one year	123	74.10%
	1-2 years	20	12.05%
	2-5 years	16	9.64%
	More than five years	7	4.22%
Engagement type	Citizen-led	102	61.45%
	Government-led	64	38.55%
Engagement output	Application	81	20.10%
	Map	70	17.37%
	Visualization	85	21.09%
	Article	67	16.63%
	News	43	10.67%
	New database	46	11.41%
	Other	2	0.50%
	No product or service created	9	2.23%

Characteristics	Category	Sample	
		N	%
OGD domain	Agriculture	25	4.20%
	Care and health	27	4.54%
	Climate	27	4.54%
	Business, economy, and finance	58	9.75%
	Defense	10	1.68%
	Ecosystems, nature, and environment	39	6.55%
	Education, science, and research	52	8.74%
	Energy	22	3.70%
	Government and management	119	20.00%
	Housing	24	4.03%
	Industry and manufacturing	16	2.69%
	Infrastructure, space, and transportation	44	7.39%
	Maritime and ocean	14	2.35%
	Public order and safety	25	4.20%
	Society and social	62	10.42%
	Tourism	29	4.87%
	Other	2	0.34%

Table 5.7 provides an overview of the characteristics of the survey respondents' attitudes and behavioral intentions toward OGD engagement. As

far as the intrinsic motivations are concerned, enjoyment of learning and studying OGD indicator (INT6) appears highly motivating the respondents to engage with OGD (n=145, 87.35%). In addition, the enjoyment of experience with OGD (INT7) (n=142, 85.54%) and intellectual challenge (INT8) (n=139, 83.73%) indicators are also dominating intrinsic motivations that drive the respondents' engagement. Regarding the extrinsic motivations, personal benefits or relative advantage (EXT1) highly likely motivate the respondents to engage with OGD (n=157, 94.58%).

Concerning the technological factor indicators, the results seem mixed in different quality aspects. The availability (SYSQ1) (n=98, 59.04%) and the responsiveness (SYSQ2) (n=95, 57.23%) of OGD portals are two technological indicators that highly likely drive the respondents' engagement. On the contrary, the accuracy (DQ1) and the completeness (DQ2) of OGD appear to be the least influential technological indicators of the respondents' intention to engage with OGD. Most respondents disagreed that DQ1 (n=106, 63.86%) and DQ2 (n=95, 57.23%) drive their engagement; these two indicators seem to resemble the barriers to OGD engagement.

Concerning social factors, benefitting society (SOC4) highly likely influences most respondents to engage with OGD (n=153, 92.17%). As for the political factors, four indicators that represent trust in OGD (TR2 and TR3), political efficacy (POL1), and interests in politics (POL2) are likely to drive the respondents' engagement with OGD. These four indicators fairly share similar distribution of the respondents' evaluation on the political factors: TR2 (n=114, 68.67%); TR3 (n=111, 66.87%); POL1 (n=118, 71.08%); and POL2 (n=115, 69.28%). Lastly, regarding behavioral intentions, the results show that most of the respondents had intentions to engage with OGD in the future.

Table 5.7. The characteristics of the respondents' attitudes and behavioral intention toward OGD (n=166).

Indicator	**Strongly disagree**		**Disagree**		**Neither disagree nor agree**		**Agree**		**Strongly agree**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Intrinsic motivations (INT)										
INT1	1	0.60%	8	4.82%	20	12.05%	75	45.18%	62	37.35%
INT2 *)	11	6.63%	25	15.06%	49	29.52%	50	30.12%	31	18.67%
INT3	6	3.61%	36	21.69%	29	17.47%	68	40.96%	27	16.27%
INT4	2	1.20%	22	13.25%	30	18.07%	72	43.37%	40	24.10%
INT5 *)	9	5.42%	29	17.47%	48	28.92%	50	30.12%	30	18.07%
INT6	0	0.00%	0	0.00%	21	12.65%	77	46.39%	68	40.96%
INT7	0	0.00%	7	4.22%	17	10.24%	90	54.22%	52	31.33%
INT8	0	0.00%	6	3.61%	21	12.65%	91	54.82%	48	28.92%

Indicator	Strongly disagree		Disagree		Neither disagree nor agree		Agree		Strongly agree	
	N	%	N	%	N	%	N	%	N	%
Extrinsic motivations (EXT)										
EXT1	1	0.60%	2	1.20%	6	3.61%	73	43.98%	84	50.60%
EXT2 *)	17	10.24%	29	17.47%	30	18.07%	62	37.35%	28	16.87%
EXT3	4	2.41%	15	9.04%	30	18.07%	74	44.58%	43	25.90%
EXT4	10	6.02%	27	16.27%	62	37.35%	45	27.11%	22	13.25%
Data quality (DQ)										
DQ1	22	13.25%	84	50.60%	33	19.88%	23	13.86%	4	2.41%
DQ2	18	10.84%	77	46.39%	33	19.88%	29	17.47%	9	5.42%
DQ3	12	7.23%	48	28.92%	40	24.10%	51	30.72%	15	9.04%
DQ4	12	7.23%	58	34.94%	41	24.70%	42	25.30%	13	7.83%
System quality (SYSQ)										
SYSQ1	4	2.41%	39	23.49%	25	15.06%	74	44.58%	24	14.46%
SYSQ2	7	4.22%	32	19.28%	32	19.28%	76	45.78%	19	11.45%
SYSQ3	10	6.02%	49	29.52%	35	21.08%	53	31.93%	19	11.45%
SYSQ4	8	4.82%	35	21.08%	37	22.29%	69	41.57%	17	10.24%
Service quality (SERVQ)										
SERVQ1	8	4.82%	48	28.92%	38	22.89%	57	34.34%	15	9.04%
SERVQ2	13	7.83%	42	25.30%	46	27.71%	51	30.72%	14	8.43%
SERVQ3	7	4.22%	33	19.88%	45	27.11%	63	37.95%	18	10.84%
SERVQ4	17	10.24%	34	20.48%	57	34.34%	45	27.11%	13	7.83%
Social influence (SOC)										
SOC1	9	5.42%	27	16.27%	61	36.75%	52	31.33%	17	10.24%
SOC2	6	3.61%	19	11.45%	43	25.90%	66	39.76%	32	19.28%
SOC3	4	2.41%	16	9.64%	55	33.13%	71	42.77%	20	12.05%
SOC4	0	0.00%	1	0.60%	12	7.23%	68	40.96%	85	51.20%
Trust in OGD (TR)										
TR1	1	0.60%	15	9.04%	50	30.12%	74	44.58%	26	15.66%
TR2	0	0.00%	11	6.63%	41	24.70%	87	52.41%	27	16.27%
TR3	1	0.60%	9	5.42%	45	27.11%	85	51.20%	26	15.66%
Political participation (POL)										
POL1	2	1.20%	6	3.61%	40	24.10%	87	52.41%	31	18.67%
POL2	3	1.81%	15	9.04%	33	19.88%	71	42.77%	44	26.51%
POL3	13	7.83%	27	16.27%	48	28.92%	54	32.53%	24	14.46%
Behavioral intention to engage with OGD (BI)										
BI1	0	0.00%	1	0.60%	8	4.82%	88	53.01%	69	41.57%
BI2	0	0.00%	1	0.60%	11	6.63%	83	50.00%	71	42.77%
BI3	0	0.00%	1	0.60%	10	6.02%	89	53.61%	66	39.76%
BI4	1	0.60%	3	1.81%	16	9.64%	76	45.78%	70	42.17%

*) The indicator's scales are reversed (as explained in Section 5.2.3)

5.3.2. Measurement model assessment

This section reports and discusses the first assessment of the PLS-SEM model. The assessment concerns the relationships between the indicators and the constructs that they reflected. Figure 5.2 illustrates the initial PLS path model that reflects the research model (see Section 5.1.1) before the measurement model assessment. Appendix D provides a complete overview of the model's indicators, represented with square shapes and labeled with an abbreviation of the constructs followed by a number. As recommended in the literature, the model assessment concerns four evaluation metrics: indicator loadings, internal consistency reliability, convergent validity, and discriminant validity (Hair, Hult, Ringle, & Sarstedt, 2014). The following subsections describe the metrics definition and acceptable value and provide a general impression of the appearance of these values in the model. At the end of this section, a detailed assessment is reported. Indicators are also evaluated against the acceptable values of the metrics. Based on the evaluation results, keeping or dropping an indicator will depend on the explained variance yielded during the assessment. PLS-SEM aims to maximize the R^2 value (i.e., the explained variance) of the dependent variables in the PLS path model (Hair et al., 2017). Therefore, assessing the quality of the measurement (and structural) models focuses on the metrics that indicate the model's predictive capabilities.

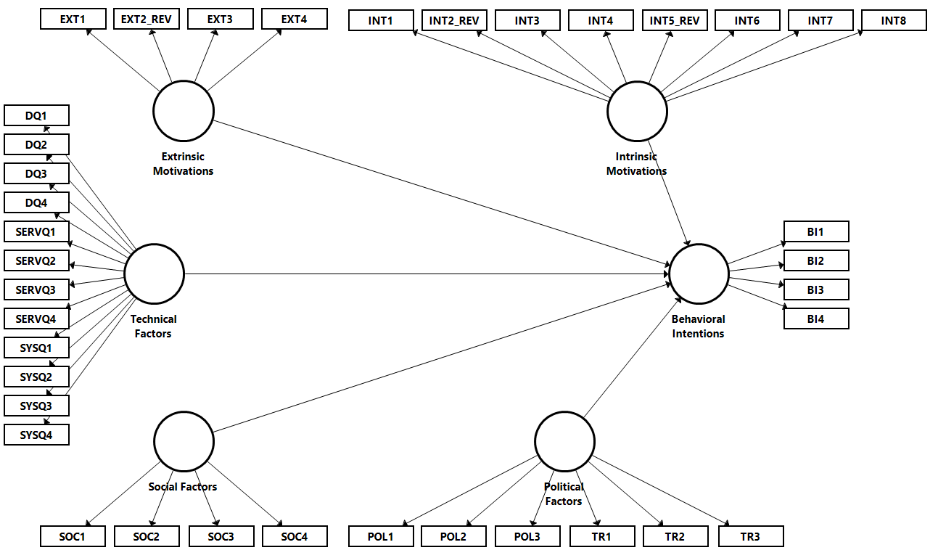


Figure 5.2. The initial path model of OGD citizen engagement, created using SmartPLS 3 (Ringle et al., 2015).

Indicator loadings

Indicator loading (or *outer loading*) refers to the total contribution of an indicator to the definition of its construct (Garson, 2016). The square of standardized outer loading portrays *communality*, which shows how much the construct explains the variation in the indicator (Hair et al., 2017). Researchers describe it as the variance derived from the indicator. Hair et al. (2019) recommend loadings above 0.708 because this value indicates that the construct explains higher than 50% of the indicator's variance. However, researchers frequently acquire weaker outer loadings (less than 0.70) in social science studies (Hulland, 1999). Therefore, rather than automatically dropping indicators whose outer loading is less than 0.70, Hair et al. (2017) suggest that researchers thoroughly examine the effects of indicator elimination on the composite reliability. An indicator with outer loading ranging from more than 0.40 to less than 0.70 should be deemed for elimination only when dropping the indicator causes an increase in the composite reliability that exceeds the threshold value (Hair et al., 2017).

Overall, the indicators' outer loadings appeared satisfactory. Most of the indicators had loadings above 0.60 (n=30; 78.95%). Among these indicators, twenty-one of them (55.26%) had outer loadings above 0.70. On the other hand, five indicators had loadings below or equaled 0.40, warranting their automatic removal from the model.

Internal consistency reliability

Internal consistency reliability can be defined as the degree of consistency with which multiple indicators on a research instrument measure a single construct (Christensen et al., 2015; Creswell & Creswell, 2018). Researchers should perform this test to evaluate whether the scale indicators measure the same construct and are highly intercorrelated (Bryman, 2012; Hair, Black, et al., 2014). If they are not, some of the indicators may not be related to the construct and, therefore, indicate something different (Bryman, 2012). The traditional criterion for internal consistency commonly reported in the literature is Cronbach's (1951) α or coefficient alpha (Henseler, Ringle, & Sinkovics, 2009; Kline, 2016). Higher values of α usually demonstrate higher levels of reliability (Hair et al., 2017), and α values higher than 0.70 indicate good reliability (Christensen et al., 2015; Creswell & Creswell, 2018; Field, 2009). Hair et al. (2017) suggest that values ranging from 0.60 to 0.70 are deemed acceptable in exploratory studies, while "satisfactory to good" values are between 0.70 and 0.90. However, reliability values of 0.95 and more are questionable because they demonstrate that the indicators are redundant (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012).

However, Cronbach's α is deemed a less precise measure (Hair et al., 2017). It usually underestimates scale reliability because the indicators are unweighted (Garson, 2016; Hair et al., 2017). Jöreskog (1971) proposed composite reliability to replace Cronbach's α (Garson, 2016). In addition, the results of composite reliability evaluation can be interpreted similarly as Cronbach's α (Henseler et al., 2009). The indicators are weighted based on their standardized loadings and measurement error (Hair et al., 2019; Henseler et al., 2009; Shook, Ketchen, Hult, & Kacmar, 2004). Subsequently, composite reliability is typically higher than Cronbach's α (Hair et al., 2019).

Nevertheless, researchers considered composite reliability and Cronbach's α tests at the extreme of the reliability continuum (Hair et al., 2017). Cronbach's α represents the lower bound of the continuum, while the composite reliability represents the continuum's upper bound (Hair et al., 2019). Alternatively, Dijkstra and Henseler (2015) developed ρ_A analysis as an approximately precise measure of internal consistency reliability. The results of ρ_A analysis usually lie between Cronbach's α and composite reliability (Hair et al., 2017). Therefore, researchers may regard ρ_A as a good compromise between the two measures. When analyzing the internal consistency reliability, researchers must evaluate whether it is significantly higher or lower than the suggested minimum or maximum thresholds (Hair et al., 2019).

As far as the data are concerned, two constructs had Cronbach's α values between 0.50 and 0.60, and one construct had an α value between 0.60 and 0.70, while three constructs had α values above 0.70. At the same time, five among the six constructs had above satisfactory ρ_A values above 0.70. Only one construct had a ρ_A value between 0.50 and 0.60. These results justified removing indicators that contributed to the lower values of Cronbach's α and ρ_A.

Convergent validity

Convergent validity refers to the extent to which a construct converges to explain its indicators' variance (Hair et al., 2019). Therefore, it indicates whether a group of indicators represents the same fundamental construct (Henseler et al., 2009). Fornell and Larcker (1981) recommend the average variance extracted (AVE) metric applied to all indicators on each construct to evaluate a construct's convergent validity. A construct's AVE is calculated by summing all of its indicators' squared loadings divided by the number of its indicators. Hence, AVE equals the communality of a construct (Hair et al., 2017). An AVE value of 0.50 or higher indicates that generally, a construct can explain at least half or more of its indicators' variance (Hair et al., 2017; Henseler et al., 2009).

On the one hand, three among six constructs of the model had AVE values higher than 0.50. On the other hand, three other constructs had AVE values below 0.50. As a result, the removal of problematic indicators that contributed to the latter AVE values is warranted.

Discriminant validity

Discriminant validity indicates the degree to which a given construct is empirically distinct and specific from other constructs in the same model (i.e., there are low correlations among the constructs) (Brown, 2015; Sarstedt & Mooi, 2019). Henseler, Ringle, and Sarstedt (2015) proposed using the heterotrait-monotrait (HTMT) ratio to assess discriminant validity. HTMT is the ratio between the intra-trait correlations and the within-trait correlations (Hair et al., 2017). The ratio is computed by comparing the mean of the indicator correlations across all constructs with the mean of the average correlations for all indicators that measure the same construct (Hair et al., 2019). When HTMT values are high, there may be discriminant validity problems. Henseler et al. (2015) proposed a threshold value of 0.85 for conceptually distinct constructs and 0.90 for conceptually very similar constructs. Values more than these thresholds suggest that the constructs are not distinct, and discriminant validity is not present. However, researchers should depend upon a bootstrapping procedure to determine the HTMT confidence interval, which represents the range of the true population value of HTMT, assuming a 95% level of confidence (Hair et al., 2017). Confidence interval that includes the value of 1 demonstrates a lacking discriminant validity. If, on the other hand, the value of 1 do not fall into the range, this indicates the empirical distinctiveness of the two constructs.

The HTMT values of the constructs appeared acceptable because all of them were below 0.85. However, the confidence interval computed by the bootstrapping procedure showed that the discriminant validity of the motivation constructs (both intrinsic and extrinsic) seemed problematic because it contained the value 1.

Model assessment

The statistical software application SmartPLS 3.0 (Ringle et al., 2015) is used to compute the PLS path model. The following settings were applied before executing the PLS algorithms: path weighting, 1000 maximum iterations, and 10^{-7} stop criterion. The path weighting scheme is the recommended default setting because it provides the highest R^2 value for endogenous latent indicators (Hair et al., 2017). 1000 was applied as the maximum number of iterations used for computing the PLS results because it is sufficiently large. At the same time, a 10^{-7} stop criterion value was also applied because it is the recommended setting of the SmartPLS software.

Figure 5.3 displays the results of the PLS algorithm in three different metrics for assessing the measurement model: the values of indicator loading, the path coefficient, and the explained variance R^2. The values that appeared on the arrow between constructs and indicators (e.g., between Extrinsic Motivations and EXT1) represent the indicator loading values (e.g., 0.773). At the same time, we can see that the arrows' values between the constructs (e.g., between Extrinsic Motivations and Behavioral Intentions) constitute the path coefficient values (e.g., 0.387). Lastly, the value that appeared on the endogenous construct (i.e., Behavioral Intentions) is the explained variance (R^2), i.e., 0.451.

Figure 5.3 also shows that non-technical factors positively correlate with behavioral intention, contrary to the technical factors that have a negative relationship (i.e., -0.137). Surprisingly, this result indicates that the lower quality of OGD, the higher the respondents' intention to engage with OGD. One possible explanation is that the respondents may have skills and knowledge needed or alternative ways such as consulting the data with the providers to figure out the problems related to OGD quality. Another possible explanation concerns the fact that low data quality results in societal problems that the respondents need to address. As a result, respondents keep engaging with OGD to expose these problems. The Kawal Pemilu case shows such an example that the citizens are motivated to find the low-quality election result data and be the first to post it on their social media accounts.

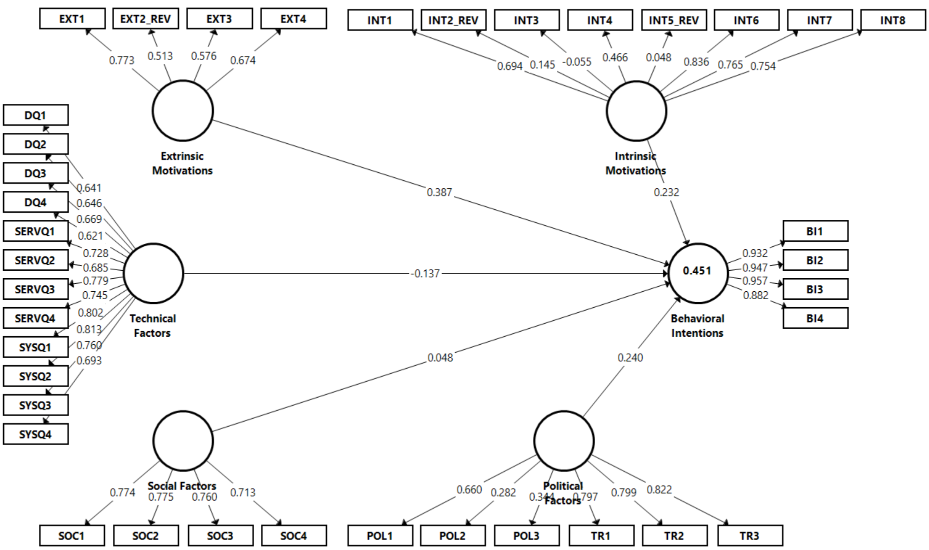


Figure 5.3. The results of the PLS-SEM analysis on the initial path model (R^2=0.451).

Table 5.8 below provides the summary of the measurement model assessment results and shows the following two indications. Firstly, the discriminant validity test indicates that two constructs, i.e., intrinsic motivations and extrinsic motivations, are not empirically distinct. The values of the HTMT confidence interval of these constructs include 1, i.e., 0.638 – 1.020 (see the previous Discriminant validity subsection for the explanation). The rightest column of Table 5.8 represents whether a construct has HTMT confidence interval values that do not include value 1. "Yes" represents the exclusion of value 1, while "No" constitutes the inclusion of value 1. It appears that these constructs should be merged into one construct, for example, motivations.

Table 5.8. The results summary for the initial OGD citizen engagement path model.

Construct	**Indicator**	**Loadings**	**Internal Consistency Reliability**		**Convergent Validity**		**Discriminant Validity**
			Cronbach's α	**Composite Reliability**	**Indicator Reliability**	**AVE**	**The HTMT confidence interval does not include 1**
Behavioral Intention	BI1	0.932	0.947	0.962	0.868	0.865	Yes
	BI2	0.947			0.896		
	BI3	0.957			0.915		
	BI4	0.882			0.778		
Intrinsic Motivations	INT1	0.694	0.563	0.711	0.481	0.322	No (with Extrinsic Motivations)
	INT2 *)	**0.145**			**0.021**		
	INT3	**-0.055**			**0.003**		
	INT4	0.466			0.218		
	INT5 *)	**0.048**			**0.002**		
	INT6	0.836			0.699		
	INT7	0.765			0.585		
	INT8	0.754			0.569		
Extrinsic Motivations	EXT1	0.773	0.524	0.732	0.597	0.412	No (with Intrinsic Motivations)
	EXT2 *)	0.513			0.263		
	EXT3	0.576			0.331		
	EXT4	0.674			0.454		
Technical Factors	DQ1	0.641	0.929	0.927	0.411	0.515	Yes
	DQ2	0.646			0.417		
	DQ3	0.669			0.447		
	DQ4	0.621			0.386		
	SERVQ1	0.728			0.529		
	SERVQ2	0.685			0.470		
	SERVQ3	0.779			0.606		
	SERVQ4	0.745			0.555		
	SYSQ1	0.802			0.643		
	SYSQ2	0.813			0.661		
	SYSQ3	0.760			0.577		
	SYSQ4	0.693			0.480		
Social Factors	SOC1	0.774	0.756	0.842	0.599	0.571	Yes
	SOC2	0.775			0.600		
	SOC3	0.760			0.578		
	SOC4	0.713			0.509		
Political Factors	POL1	0.660	0.694	0.801	0.436	0.431	Yes
	POL2	**0.282**			**0.080**		
	POL3	**0.344**			**0.118**		
	TR1	0.797			0.636		
	TR2	0.799			0.638		
	TR3	0.822			0.676		

*) The indicator's scales are reversed (as explained in Section 5.2.3)

Secondly, according to Hair et al. (2017), some indicators have significantly low outer loadings (below 0.40), i.e., INT2, INT3, INT5, POL2, and POL3 (bold printed in Table 5.8). INT2, INT3, and INT5 are items of the respondents' value system (Alathur, Ilavarasan, & Gupta, 2014) used to modify Jurisch et al.'s (2015) closeness of the topic construct. Their questionnaire asked the respondents to evaluate whether they will engage in open government online if the topic of interest is local or regional. Jurisch et al. (2015) found that the closeness of the topic significantly influenced the respondents' intention to use open government. Instead of adopting this construct, questions were designed to evaluate the compatibility of OGD engagement, assumed as an innovation, with the respondents' existing values. Political participation indicators, i.e., POL2 and POL3, were adopted from Jurisch et al. (2015). Jurisch et al. (2015) found that political participation significantly influenced the respondents' intention to use open government. However, the results show that INT2, INT3, INT5, POL2, and POL3 are not empirically reliable for measuring intrinsic motivations and political factors, respectively. The results also indicate that these indicators may be merged to measure different constructs, for example, value system construct (i.e., INT2, INT3, and INT5) and political participation construct (i.e., POL2 and POL3). Due to the unreliability of these indicators, they were dropped from the model.

Figure 5.4 illustrates the merging of motivation-based constructs and the removal of individual value systems and political participation indicators. Similar to Figure 5.3, this figure also presents the values of the outer loadings of all indicators, the path coefficient between constructs, and the explained variance of the Behavioral Intentions construct (i.e., R^2=0.418). Table 5.9 provides the summary of the assessment results on the modified model. The results show that two indicators, EXT2 and INT4, have loadings below 0.4 and, therefore, are not empirically reliable to measure the motivation construct. Therefore, EXT2 and INT4 were dropped from the modified model. EXT2 is related to the respondents' evaluation of whether their activities require them to engage with OGD. Based on the responses, it can be inferred that only slightly more than half of the respondents stated that engaging with OGD is required (see Table 5.7). This result does not provide reliable support for EXT2.

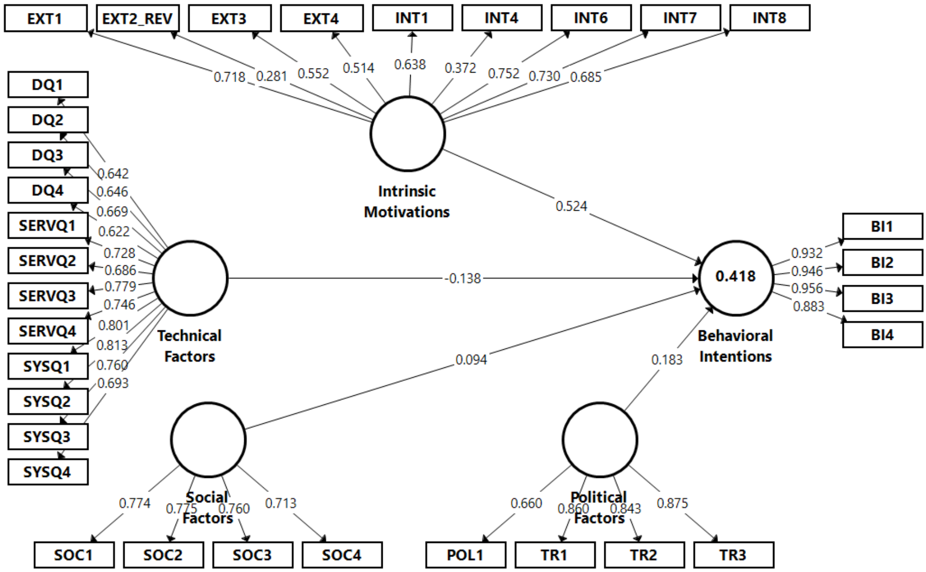


Figure 5.4. The overview of the modified model merges intrinsic motivations and extrinsic motivations constructs.

INT4 is part of the individual value system indicators (i.e., INT2, INT3, and INT5). The unreliability of these indicators shows that the respondents' value systems, such as beliefs, ideologies, and religions, do not influence their intention to engage with OGD. Although prior studies advised that the government can use OGD to oppress people who do not have power over their rights (Raman, 2012), generally, the respondents are agnostic toward the ideology that may fuel OGD. This finding contradicts previous research in the context of e-participation in India, in which citizens' dissatisfaction towards the government is influential (Alathur et al., 2014). The difference may be due to the type of data and information the citizens engage. In the e-participation context, citizens are highly likely consumers of the already interpreted data (or information) and feel urged to make decisions based on the interpretation of the data. In the OGD engagement context, contradictorily, respondents have to understand and interpret data into meaningful information or other outcomes before others can use it. The need to make ideology-related decisions heavily depends on the outcomes of an OGD engagement.

Table 5.9. The results summary for the modified OGD citizen engagement path model.

Construct	Indicator	Loadings	Internal Consistency Reliability		Convergent Validity		Discriminant Validity
			Cronbach's α	Composite Reliability	Indicator Reliability	AVE	The HTMT confidence interval does not include 1
Behavioral Intention	BI1	0.932	0.947	0.962	0.869	0.865	Yes
	BI2	0.946			0.895		
	BI3	0.956			0.915		
	BI4	0.883			0.780		
Motivations	EXT1	0.718	0.761	0.828	0.516	**0.364**	Yes
	EXT2 *)	**0.281**			**0.079**		
	EXT3	0.552			0.305		
	EXT4	0.514			0.264		
	INT1	0.638			0.407		
	INT4	**0.372**			**0.138**		
	INT6	0.752			0.566		
	INT7	0.730			0.533		
	INT8	0.685			0.470		
Technical Factors	DQ1	0.642	0.929	0.927	0.412	0.516	Yes
	DQ2	0.646			0.418		
	DQ3	0.669			0.448		
	DQ4	0.622			0.387		
	SERVQ1	0.728			0.530		
	SERVQ2	0.686			0.470		
	SERVQ3	0.779			0.607		
	SERVQ4	0.746			0.556		
	SYSQ1	0.801			0.642		
	SYSQ2	0.813			0.661		
	SYSQ3	0.760			0.577		
	SYSQ4	0.693			0.481		
Social Factors	SOC1	0.774	0.756	0.842	0.599	0.571	Yes
	SOC2	0.775			0.600		
	SOC3	0.760			0.578		
	SOC4	0.713			0.508		
Political Factors	POL1	0.660	0.828	0.886	0.435	0.663	Yes
	TR1	0.860			0.739		
	TR2	0.843			0.711		
	TR3	0.875			0.766		

*) The indicator's scales are reversed (as explained in Section 5.2.3)

Final path model

Even after EXT2 and INT4 indicators were removed from the modified model, the motivation constructs still suffer from the convergent validity problem with AVE values below 0.5 (AVE=0.459). The effects of dropping EXT3 and EXT4, which have lower loading values among motivation indicators, on the total explained variance (R^2) were compared. R^2 is higher when EXT4, which has the lowest loading value, is dropped. Therefore, EXT3 was retained in the final model. EXT3 indicates a motivation to know new people derived from Hutter et al.'s (2011) work on citizen engagement in open government platforms. This indicator played a role in the multiple case study (see Section 4.4). The EXT4 indicator is related to the respondents' evaluation of whether career concerns influence their intention to engage with OGD. This indicator was adopted from the open innovation studies in which innovators outside of a business entity participate in collaborative projects (Boudreau & Lakhani, 2009). Open data researchers have barely applied the indicator in the OGD context, and therefore, further research on the indicator is needed.

Furthermore, the removal of EXT2 and EXT4 shows that performance expectancy-related indicators do not influence the respondents' intention to engage with OGD. Respondents who stated that they are required to engage with OGD in activities are mostly researchers followed by managers and consultants, while OGD engagement influences those whose careers are primarily managers and consultants. Table 5.10 summarizes the measurement model assessment, while Figure 5.5 illustrates the final path model, whose structural model is assessed in the next section.

Table 5.10. The results summary for the final OGD citizen engagement path model.

Construct	**Indicator**	**Loadings**	**Internal Consistency Reliability**		**Convergent Validity**		**Discriminant Validity**
			Cronbach's α	**Composite Reliability**	**Indicator Reliability**	**AVE**	**The HTMT confidence interval does not include 1**
Behavioral Intention	BI1	0.933	0.947	0.962	0.870	0.864	Yes
	BI2	0.945			0.894		
	BI3	0.956			0.913		
	BI4	0.884			0.781		
Motivations	EXT1	0.724	0.804	0.858	0.525	0.504	Yes
	EXT3	0.555			0.308		
	INT1	0.679			0.462		
	INT6	0.787			0.619		
	INT7	0.759			0.576		
	INT8	0.732			0.536		
Technical Factors	DQ1	0.642	0.929	0.927	0.413	0.516	Yes
	DQ2	0.647			0.419		
	DQ3	0.670			0.448		
	DQ4	0.622			0.387		
	SERVQ1	0.728			0.531		
	SERVQ2	0.686			0.471		
	SERVQ3	0.779			0.607		
	SERVQ4	0.746			0.557		
	SYSQ1	0.801			0.641		
	SYSQ2	0.812			0.660		
	SYSQ3	0.760			0.577		
	SYSQ4	0.694			0.481		
Social Factors	SOC1	0.774	0.756	0.842	0.599	0.571	Yes
	SOC2	0.775			0.600		
	SOC3	0.760			0.578		
	SOC4	0.713			0.508		
Political Factors	POL1	0.660	0.828	0.886	0.435	0.663	Yes
	TR1	0.860			0.739		
	TR2	0.843			0.711		
	TR3	0.875			0.766		

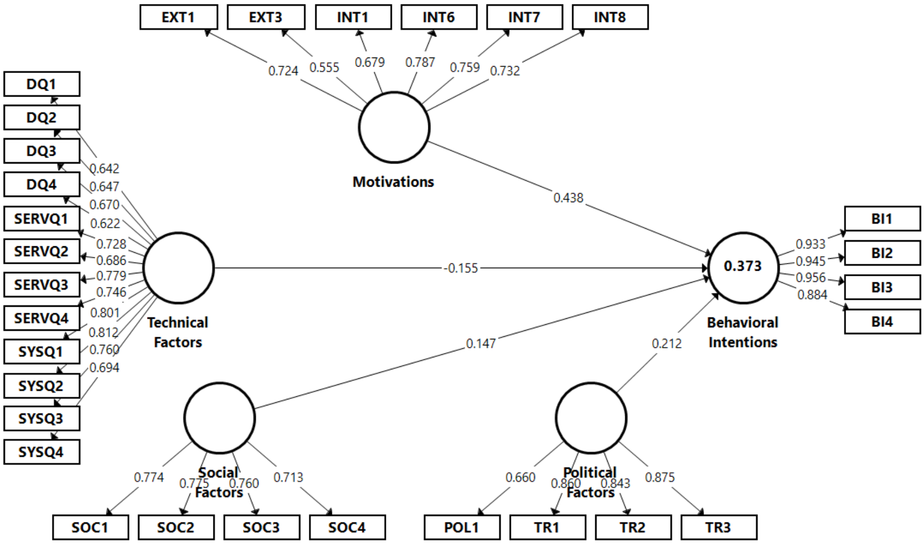


Figure 5.5. The final PLS path model after the measurement assessment (R^2=0.373).

5.3.3. Structural model assessment

This section reports and discusses the structural model assessment of the final PLS-SEM model. The relationships between the constructs were evaluated after undergoing measurement model assessment presented in Section 5.3.3. The researchers should validate their assumption that the construct indicators are reliable and valid in the measurement model assessment before conducting the structural model assessment. The researcher followed Hair et al.'s (2017) seven evaluation stages to assess the structural model. The stages comprise assessments of collinearity, significance, and relevance of relationships, level of R^2, f^2 effect size, predictive relevance Q^2, q^2 effect size, and predictive validity. The results of these assessments provide insights into the predictive capabilities of the proposed OGD citizen engagement model, which will be reflected upon in the last part of this section. The following sections discuss the seven stages of assessments.

Collinearity assessment

Collinearity is the extent to which (independent) variables (or indicators) are correlated (Gefen et al., 2000). The statistical assumption that the indicators are genuinely independent of each other is questionable when excessively high collinearity (namely, multicollinearity) exists. Multicollinearity occurs when two or more indicators strongly correlate in a regression model (Field, 2009). Researchers have to measure the *variance inflation factor* (VIF) to evaluate the level of collinearity (Hair et al., 2017). VIF refers to reciprocal tolerance.

Researchers define tolerance as the amount of variance of one indicator not explained by all other indicators of the same construct. A VIF value of 5 and higher is likely an indication of a collinearity problem (Hair et al., 2011). For example, an indicator's VIF value of 5 shows that 80% of its variance is explained by the remaining indicators related to the same construct (Field, 2009). Therefore, scholars recommend eliminating one of the corresponding indicators if a very high level of collinearity exists (Hair et al., 2011).

Before assessing the structural model, the final model's indicators' level of collinearity was checked. The VIF level of two indicators constituting the behavioral intention construct, i.e., BI2 and BI3, is higher than 5, i.e., 6.420 and 7.307, respectively. These values indicate that these indicators have high collinearity and represent similar measures, and therefore, one of them may be removed. BI2 relates to the respondents' fixed intention to engage with OGD, while BI3 concerns their anticipation of engaging with OGD in the near future. The researcher systematically experimented by eliminating one of the indicators at a time and compared the R^2 values. Removing BI2 led to a higher value of R^2, i.e., 0.385, and therefore, BI3 was retained in the final model illustrated in Figure 5.6. This figure displays the results of the PLS algorithm in three different metrics: the outer loading values of all indicators, the path coefficient between constructs, and the explained variance of the Behavioral Intentions construct.

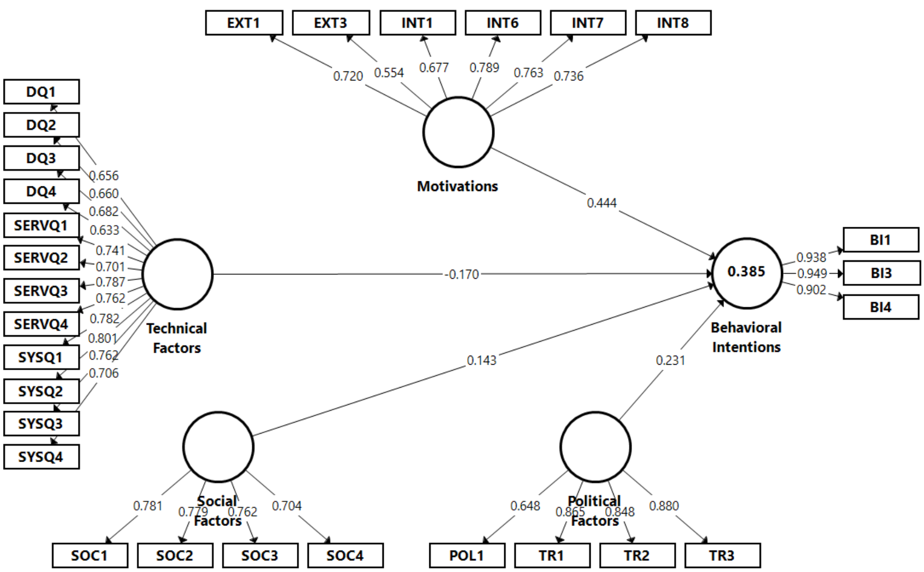


Figure 5.6. The final PLS path model after the collinearity problem has been addressed (R^2=0.385).

Path coefficients' significance assessment

The path coefficients are estimates that represent the hypothesized relationships among the constructs in a research model, i.e., the structural model (Hair et al., 2017). The coefficients' standardized values usually range from -1 to +1. Values very nearly to +1 or -1 constitute strong positive or negative relationships respectively and are typically statistically significant. Conversely, values closer to 0 usually represent weaker relationships; values close to 0 are typically not statistically significant.

Researchers rely on a nonparametric bootstrapping procedure (Davison & Hinkley, 1997) for assessing the path coefficients' significance (Hair et al., 2017; Hair et al., 2011). Bootstrap is fundamentally a method for replicating data sets utilizing resampling from the original data (Chernick, 2008; Davison & Hinkley, 1997). Bootstrap samples are a copious number of samples extracted from the original sample with replacement over which each time a case (observation) is drawn at random (Hair et al., 2017). Hair et al. (2011) recommended at least 5,000 bootstrap samples with the minimum number of cases equal to the number of observations in the original data set to run the bootstrapping algorithm in SmartPLS. The 5,000 bootstrap samples mean that 5,000 path models are estimated (Hair et al., 2017).

Table 5.11 summarizes the structural model's path coefficients' significance assessment using a bootstrapping procedure. The table's second column represents the path coefficients, indicating the strength and direction of the relationships between constructs. The decision of whether a path coefficient is significant relies heavily on its *t*-test result, significance level (i.e., *p*-value), and confidence interval values. The table's third, fourth and fifth columns represent these values, respectively.

Table 5.11. The results summary for the path coefficients' significance assessment.

Relationship	**Path coefficient**	***t* Value**	***p*-Value**	**95% Bca Confidence Interval**	**Significance ($p < 0.05$)?**
Motivations → Behavioral	**0.444**	**5.980**	**0.000**	**[0.303, 0.597]**	**Yes**
Political Factors → Behavioral	**0.231**	**2.631**	**0.009**	**[0.081, 0.424]**	**Yes**
Social Factors → Behavioral Intentions	0.143	1.846	0.065	[-0.015, 0.292]	No
Technical Factors → Behavioral Intentions	-0.170	1.284	0.199	[-0.456, 0.006]	No

The critical Student's *t* values for significance assessment (two-tailed test) using the normal (Gaussian) quantiles were determined as follows: 1.65 (significance level of 10%; $\alpha = 0.10$), 1.96 (significance level of 5%; $\alpha = 0.05$), and 2.57 (significance level of 1%; $\alpha = 0.01$) (Hair et al., 2011). The path coefficients' significance levels were assessed using *p* values representing the probability of incorrectly rejecting a true null hypothesis (Hair et al., 2017). Assuming a significance level of 5%, one can infer that the relationships under consideration are significant when the *p* values are smaller than 0.05. The bootstrap confidence interval can also be used to assess whether path coefficients are significantly different from zero. Hair et al. (2017) recommended using Efron and Tibshirani's (1986) Bca (Bias-Corrected and Accelerated Bootstrap) approach. The confidence interval gives information about the stability of the estimated coefficients by providing a range of possible population values based on the variation in the data and the sample size. It can be assumed that an estimated path coefficient significantly affects if the confidence interval does not contain zero.

Assuming a 5% significance level (Hair et al., 2017), barely all relationships in the structural model, except Social Factors → Behavioral Intentions and Technical Factors → Behavioral Intentions, are significant. The assessment results show that these relationships, Motivations → Behavioral Intentions, and Political Factors → Behavioral Intentions, have *t* values more than 1.96 (significance level of 5%). The *p* values of these relationships are also smaller than 0.05. The results also show that these relationships' bootstrapping confidence intervals, i.e., between 0.303 and 0.597 for Motivations → Behavioral Intentions and between 0.081 and 0.424, do not include zero. When a study is exploratory, researchers typically assume a significance level of 10%. Given this assumption, it can be considered that the relationship Social Factors → Behavioral Intentions path coefficient is significant because its *t* value is more than 1.65, and the *p*-value is smaller than 0.10. However, its bootstrap confidence interval, i.e., between -0.015 and 0.292, includes zero. At the same time, the results of the significance assessment on Technical Factors → Behavioral Intentions relationship show that its *t* value is less than 1.96, the *p*-value is larger than 0.05, and bootstrapping confidence interval between -0.456 and 0.006 includes zero. Therefore, the researcher concludes that Social Factors → Behavioral Intentions and Technical Factors → Behavioral Intentions relationships are not significant. As a result, both Social Factors and Technical Factors should be removed from the model.

Contrary to the previous quantitative OGD research (e.g., Saxena & Janssen, 2017; Zuiderwijk, Janssen, et al., 2015), the researcher found that social factors do not significantly influence the respondents' intention to engage with

OGD. This finding means that the respondents' social relationships with their friends, colleagues, communities, and society do not stimulate them to engage with OGD. The difference in the findings may be due to the voluntariness of OGD engagement in the studies. Saxena and Janssen (2017) and Zuiderwijk, Janssen, et al. (2015) found that voluntariness significantly negatively influences respondents' intention to use OGD and open data technology. The persons influential to the respondents likely made open data use a mandatory part of their daily activities. Saxena and Janssen's (2017) samples mainly constituted students, faculty, and bureaucrats, while Zuiderwijk, Janssen, et al. (2015) were social science researchers. Teachers and lecturers can combine open data in their courses by introducing and reinforcing using tools and techniques for open data processing. For example, the Master of Public Administration program at the University at Albany has included a mandatory course that involves open data analysis (Gascó-Hernández et al., 2018). Lecturers and researchers' supervisors, colleagues, and research groups usually influence them to re-use open data (Zuiderwijk & Spiers, 2019). On the contrary, IT professionals, non-IT professionals, academia, and others with a small part of journalism-related jobs (see Table 5.5) fairly represent the respondents' occupations. Furthermore, only slightly more than half of the respondents (n=90; 54.22%) stated that their activities require them to engage with OGD (EXT2).

In contrast to previous research (e.g., Talukder et al., 2019), the researcher discovered that technical factors, i.e., the quality of OGD constituted of data quality, system quality, and service quality, do not significantly influence the respondents' intention to engage with OGD. There exists a possibility that OGD quality may indirectly affect citizens' intentions through one or more mediating variables. Fitriani et al. (2019) found that information quality indirectly influences users' continuing intention to use open data websites through the variable of trust to open data websites. In the preceding section, based on latent variable correlations analysis results, the researcher proposes an alternative model over which the relationship between the OGD quality and behavioral intention is mediated through political factors.

Level of R^2 assessment

Researchers typically use the coefficient of determination or R^2 value to evaluate a structural model (Hair et al., 2017). R^2 measures the model's predictive power and represents "the amount of variance in the endogenous constructs explained by all of the exogenous constructs linked to it" (Hair et al., 2017, p. 198). Researchers also consider R^2 as a metric of in-sample predictive power (Rigdon, 2012). Its value varies from 0 to 1. Higher values indicate higher levels of predictive accuracy (Hair et al., 2017). The recommended R^2

values usually depend on the research discipline. For example, in the international marketing field, R^2 values of 0.75, 0.50, or 0.25 for endogenous latent variables can be respectively considered as substantial, moderate, or weak (Henseler et al., 2009). At the same time, in the information system research domain, Chin (1998b) suggest that R^2 values of 0.67, 0.33, and 0.19 indicate substantial, moderate, and weak level, respectively. However, social science researchers (e.g., Achen, 1982; King, 1986) advise caution in interpreting the R^2 value in terms of its strength and drawing inferences. Models with low R^2 value can still produce outstanding goodness of fit (Chin, 1998a).

The model's R^2 value is 0.3851 (see Figure 5.6), suggesting that 38.51% of the variance in the endogenous construct, i.e., the respondents' intention to engage with OGD, can be explained by the exogenous constructs (i.e., the factors). According to Chin (1998b), the model's R^2 value can be deemed as moderate. Previous quantitative open data studies on the intention to use and accept open data-related technologies show a varying level of R^2 values, ranging from 0.380 (Jurisch et al., 2015) to 0.580 (Weerakkody, Kapoor, et al., 2017). One noticeable outlier is Saxena and Janssen's (2017) study that yields an R^2 value of 0.940. Compared to these studies, it is justifiable to argue that the research model yields an acceptable value of R^2 in predicting the intention to engage with OGD.

F^2 effect size assessment

Researchers can use the change in the R^2 value when they exclude a particular exogenous construct from the model to assess whether the exclusion has a considerable impact on the endogenous constructs (Hair et al., 2017). This metric is denoted as the f^2 effect size and computed by estimating the PLS path model twice. The estimation is performed firstly with the exogenous construct included and then without the construct. The rule of thumb for assessing the f^2 values is that values of 0.02, 0.15, and 0.35 indicate small, medium, and large effects, respectively (Cohen, 1988). The values of effect size that are less than 0.02 represent no effect (Hair et al., 2017).

Table 5.12 shows the effect size of each exogenous construct on the intention to engage with the OGD construct. Based on these results, it can be inferred that the small impacts of political factors (f^2=0.051), social factors (f^2=0.022), and technical factors (f^2=0.032) on the behavioral intention are not statistically significant. In contrast, the medium effect of motivations on behavioral intention has statistical significance (f^2=0.196; p=0.014). Subsequently, these results reveal that extrinsic and intrinsic motivations have the most significant impact on the behavioral intention in the research model compared to other factors.

Table 5.12. The f^2 effect size of the exogenous constructs on the intention to engage with the OGD construct.

Relationships	f^2 value	t Value	p-Value	95% Bca Confidence Interval	Significance ($p < 0.05$)?
Motivations → Behavioral Intentions	**0.196**	**2.468**	**0.014**	**[0.163, 0.163]**	**Yes**
Political Factors → Behavioral Intentions	0.051	1.357	0.175	[-0.073, 0.070]	No
Social Factors → Behavioral Intentions	0.022	0.756	0.449	[-0.114, 0.042]	No
Technical Factors → Behavioral Intentions	0.032	0.557	0.578	[-0.084, 0.335]	No

Predictive relevance assessment

Furthermore, researchers have to also investigate Stone-Geisser's Q^2 (Geisser, 1974; Stone, 1974) value (Hair et al., 2017). Q^2 metric combines aspects of the model's out-of-sample predictive power and in-sample explanatory power (Shmueli, Ray, Estrada, & Chatla, 2016). Q^2 value can be obtained using the blindfolding algorithm with a particular omission distance D (Hair et al., 2017). Blindfolding involves an iterative process of omitting each dth data point in the endogenous construct's indicators and model re-estimation based on the remaining data points (Chin, 1998b; Tenenhaus, Vinzi, Chatelin, & Lauro, 2005). The blindfolding procedure enables researchers to evaluate the original and estimated values (Hair et al., 2017). The closer the prediction values are to the original ones, the higher the predictive accuracy of the path model (Chin, 1998b). According to Hair et al. (2017), a construct's Q^2 values larger than zero indicate that the model has predictive relevance, while values equal to zero and below suggest a lack of predictive relevance. Furthermore, Hair et al. (2017) suggest using the cross-validate redundancy approach to compute the Q^2 values. This approach develops the path model estimates of both the scores of the predictor constructs (structural model) and target endogenous construct (measurement model) (Hair et al., 2017).

The results of the blindfolding algorithm show that the model's Q^2 value is 0.314. Hair et al. (2019) suggest that a Q^2 value higher than 0, 0.25, and 0.50 indicates the model's small, medium, and large predictive accuracy, respectively. Thus, the model's Q^2 value signals that the research model has a predictive relevance with a medium level of accuracy. Moreover, the value is used to compute the exogenous constructs' level of predictive relevance, i.e., q^2 effect size.

Q^2 effect size assessment

While researchers assess the impact of the R^2 value using the f^2 effect size, the relative impact of relationships between exogenous and endogenous constructs (the q^2 effect size) can be computed by the difference of the Q^2 values. Hair et al. (2017) formulate the q^2 effect size as follows:

$$q^2 = \frac{Q^2_{included} - Q^2_{excluded}}{1 - Q^2_{included}}$$

Hair et al. (2017) advocate that an exogenous construct's q^2 values of 0.02, 0.15, and 0.35 demonstrate a small, medium, or large predictive relevance, respectively, for a particular endogenous construct.

Table 5.13 provides an overview of the exogenous constructs' q^2 computation. These results show that social factors lack predictive relevance on behavioral intention while political and technical factors have a small predictive relevance. In contrast, motivations have a medium predictive relevance on the behavioral intention to engage with OGD. This finding, once more, indicates the importance of the intrinsic and extrinsic motivations in driving the respondents' intention.

Table 5.13. The q^2 effect size of the exogenous constructs on the intention to engage with the OGD construct.

Relationships	**q^2 value**	**Level of relevance**
Motivations → Behavioral Intentions	**0.152**	**Medium**
Political Factors → Behavioral Intentions	**0.035**	**Small**
Social Factors → Behavioral Intentions	0.013	-
Technical Factors → Behavioral Intentions	**0.028**	**Small**

Predictive validity assessment

A research model's predictive power (or accuracy) concerns its ability to generate accurate predictions of either new temporally or cross-sectionally interpreted observations (Shmueli & Koppius, 2011). Predictive validity indicates the relationships between measures and constructs over which a provided set of measures for a particular construct can predict a given outcome indicator (Straub, Boudreau, & Gefen, 2004). The R^2 value merely assesses a model's in-sample explanatory power without indicating its out-of-sample predictive power, i.e., the ability to predict an outcome from new observations not included in the estimation procedures (Shmueli et al., 2019). At the same time, the Q^2 value combines in-sample and out-of-sample prediction, but its computation does not make use of holdout samples (Sarstedt, Ringle, & Hair, 2017). Shmueli et al. (2016) invented a holdout-sample-based algorithm, namely PLSpredict, that produces observation-level predictions on an indicator or a construct level. In contrast to the R^2 and Q^2 metrics, PLSpredict offers a method to assess a model's out-of-sample predictive power (Shmueli et al., 2019).

PLSpredict performs *k*-fold cross-validation. Shmueli et al. (2016) define a fold as a group of the total sample, while *k* refers to the number of the sub-sample groups into which the sample data are randomly equally split (Shmueli et al., 2016). PLSpredict merges *k*-1 subsets into a single sample to predict the remaining *k*th data subset. Shmueli et al. (2016) label the predicted subset as *the holdout sample*. The cross-validation procedure is reiterated *k* instances in which each subset is employed as the holdout sample. Subsequently, individual observation (case) in each holdout sample has a predicted value determined using a sample in which the observation was not included to calculate the model parameters. Shmueli et al. (2019) propose applying *k* value of 10 or another value inasmuch as each fold's sample meets the minimum required sample size.

Researchers should first identify their model's key target construct when assessing its predictive power (Shmueli et al., 2019). Then, they should interpret the $Q^2_{predict}$ statistic to ensure that the PLS predictions outperform alternative benchmark predictions that use a linear regression model (LM) (Evermann & Tate, 2016). A positive $Q^2_{predict}$ value indicates that the model's predictive relevance is confirmed while a negative one signals that the model lacks predictive power (Shmueli et al., 2016). The researchers then assess the degree of prediction error based on its distribution (Shmueli et al., 2019). The researchers must use the root mean squared error (RMSE) if they deem the error highly symmetrically distributed. On the contrary, researchers should use mean absolute error (MAE) when the predicted error's distribution is greatly non-symmetric. Next, the researchers should compare each indicator's RMSE (or MAE) value with the LM value (Hair et al., 2019). They then assess the PLS analysis-LM RMSE (or MAE) value comparison. Lower PLS' RMSE (or MAE) values for every indicator demonstrate a high predictive power while lower values for the majority of the indicators suggest a medium predictive power (Shmueli et al., 2019). At the same time, lower values for a minority of the indicators signal a low predictive power. Lastly, the researchers should evaluate the distribution of the PLS prediction errors: a left-tailed distribution signals over-prediction; a right-tailed distribution indicates under-prediction.

The default recommended settings for the PLSpredict procedure execution using the SmartPLS software (Ringle et al., 2015) were applied. The number of folds equals ten, and the number of repetitions equals ten as well (Shmueli et al., 2016). Since the partitioning of the data is random, the outcomes of the algorithm execution may vary at different points in time. Therefore, the execution was repeated ten times, and the average of the outcomes was computed to ensure a more stable estimate of the predictive performance. Table 5.14 provides the averaged outcomes of the PLSpredict execution. At the construct level, the $Q^2_{predict}$ value, i.e., 0.351 confirms the model's predictive relevance. Based on the indicators' (i.e., BI1, BI3, and BI4) RMSE values comparison between those obtained in PLS analysis and LM, it can be inferred that all indicators' PLS RMSE values are lower than their LM RMSE values. Thus, it is safe to suggest that the model has high predictive power.

Table 5.14. The PLSpredict assessment results.

Construct Prediction Summary						
	$Q^2_{predict}$					
Behavioral Intentions	0.351					
Indicator Prediction Summary						
	PLS			**LM**		
	RMSE	MAE	$Q^2_{predict}$	RMSE	MAE	$Q^2_{predict}$
BI4	0.646	0.487	0.279	0.717	0.531	0.112
BI3	0.522	0.421	0.283	0.549	0.432	0.207
BI1	0.494	0.406	0.335	0.538	0.432	0.208

Reflections on the structural model assessment

Overall, the structural model assessment shows that the model has high predictive quality. However, it also reveals that two factors, namely social factors, and technical factors, do not significantly affect the citizens' intention to engage with OGD. Conversely, (extrinsic and intrinsic) motivations and political factors significantly influence the citizens' intention to engage with OGD. Furthermore, the assessment results indicate that (extrinsic and intrinsic) motivation has the most significant impact on the behavioral intention in the research model, compared to other factors. Lastly, three factors, namely (extrinsic and intrinsic) motivations, political factors, and technical factors, are relevant to predict the citizens' intention to engage with OGD. Although technical factors do not significantly influence the intention, the latter findings indicate a plausible alternative model that can justify the inclusion of technical factors.

5.3.4. Effects of citizens' profiles as moderator

This section reports the moderating effects of citizens' profiles, i.e., age, educational level, gender, and experience, on the relationships between intrinsic motivations, technical factors, social factors, political factors, and citizens' behavioral intention. A profile comprises four indicators with particular measurements (i.e., continuous data for age, ordinal data for educational and experience level, and categorical data for gender). Combining these indicators as a composite measure for a citizen's profile offers a meaningless moderating relationship and difficulty examining each measurement's effects. Four moderating variables were proposed: age, educational level, gender, and experience level. These variables are measured using a single item indicator. To simulate the moderation relationship, sixteen models in which each moderator interacts with each factor influencing the citizens' intention to engage with OGD were created, as Henseler and Chin (2010) suggested. In the end, the effects of the moderator variables in the sixteen individual models were also assessed.

The primary objective of this research phase is to determine whether or not the indicators of citizens' profiles significantly affect the relationship between the factors (i.e., motivational, political, social, and technical factors) and the intention to engage with OGD. Chin, Marcolin, and Newsted (2003) proposed the two-stage approach for this particular objective when running a moderation analysis. The approach consists of the estimation of the main effects model without the interaction term (Stage 1) and the creation of a single-item measure generated with a multiplication between the scores of exogenous and moderator variables (Stage 2) (Chin et al., 2003).

Researchers should pay particular attention to the f^2 effect size of the interaction (moderation) effect when analyzing the moderation (Hair et al., 2017). Traditionally, researchers use Cohen's (1988) rule of thumb to assess the level of f^2 value, i.e., values of 0.02, 0.15, and 0.35 demonstrate the small, medium, and large effect sizes.

Table 5.15 summarizes the results of the moderation analysis of four indicators of the citizen's profiles on the relationships between the factors and the intention to engage with OGD. Based on the results, it can be inferred that most respondents' profiles do not moderate these relationships. The results support the moderating effect of education level on the technical factors – intention relationship and experience level on the social factors – intention relationship. Figure 5.7 and Figure 5.8 depict the simple slope plots of these two two-way moderation effects.

Table 5.15. The moderation analysis of the citizen's profiles: age, gender, education level, and experience level in OGD engagement.

Relationships	**Path coefficient**	***t* Value**	***p* Value**	**95% Bca CI**	**Sign. (*p* < 0.05)?**	**f^2 value**	**Effect size**
Moderator: Age							
Motivations → Behavioral Intentions	-0.092	0.906	0.365	[-0.289, 0.106]	No	0.011	-
Political Factors → Behavioral Intentions	0.021	0.246	0.806	[-0.143, 0.194]	No	0.001	-
Social Factors → Behavioral Intentions	-0.128	1.400	0.162	[-0.298, 0.071]	No	0.023	-
Technical Factors → Behavioral Intentions	-0.016	0.157	0.875	[-0.230, 0.155]	No	0.000	-
Moderator: Gender							
Motivations → Behavioral Intentions	-0.132	1.945	0.052	[-0.257, 0.021]	No	0.042	-
Political Factors → Behavioral Intentions	-0.008	0.082	0.935	[-0.129, 0.223]	No	0.000	-
Social Factors → Behavioral Intentions	-0.152	1.854	0.064	[-0.300, 0.018]	No	0.037	-
Technical Factors → Behavioral Intentions	-0.087	0.921	0.357	[-0.261, 0.108]	No	0.011	-
Moderator: Education level							
Motivations → Behavioral Intentions	-0.037	0.360	0.719	[-0.257, 0.148]	No	0.001	-
Political Factors → Behavioral Intentions	0.072	0.781	0.435	[-0.113, 0.	No	0.006	-
Social Factors → Behavioral Intentions	-0.056	0.597	0.550	[-0.231, 0.141]	No	0.005	-
Technical Factors → Behavioral	**0.208**	**2.039**	**0.042**	**[0.019, 0.403]**	**Yes**	**0.062**	**Small**
Moderator: Experience level							
Motivations → Behavioral Intentions	0.097	1.515	0.130	[-0.029, 0.229]	No	0.013	-
Political Factors → Behavioral Intentions	0.138	1.843	0.065	[0.011, 0.309]	No	0.017	-
Social Factors → Behavioral	**0.194**	**2.737**	**0.006**	**[0.072, 0.363]**	**Yes**	**0.042**	**Small**
Technical Factors → Behavioral Intentions	0.066	0.765	0.444	[-0.092, 0.245]	No	0.003	-

The results of the moderation analysis on the education level show that the relationship between technical factors (OGD quality) and behavioral intention to engage with OGD is -0.152 for an average level of the respondents' educational background. For higher levels of education (e.g., education level increases one standard deviation unit), the relationship between OGD quality and behavioral intention increases (i.e., -0.152 + 0.208 = 0.056). In contrast, for lower levels of education (e.g., education level decreases one standard deviation point), the relationship between OGD quality and behavioral intention becomes -0.152 – 0.208 = -0.360. As shown in Figure 5.7, the relationship between OGD quality and behavioral intention is positive for the higher education level, as indicated by their positive slope. Hence, the higher the quality of the OGD, the higher the level of intention to engage with it.

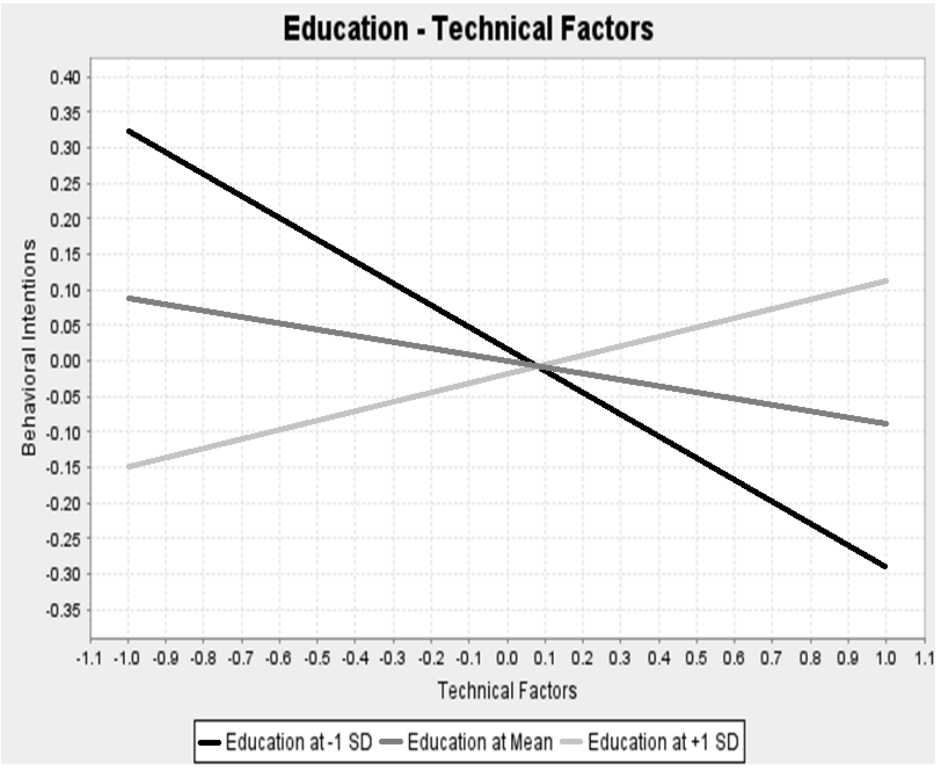


Figure 5.7. The simple slope plot of the two-way interaction effect of education level on technical factors – behavioral intention relationship.

On the contrary, the relationship is negative for the lower education level indicated by the negative slope: higher quality of OGD corresponds with lower intention. From the figure, we can see that the higher level of education has a flatter slope while the lower education has a steeper negative slope. Hence, it can be inferred that the overall relationship between the technical factors and behavioral intention is negative. We can also infer that higher education levels result in a stronger relationship between OGD quality and behavioral intention, while lower levels of education entail a weaker relationship between OGD quality and behavioral intention. The following plausible explanation about this difference was proposed. Respondents with a higher level of education may not be willing to expend valuable and scarce resources handling low-quality data and, therefore, prefer to engage with a higher-level quality of OGD. On the contrary, respondents with a lower level of education may be more critical of governments and want to scrutinize and expose low-quality OGD.

For an average level of the respondents' experience with OGD engagement, the results of the moderation analysis on experience level show that the relationship between social factors (social influence) and behavioral intention to engage with OGD is 0.185. The relationship between social influence and behavioral intention increases by the size of the interaction term (i.e., 0.185 + 0.194 = 0.379) for longer prior experienced respondents (e.g., the last engagement occurs more than two years before the data collection). Conversely, the relationship between social influence and behavioral intention becomes 0.185 – 0.194 = -0.009 for those who have recently experienced respondents (e.g., the last engagement happened less than one year before the survey). As shown in Figure 5.8, the relationship between social influence and behavioral intention is positive for the longer prior experienced respondents, as indicated by their positive slope. Therefore, more significant social influence corresponds with a higher level of intention to engage with OGD.

On the contrary, the relationship is negative for more recent experienced respondents as indicated by the negative slope: greater social influence correlates with lower intention. The figure shows that the more recent experience has a substantially flatter negative slope while the longer prior experience has a steeper positive slope. Hence, it can be inferred that the overall relationship between social influence and behavioral intention is positive. We can also infer that more recent experience leads to a weaker relationship between social influence and behavioral intention. In contrast, longer prior experience results in a stronger relationship between social influence and behavioral intention. The following plausible explanation about this difference was proposed. Respondents who have recently or for the first time engaged with OGD do not need to be influenced by their colleagues, friends, supervisors, or other important persons because they want to try out exploring OGD. On the other hand, respondents who have long experience with OGD may have a bad experience with OGD engagement and may need to be stimulated by their social environment to engage again with OGD.

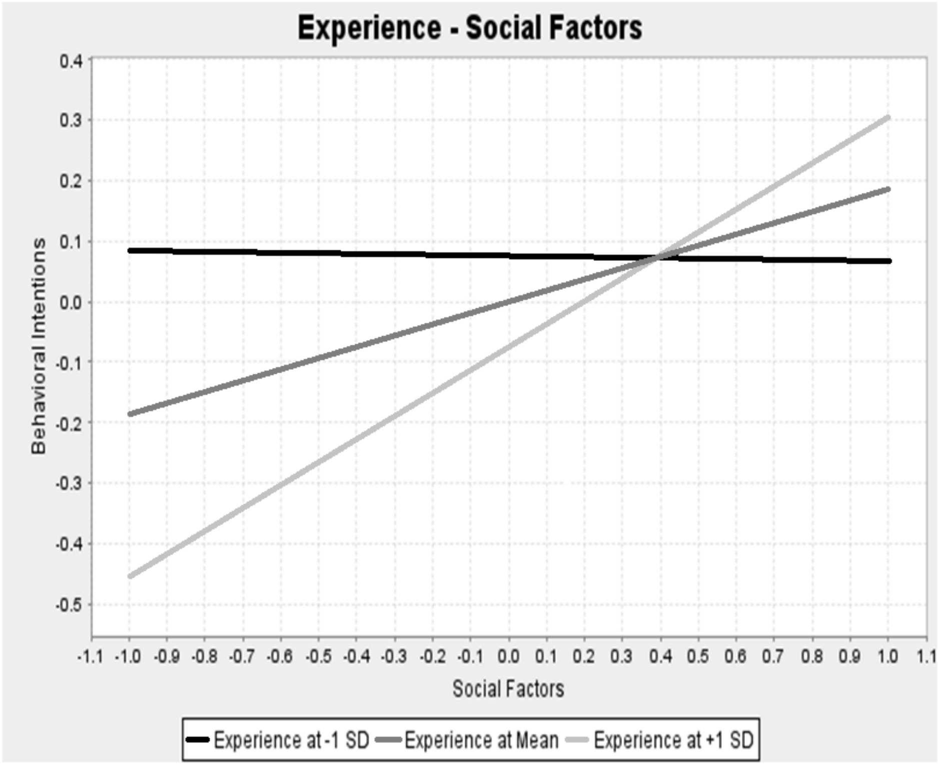


Figure 5.8. The simple slope plot of the two-way interaction effect of experience level on social factors – behavioral intention relationship.

5.3.5. Multigroup analysis

Applications of PLS-SEM usually analyze the relationships between exogenous and endogenous latent variables, assuming that the underlying data stem from a homogenous population (Hair et al., 2017; Hair et al., 2018). However, this assumption is often unrealistic because individuals are frequently different (Hair et al., 2017). Inadequacy to take such heterogeneity into account can be a validity threat to the PLS-SEM result because it can result in incorrect conclusions (Becker, Rai, Ringle, & Völckner, 2013). Based on the case study (see Chapter 4), the reasons to engage with OGD in a citizen-led initiative can be different from those in a government-led initiative. For example, citizens are motivated to engage with government-led OGD engagement because of their future career concerns and financial benefit. At the same time, those who engage with citizen-led OGD engagement are motivated by learning and developing skills. Therefore, it can be assumed that there might be significant differences in perceptions and evaluations on the factors that stimulate OGD engagement between those who engage in the citizen-led initiative and those in the government-led initiative. In this section, multigroup analysis using SmartPLS (Ringle et al., 2015) was reported and discussed to assess whether the researcher's assumption is supported.

Before using the multigroup analysis to examine group-specific parameter estimates for significant differences, researchers should ensure that measurement invariance (or measurement equivalence) exists (Hair et al., 2017). Researchers can use the measurement invariance of the composite models (MICOM) procedure developed by Henseler, Ringle, and Sarstedt (2016) to examine measurement invariance in the PLS-SEM context. The procedure is developed based on the latent constructs' scores. It involves three hierarchically interrelated steps: (1) configural invariance, (2) compositional invariance, and (3) equality of composite mean values and variances (Henseler et al., 2016). Researchers should assess configural invariance before assessing compositional invariance (Hair et al., 2017). When the researchers can confirm the configural and compositional variance, partial measurement invariance is supported, which in turn permits the multigroup analysis.

Researchers perform the MICOM procedure using the permutation algorithm developed and substantiated by Chin and Dibbern (2010) in SmartPLS 3 (Ringle et al., 2015). Table 5.24 shows the results of the second and third steps of the MICOM procedure, i.e., the compositional invariance and composite mean values and variances equality assessment. Schlägel and Sarstedt (2016) urge researchers to examine the original composite score correlations (c) and the empirical distribution of the composite score correlations computed from the permutation (c_u). Compositional invariance is established when c exceeds the 5% quantile of c_u. The results in Table 5.16 show that the compositional invariance applies to both citizen-led and government-led OGD engagement samples. The permutation confidence intervals of differences should be examined to assess the equality of a construct's mean values and variances across groups (Hair et al., 2018). The construct's mean values and variances can be assumed equal when the permutation confidence intervals of differences between the first and second group's construct scores include the original difference. The results show that partial measurement invariance is established because one construct, i.e., the motivation construct, indicates differences of mean values across the citizen-led and government-led OGD engagement. Based on these results, the model appears to meet the requirement for multigroup analysis.

Table 5.16. The research model's MICOM results.

Constructs	**Citizen-led vs. Government-led**			
	c		**5% quantile of c_u**	
Behavioral Intentions	0.999		0.998	
Motivations	0.993		0.964	
Political Factors	0.991		0.935	
Social Factors	0.989		0.948	
Technical Factors	0.843		0.103	
	Mean difference	**95% confidence interval**	**Logarithm of variances**	**95% confidence interval**
Behavioral Intentions	-0.323	[-0.327, 0.297]	0.299	[-0.451, 0.455]
Motivations	**-0.446**	**[-0.294, 0.324]**	0.428	[-0.533, 0.576]
Political Factors	-0.159	[-0.303, 0.311]	-0.006	[-0.431, 0.524]
Social Factors	-0.236	[-0.327, 0.303]	-0.060	[-0.453, 0.469]
Technical Factors	-0.283	[-0.313, 0.326]	-0.107	[-0.424, 0.482]

To analyze the significance of differences between respondents who prefer to engage with OGD in citizen-led initiatives and those who are inclined towards government-led initiatives, the output of the permutation test on the model's path coefficients should be assessed (Hair et al., 2018). Notably, the path coefficients original difference between these two groups of respondents was examined to determine whether it is included in the permutation confidence intervals of differences (Schlägel & Sarstedt, 2016). When the confidence interval excludes a relationship's path coefficients original differences, it can be assumed that the relationship is statistically different across groups. Table 5.17 exhibits the results of the permutation procedure on the path coefficients; the path coefficient difference of all relationships is included in the 95% confidence interval. Based on these outcomes, it can be safely assumed that the relationships between the factors and behavioral intentions to engage with OGD is not statistically different across two groups of OGD engagement.

Table 5.17. The permutation output for the model's path coefficients.

Relationship	**Path coefficient difference**	**95% confidence interval**	***p*-Value**
Motivations → Behavioral Intentions	0.024	[-0.293, 0.297]	0.886
Political Factors → Behavioral Intentions	0.143	[-0.313, 0.326]	0.418

Relationship	Path coefficient difference	95% confidence interval	*p*-Value
Social Factors → Behavioral Intentions	-0.214	[-0.333, 0.320]	0.192
Technical Factors → Behavioral Intentions	-0.060	[-0.469, 0.368]	0.804

5.3.6. Importance-performance map analysis

The importance-performance map analysis (IPMA) expands the PLS-SEM's path coefficient estimates report and adds a new dimension based on the construct's scores average values (Höck, Ringle, & Sarstedt, 2010). Specifically, IPMA compares the total effects of a structural model on a particular endogenous construct that indicates the antecedent construct's importance with the average latent variable scores of the construct's antecedents representing their performance (Hair et al., 2017; Ringle & Sarstedt, 2016). The goal of IPMA is to determine antecedents that have high importance (strong total effect) for the endogenous construct but at the same time have a somewhat low performance (low scores on the average latent variable) (Ringle & Sarstedt, 2016). In addition, the underlying aspects of these constructs indicate potential improvement areas that should get significant attention (Hair et al., 2017).

Ringle and Sarstedt (2016) suggest that the research design should meet three following requirements before applying IPMA. First, all indicators in the PLS path model have to employ a metric or quasi-metric scale such as an ordinal scale (Sarstedt & Mooi, 2019). Second, all indicators' coding must have the same scale direction in which their minimum values represent the worst outcome while the maximum values represent the best outcome (Ringle & Sarstedt, 2016). Third, all indicator's outer weights estimates have to be positive. Before conducting IPMA, all indicators in the research model should be checked and ensured that they had met the specified requirements. All of the indicators were measured in a quasi-metric scale, i.e., a five-point Likert scale that ranges from a minimum value of 1 (i.e., strongly disagree; the worst) to a maximum value of 5 (i.e., strongly agree; the best). The coding of three indicators, i.e., INT2, INT5, and EXT2, that were negatively worded, was reversed before data analysis (see Section 5.2.3). In the end, all indicators have the same scale direction. The results of the measurement model assessment (see Section 5.3.2) show that the indicators' outer weight estimates have positive values. Based on these assessments, it is safe to conclude that the research model meets the IPMA requirements.

Table 5.18 depicts the results of performing the IPMA procedure in the SmartPLS 3 software (Ringle et al., 2015). Based on these results, a construct- and an indicator-level map of the importance-performance matrix can be developed as shown in Figures 5.9 and 5.10, respectively. We can see that the constructs in the lower right area of the map indicate high importance for the target construct (i.e., endogenous construct) but have a low performance (Hair et al., 2017). Thus, there is a specifically high potential for improving these constructs' performance, whereas constructs that indicate lower importance have a lower potential for improvement. Furthermore, we can see that among the constructs, motivations have the highest effect and performance on behavioral intentions to engage with OGD.

Table 5.18. The importance-performance matrix of the constructs and indicators in the research model.

Construct		**Total effect (importance)**		**Avg. score (performance)**	
Motivations		0.497		74.097	
Political Factors		0.213		66.812	
Social Factors		0.134		69.456	
Technical Factors		-0.129		53.802	
Indicator	**Total effect (importance)**	**Avg. score (performance)**	**Indicator**	**Total effect (importance)**	**Avg. score (performance)**
Motivations			**Technical Factors**		
EXT1	0.120	85.693	DQ1	-0.013	35.392
EXT3	0.046	70.633	DQ2	-0.009	40.060
INT1	0.083	78.464	DQ3	-0.002	51.355
INT6	0.106	64.157	DQ4	-0.007	47.892
INT7	0.065	70.884	SERVQ1	-0.002	53.464
INT8	0.076	69.679	SERVQ2	-0.003	51.657
Social Factors			SERVQ3	-0.016	57.831
SOC1	0.024	56.175	SERVQ4	-0.010	50.452
SOC2	0.031	64.910	SYSQ1	-0.031	61.295
SOC3	0.025	63.102	SYSQ2	-0.018	60.241
SOC4	0.054	80.924	SYSQ3	-0.012	53.313
Political Factors			SYSQ4	-0.008	57.831
POL1	0.064	70.934			
TR1	0.049	66.416			
TR2	0.048	59.438			
TR3	0.051	68.976			

In contrast, technical factors have the lowest effect and performance. On the other hand, the IPMA procedure suggests that political factors indicate relatively high importance but lower performance than social and technical factors. Therefore, it can be inferred that there exists a high potential for improving the performance of political factors on citizens' behavioral intentions to engage with OGD. Notably, priority should be given to assessing the indicator-level IPMA map and identifying the indicators constituting political factors with a relatively high potential for improvements.

We can see that the indicators constituting motivations (i.e., EXT1, EXT3, INT1, and INT7) and political factors (i.e., SOC4 and POL1) have the highest effects and performance on behavioral intentions to engage with OGD at the indicator level. EXT1 concerns the relative advantage (or perceived benefits) of engaging with OGD, and EXT3 is related to the desire to get to know new people. INT1 and INT7 concern the self-efficacy of OGD engagement and enjoyment of engaging with OGD, respectively. SOC4 is about social influence to engage with OGD to benefit the society, while POL1 has to do with OGD engagement political efficacy. We can also see in Figure 5.10 that intrinsic motivation indicators such as INT6, INT8, and trust indicators including TR1, TR2, and TR3 are in the right lower area than the previous indicators, showing a relatively high potential of improvements. These constructs are highly likely relevant for public servants in charge of OGD to create appropriate intervention and apply managerial actions to enhance the citizens' behavioral intention to engage with OGD.

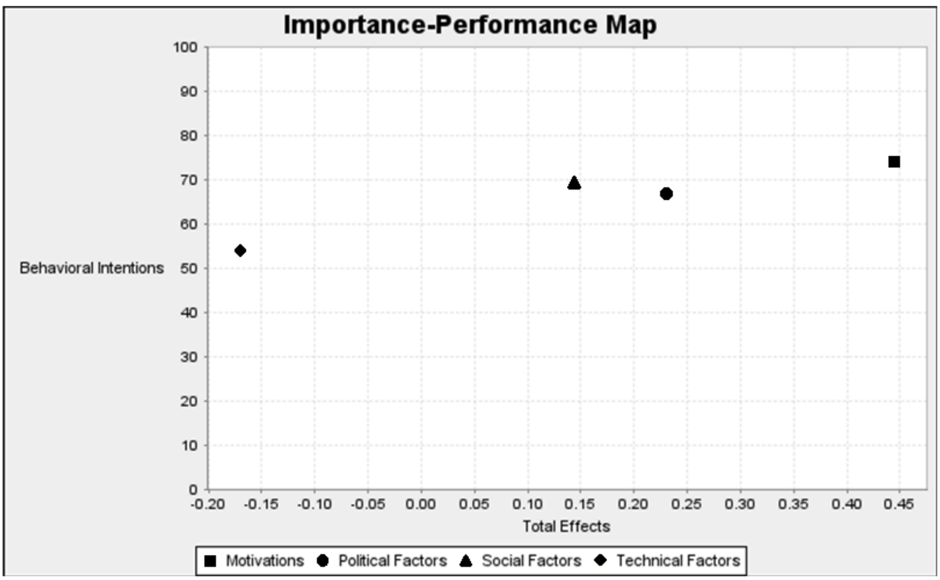


Figure 5.9. The construct-level importance-performance matrix map of the research model.

INT6 and INT8 are related to the enjoyment of studying OGD and the intellectual challenge of OGD engagement, respectively. When promoting and organizing OGD engagement events such as a hackathon, public servants should emphasize creating an atmosphere that enables citizens to learn and study OGD. For example, they can invite prominent data scientists that have been known for using OGD and creating something out of it to allow citizens to interact with the scientist and reap benefits from the interaction. On the other hand, public servants should also offer relevant societal problems and related OGD to challenge citizens to solve those problems with their capacity. For example, formulating a hackathon that evolves around solving real-life social problems relevant for the community of interest and at the same time provides relevant OGD that have a relatively high potential for the solutions competed in the event.

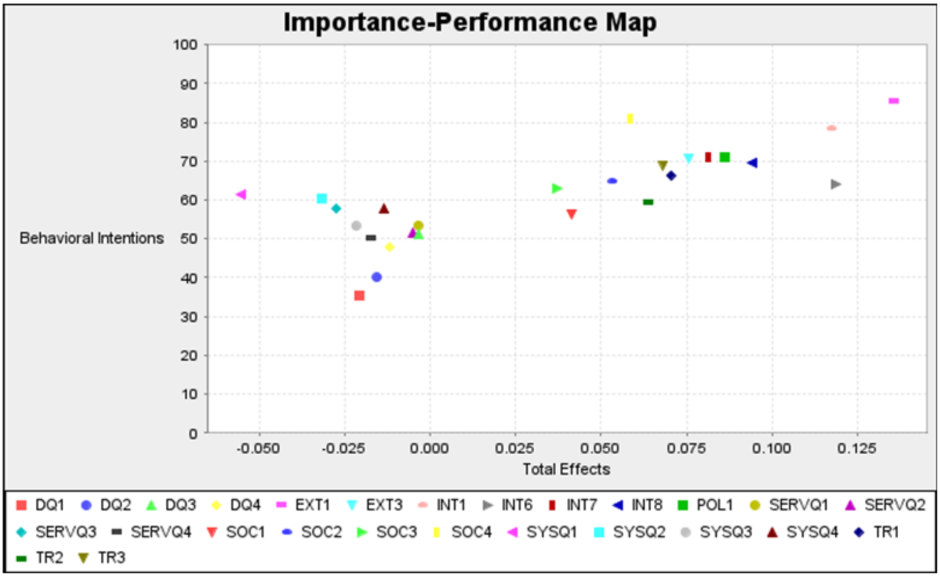


Figure 5.10. The indicator-level importance-performance matrix map of the research model.

TR1, TR2, and TR3 are related to the citizens' trust toward OGD and OGD providers. Purwanto et al. (2020) indicate that the quality of OGD viewed from three different dimensions, i.e., system quality, data quality, and service quality, influence this type of trust. Therefore, improving the OGD quality will highly likely enhance the citizens' trust. Public servants should increase the quality of the systems providing access to OGD (e.g., website, portal, tools), the OGD itself (e.g., accuracy, completeness, format, and interoperability). For example, ensuring that the links to download open data sets work well and provide different file format options for downloading data. At the same time, they should also actively offer help, guidance, and support to those who engage with OGD—for instance, they should establish a dedicated team that offers OGD related services and provides complete documentation on how to access and use OGD.

5.3.7. Hypotheses evaluation

Based on the PLS-SEM assessments, the significant test results of the main hypotheses that predict the influences of motivations, social factors, technical factors, and political factors on citizens' behavioral intention to engage with OGD were summarized in Table 5.19. The significance test results of hypotheses that predict the moderation of citizens' profiles, namely age, gender, education level, and OGD experience on the relationships between factors and behavioral intention, were also summarized in Table 5.21. It is important to note that the intrinsic motivation and extrinsic motivation constructs were merged into motivations construct because of discriminant validity problems between the two constructs. The problems refer to the validity assessment results indicating that intrinsic and extrinsic motivation constructs are not empirically unique and distinct. As a result, hypothesis *H1* was changed from representing the prediction of intrinsic motivation influence on behavioral intention relationship to postulating the positive effects of (both intrinsic and extrinsic) motivations in the final path model.

Furthermore, the significance test on hypothesis *H2* was not performed. Correspondingly, significance tests were not performed on hypotheses H7a – H7d that posit the moderating effects of citizens' profiles on the relationship between extrinsic motivation and behavioral intention. Instead, significance tests on hypotheses *H6a – H6d* that postulated the moderating effects of citizens' profiles on the relationship between (both intrinsic and extrinsic) motivations and behavioral intention were carried out.

Table 5.19. The results of the hypotheses significance test.

Hypothesis	Relationship	Significance	*t* Value	Status
H1	**Motivations → Behavioral Intention**	**Yes**	**5.980**	**Supported**
*H2**				
H3	Social Factors → Behavioral Intention	No	1.846	Not supported
H4	Technical Factors → Behavioral Intention	No	1.284	Not supported
H5	**Political Factors → Behavioral Intention**	**Yes**	**2.631**	**Supported**

* *H2* was not evaluated because intrinsic motivation and extrinsic motivation constructs have been merged into motivations construct

Table 5.19 indicates that among the tested hypotheses, the statistical analysis only supports H1 and H5. These results mean that (both intrinsic and extrinsic) motivations and political factors significantly positively influence citizens' behavioral intention to engage with OGD. These results also support the findings of previous empirical research. Previous studies on the intention to use OGD or OGD technologies or to participate in OGD hackathons show that both intrinsic (e.g., Juell-Skielse et al., 2014; Khayyat & Bannister, 2017; Wirtz et al., 2018) and extrinsic motivations influence the intention (e.g., Jurisch et al., 2015; Weerakkody, Irani, et al., 2017; Weerakkody, Kapoor, et al., 2017; Zuiderwijk et al., 2012). Moreover, previous research found that political factors influence the intention to use OGD (e.g., Cranefield et al., 2014; Hutter et al., 2011; Ruijer et al., 2017; Wijnhoven et al., 2015). The study respondents were drawn from OGD user communities such as the regional Open Knowledge Foundation Network and Kawal Pemilu 2019, who are especially active in pushing their government to open public data and interpret OGD, respectively. This type of respondent likely constitutes politically active citizens who demand transparency and the accountability of their governments (Peixoto, 2013). They trusted the available OGD, engaged with it, and created something out of it despite technical-related problems surrounding the OGD.

Contrary, the statistical evidence of this research indicates that the citizens' behavioral intention to engage with OGD is not significantly related to social and technical factors. These results contrast the findings of previous empirical research (e.g., Choi & Tausczik, 2017; Saxena & Janssen, 2017; Weerakkody, Kapoor, et al., 2017; Zuiderwijk, Janssen, et al., 2015). Unlike previous research (e.g., Choi & Tausczik, 2017; Saxena & Janssen, 2017; Weerakkody, Kapoor, et al., 2017; Zuiderwijk, Janssen, et al., 2015), the profiles of the respondents showed that they have a fair representation of different daily occupations that do not require them to engage with OGD. As a result, the respondents' important others in their social relationships might not influence them. The demographic data indicates that most of the respondents had a higher degree of education. At the same time, slightly more than half of the respondents claimed that they created visualization such as statistical charts or infographics based on OGD. Also, almost half of the respondents stated that they developed the mobile, computer, or web-based applications on top of OGD. Therefore, they likely possess sufficient knowledge and skills to overcome problems related to technical factors. In addition, based on the statistical analysis shown in Table 5.20, there exists a possibility that social factors and technical factors may indirectly affect citizens' intention through one or more mediating variables. Previous empirical research has also shown such indication (e.g., Fitriani et al., 2019).

Table 5.20 shows the results of latent variable correlations (LVCs) analysis, which is part of the discriminant validity assessment. These results indicate empirical supports for the proposition that some factors indirectly affect behavioral intention by mediating other factors. As we can see, the LVCs assessment indicates that political factors correlate with motivations, while social factors correlate with motivations. At the same time, political factors and technical factors correlate with motivations, political factors, and social factors. Furthermore, this evaluation shows that motivations and political factors partially mediate the relationship between social factors and citizens' intentions. At the same time, the political factors – behavioral intention relationship is partially mediated by motivations, while the technical factors – behavioral intention relationship is partially mediated by political factors.

Table 5.20. The latent variable correlations (LVCs) assessment.

Relationship	**LVC**	***t* Value**	***p*-Value**	**95% BCa CI**	**Significant?**
Political Factors → Motivations	**0.474**	**6.759**	**0.000**	**[0.316, 0.594]**	**Yes**
Social Factors → Motivations	**0.560**	**8.870**	**0.000**	**[0.409, 0.665]**	**Yes**
Social Factors → Political Factors	**0.438**	**7.094**	**0.000**	**[0.293, 0.542]**	**Yes**
Technical Factors → Motivations	0.365	2.759	0.006	[-0.103, 0.495]	No
Technical Factors → Political Factors	**0.549**	**3.657**	**0.000**	**[0.121, 0.679]**	**Yes**
Technical Factors → Social Factors	0.287	2.285	0.022	[-0.038, 0.470]	No

Furthermore, we can see that among the tested moderation hypotheses, the statistical analysis only supports *H8d* and *H9c* (see Table 5.21). Hypothesis *H8d* predicts that experience with OGD moderates the influence of social factors on behavioral intention, while hypothesis *H9c* posits that education level moderates the effect of technical factors on behavioral intention. These findings support previous empirical studies investigating the effects of citizens' profiles on the intention to engage with OGD (e.g., Hutter et al., 2011; Wang et al., 2019). The significance and hypotheses evaluation results were taken into account in developing the final research model explained in the next section.

Table 5.21. The significance and moderating hypotheses testing results.

Hypothesis	**Moderator**	**Relationship**	**Significance**	***t* Value**	**Status**
H6a	Age	Motivations → Behavioral Intention	No	0.906	Not supported

Hypothesis	**Moderator**	**Relationship**	**Significance**	***t* Value**	**Status**
H6b	Gender	Motivations → Behavioral Intention	No	1.945	Not supported
H6c	Education level	Motivations → Behavioral Intention	No	0.360	Not supported
H6d	Experience	Motivations → Behavioral Intention	No	1.515	Not supported
*H7a – H7d**					
H8a	Age	Social Factors → Behavioral Intention	No	1.400	Not supported
H8b	Gender	Social Factors → Behavioral Intention	No	1.854	Not supported
H8c	Education level	Social Factors → Behavioral Intention	No	0.597	Not supported
H8d	**Experience**	**Social Factors → Behavioral Intention**	**Yes**	**2.737**	**Supported**
H9a	Age	Technical Factors → Behavioral Intention	No	0.157	Not supported
H9b	Gender	Technical Factors → Behavioral Intention	No	0.921	Not supported
H9c	**Education level**	**Technical Factors → Behavioral Intention**	**Yes**	**2.039**	**Supported**
H9d	Experience	Technical Factors → Behavioral Intention	No	0.765	Not supported
H10a	Age	Political Factors → Behavioral Intention	No	0.246	Not supported
H10b	Gender	Political Factors → Behavioral Intention	No	0.082	Not supported
H10c	Education level	Political Factors → Behavioral Intention	No	0.781	Not supported
H10d	Experience	Political Factors → Behavioral Intention	No	1.843	Not supported

* *H7a*, *H7b*, *H7c*, and *H7d* were not evaluated because intrinsic motivation and extrinsic motivation constructs have been merged into motivation construct

5.4. Conclusion and answer to the third research question

Chapter 4 reported multiple case studies to identify the reasons and motivations behind citizen engagement with OGD and helped generate a research model that hypothesizes the relationships between factors and intention to engage with OGD. This chapter reported and discussed the quantitative investigation of the proposed research model using a PLS-SEM approach. The chapter reported the quantitative study design, including the survey development, instruments used to measure the survey, and data collection strategy. It also reported the preparation processes of the collected data before applying the PLS-SEM technique for data analysis. Furthermore, it reported and discussed the results of the PLS-SEM assessment. More importantly, this chapter answers RQ3: *what model explains citizens' intention to engage with OGD?* Ultimately, this chapter presents the primary outcome of this research: the final model of OGD citizen engagement (OGD-CEM).

Based on the results of the PLS-SEM assessment, the following inferences can be drawn. First, among the constructs examined in the quantitative study, motivations and political factors are the most critical constructs that influence citizens' intention to engage with OGD. It can also be inferred that respondents' motivations have a larger effect on intention than political factors; there is a large room for improvement in increasing citizens' perception of political factors. Second, at the indicator level, relative advantage (EXT1), desire to get to know new people (EXT3), self-efficacy (INT1), enjoyment (INT7), and desire to benefit society (SOC4) are important indicators that have the most significant effects on the intention. On the other hand, a potential improvement exists to increase citizens' perception of enjoyment of studying OGD (INT6), intellectual challenge (INT8), and trust in OGD and OGD providers (TR1, TR2, and TR3). Third, citizens' experience with OGD moderates the influence of social factors on behavioral intention, while citizens' education level moderates the effect of technical factors on behavioral intention. Lastly, the relationships between the factors and behavioral intentions to engage with OGD are not statistically different across two OGD engagement groups, namely citizen-led and government-led engagement.

The outcome of this research stage is the final version research model, namely the OGD Citizen Engagement Model (OGD-CEM) (see Figure 5.11). The model is built based on the statistical analysis reported in the previous sections. As we can see in Table 5.11, the path coefficients' significance assessment results indicate that social factors and technical factors do not have a statistically significant influence on behavioral intention to engage with OGD. Therefore, the final model comprises only motivations and political factors that influence citizens' behavioral intention to engage with OGD. Both relationships: social factors → behavioral intention, and technical factors → behavioral intention were removed from the model. No moderation effects of citizens' profiles exist in the OGD-CEM model because such effects apply only to removed relationships.

The final OGD-CEM model explains that (both extrinsic and intrinsic) motivations toward the engagement and perceived political factors toward OGD and its provider determine citizens' behavioral intention to engage with OGD. Notably, in the extrinsic motivation context, the more citizens perceive that engaging with OGD will give them an advantage and provide an opportunity to broaden their social networks, the more they will be inclined to engage with OGD. In the intrinsic motivation context, the more citizens perceive that they can engage with OGD easily, that engaging with OGD is enjoyable, and that OGD engagement challenges them intellectually, the more they will likely engage with OGD. Furthermore, the more citizens perceive that their engagement with OGD will influence public policy, and the higher citizens' trust in OGD and the governmental organizations that provide it, the more they will be inclined to engage with OGD. Researchers have widely investigated the effect of motivations on citizens' intention to engage with OGD in the open data domain, and the model reinforces the findings of these studies.

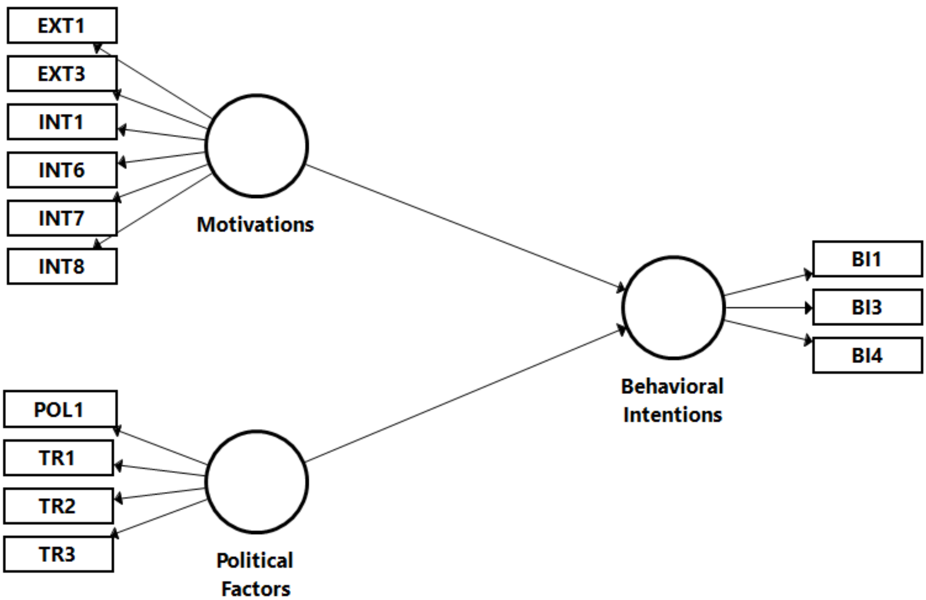


Figure 5.11. The final version of the OGD-CEM.

On the other hand, researchers barely study the influence of political factors on citizens' intention in the open data domain, and thus, the model is among the firsts to explain this relationship. However, the model does not predict that social and technical factors determine citizens' behavioral intention; the influence of citizens' social relationships and their perceived OGD quality do not affect the intention to engage with OGD. Thus, the model contradicts the findings of previous research that infer that social influence and three dimensions of IS quality (i.e., data quality, system quality, and service quality) affect citizens' behavioral intention. However, the statistical analysis results indicate a plausible explanation: social factors have indirect effects on intention via motivations and political factors. In contrast, technical factors have direct and indirect effects (partial effect) on intention via political factors. Researchers have widely studied these relationships in the e-commerce and e-government literature.

6. Conclusions, Limitations, and Further Research

This research examined the factors influencing citizens’ intention to engage with Open Government Data (OGD) to create artifacts that they envision solving societal problems. In this dissertation, citizen engagement with OGD is defined as the citizens' collaborative activities to convert OGD into valuable artifacts that are important and relevant to them and society. OGD engagement can benefit the government, for example, by facilitating policy implementation and society by generating ideas, information, and service innovation. This study focuses on a particular type of citizen, namely digitally literate citizens who are not government officials. Outside the scope of this research are OGD providers (e.g., governmental organizations), OGD users from the private (e.g., companies) and public sectors (e.g., civil society organizations), and OGD end-users (e.g., society).

OGD programs’ success is contingent upon, among other factors, citizen engagement with OGD. However, researchers often overlook the comprehension about citizens who engage with OGD. Furthermore, the current open data literature provides a fragmented insight into the factors influencing citizens’ intention to engage with OGD and rarely integrates different adoption and acceptance theories applied in the IS research domain. Therefore, this study aims *to develop a model for understanding factors contributing to citizen engagement with OGD*.

This chapter first addresses the key findings from this study in Section 6.1, followed by its scientific contributions and its implications to OGD practice (Section 6.2). Then the limitations of this research and the corresponding agendas for future research are addressed in Section 6.3 and 6.4, respectively. Lastly, this chapter provides reflections on the substance of this research in Section 6.5.

6.1. Findings from this research

This section first discusses the findings of this research related to each of the formulated three research questions (Section 6.1.1-6.1.3). At the end, this section discusses the overall attainment of the research objectives (Section 6.1.4).

6.1.1. RQ1: driving and inhibiting factors of OGD citizen engagement

The first research question (i.e., what drivers and inhibitors for citizen engagement with OGD have been identified in previous research?) aimed to identify driving and inhibiting factors of citizen engagement with OGD. This question helped the researcher understand what we have empirically known regarding citizen engagement with OGD and the factors that drive citizens to engage with OGD and inhibit them from doing so. A systematic literature review was conducted to answer this question.

Based on the literature synthesis, there are two types of factors associated with citizen engagement: 1) factors that directly influence OGD engagement and 2) citizen's profiles that moderate the relationships between factors and OGD engagement. Driving factors of OGD engagement related to the following six groups were found: intrinsic motivations, extrinsic motivations, economic factors, social factors, technical factors, and political factors. At the same time, inhibiting factors were identified, and they were categorized into two groups: technical and political factors. Six dimensions of intrinsic motivations that drive citizens to engage with OGD were identified: fun and enjoyment in exploring data, perceived ease of use, altruism, intellectual challenge, the relevance of OGD engagement to citizens' beliefs, and wanting to learn new things. At the same time, citizens are extrinsically motivated to engage with OGD because they perceive that engaging with OGD offers relative advantage or usefulness to job performance and future career concerns. Economic factors such as expecting monetary/financial rewards and potential gains from OGD engagement can also drive citizens to engage with OGD.

Furthermore, social factors, particularly the influence of social relationships and desire to benefit society, can drive citizens to engage with OGD. Three technical factors that drive OGD engagement were identified: OGD system quality, OGD quality, and OGD service quality. OGD system quality refers to the citizen's perception of the quality of the systems/platforms/technologies providing access to data in terms of functionalities/features, user-friendliness, availability, and response time. OGD quality is related to the citizen's perception of the OGD's quality, such as relevance, completeness, timeliness, and reliability. OGD service quality refers to the citizen's perception about the quality of the support and services provided for the OGD and OGD system usage regarding the availability of help or documentation and the responsiveness to citizen feedback. Finally, concerning the political factors, citizens' trust in government, need for change, and participation in public issues can drive them to engage with OGD.

Technical and political factors can both be drivers and inhibitors of OGD engagement at ends of the same seesaw, showing an opposed relationship; the increase of drivers will decrease inhibitors, and vice versa. Task complexity and the low quality of OGD, OGD system, and OGD service likely inhibit OGD engagement. The absence of five dimensions of OGD system quality concerning system documentation, functionality, user-friendliness, integration, and responsiveness can also inhibit citizens from engaging with OGD. At the same time, the lack or the absence of timely, interoperable, well-formatted, complete, accessible, available, accurate data and metadata can be inhibitors to OGD engagement. Regarding OGD service quality, the lack of support for the use of data, communication between data user and data provider, and feedback mechanism can inhibit citizens from engaging with OGD. Finally, the lack or the absence of trust is the primary inhibitor of OGD engagement in political factors. It is also important to note that the absence or the lack of certain driving factors such as intrinsic motivations, extrinsic motivation, economic factors, and social factors does not automatically equate to the presence of inhibiting factors.

The second type of factor is citizen's profiles that moderate the relationships between factors and OGD engagement. Citizens with particular characteristics (profiles) correlate with either more or less tendency to engage with OGD. The citizen profiles that influence her or his intention to engage with OGD can be differentiated based on nine factors: age, gender, education level, capabilities, resources, competency, experience, awareness, and voluntariness.

6.1.2. RQ2: citizens' motivations to engage with OGD

The second research question (i.e., why do citizens engage with OGD in existing government-led and citizen-led OGD initiatives?) aimed to explore the identified driving and inhibiting factors influencing OGD citizen engagement and identify new factors that might be missing from the literature. A multiple case study approach involving real-life settings of OGD engagement was employed to attain this objective. The clusters of factors derived from the systematic literature review (see Section 3.2.2, 3.4, and 3.5) were used as a framework to elicit citizens' thoughts, feelings, or experiences toward their engagement with OGD from two case studies. The two selected cases concerned the government-led citizen engagement with OGD from the primary education and education inspection domains (Hack de Valse Start) and the citizen-led engagement with OGD from the election domain (Kawal Pemilu). In the former case, citizens engaged with open education inspection data provided by the Dutch government's Inspectorate of Education to solve the challenges competed in the Hack de Valse Start hackathon concerning inequality in primary education. In the latter case, citizens engaged with open election data published by the Indonesian General Election Commission to digitize the election results, make them available for the public on the Internet, and provide reports on anomalous results to KPU. Semi-structured interviews, documents, web pages, data sets, and participant observations were examined in both cases. The cases allowed the researcher to analyze why citizens engaged with the open education inspection data and open election data.

Fifteen factors from the theoretical framework derived from the literature review study were identified, including intrinsic and extrinsic motivations, social, technical, and political factors, influential on government-led and citizen-led OGD engagement. Citizens who participated in both cases were found motivated by the feeling of fun and enjoyment (intrinsic motivations) and the desire to get to know new people (extrinsic motivations). Wider social relationships and desire to benefit society (social factors) can also drive citizens to engage with OGD. Concerning technical factors, citizens were more likely to engage with OGD when the OGD meets their expectations of accurate, well-formatted, up-to-date, and easy-to-understand data and reliable, assuring, and responsive support. Furthermore, citizens' trust in OGD, interests in politics, political change expectations, and involvement in political activities (political factors) can drive citizen engagement in both studied cases. In contrast, two factors categorized in intrinsic motivations and social factors did not influence both engagement types: status and reputation and the influence of close social relationships. Citizens were not motivated by the desire to increase or conserve their status and reputation. At the same time, citizens' close social relationships did not influence them to engage with OGD.

Furthermore, some factors were found to play a role only in a particular case. Eight factors from five categories, i.e., intrinsic and extrinsic motivations, economic, technical, and political factors, were related to only one type of engagement. Five among these factors play a role only in the government-led engagement, while the other three play a role in the citizen-led engagement. On the one hand, citizens were motivated by the intellectual challenge (intrinsic motivations), future career concerns (extrinsic motivations), financial reward (economic factors), data completeness, and data interoperability (technical factors) in the Hack de Valse Start case. On the other hand, citizens were motivated by the desire to learn and develop skills (intrinsic motivations), OGD system reliability (technical factors), and government responsiveness (political factors) in the Kawal Pemilu case.

Only one factor was found missing from the current literature: the novelty of OGD engagement. This factor was identified in the Kawal Pemilu case. Novelty refers to the impression of a new experience in engaging with OGD. Citizens wanted to engage with OGD because the way they do so is something they had never experienced before.

In the case studies, citizens' profiles, assuming to moderate the relationships between factors found in the literature, were not examined. The collected evidence did not show whether a particular factor influences a citizen with a specific profile more than a citizen with a different profile. In addition, the number of citizens who participated in the case study is low, and thus, investigating their age, gender, and other background factors is impractical. Finally, studying these citizens' profiles might also be considered violating their privacy from an ethical perspective.

6.1.3. RQ3: the model that explains the factors influencing citizens' intention to engage with OGD

The third research question (i.e., what model explains citizens' intention to engage with OGD?) aimed to evaluate the proposed hypotheses (and research model) of the OGD citizen engagement model (OGD-CEM) using a larger sample of citizens. This question helped to resemble the outcome of this research, i.e., the final OGD-CEM model. The model hypothesized that intrinsic motivation, extrinsic motivation, social factors, technical factors, and political factors determine citizens' intention to engage with OGD. The model also hypothesized that citizens' profiles moderate the relationships between the factors and intention. A multivariate analysis using a PLS-SEM technique was used to achieve this objective. An online questionnaire was developed and distributed to various open data user communities. Data from 627 respondents were collected using a non-probability sampling technique; 461 of these responses were excluded due to missing data. Overall, 166 usable responses were analyzed, representing a 26.5% survey completion rate.

Five types of assessment in the PLS-SEM analysis were carried out: measurement model assessment, structural model assessment, moderation effect analysis, multigroup analysis, and Importance-Performance Map Analysis (IPMA). Measurement model assessment concerns evaluating the relationships between the indicators (the question items asked in the survey) and the factors they measure. At the same time, structural model assessment evaluates the relationships between the factors and citizens' intention to engage with OGD. These two assessments were the primary means of testing the proposed research model's hypotheses and suggesting the final research model after testing the hypotheses. Moderation effect analysis evaluates whether particular citizens' profiles moderate the relationships between the factors and citizens' behavioral intention to engage with OGD. Multigroup analysis was employed to assess whether there are significant differences of perceptions and evaluations on the factors that stimulate OGD engagement between those who engage in the citizen-led initiative and those in the government-led initiative. Finally, IPMA concerns identifying factors and indicators that have relatively high importance (strong total effect) for citizens' intention to engage with OGD but at the same time have a relatively low performance. IPMA indicates potential improvement areas that should receive significant attention.

The following results were attained from the PLS-SEM assessments:

1. (Both extrinsic and intrinsic) motivations and political factors are the most critical constructs that influence citizens' intention to engage with OGD. Motivations have a larger effect on intention than political factors; there is considerable room to improve citizens' perception of political factors.
2. At the indicator level, relative advantage, the desire to get to know new people, self-efficacy, enjoyment, and the desire to benefit society are important indicators of motivations and political factors that affect the citizens' intentions. However, there exists a potential improvement for increasing citizens' willingness to engage with OGD by encouraging more enjoyable and intellectually challenging OGD engagement. Increasing citizens' trust in OGD and OGD providers can also improve citizens' intentions to engage with OGD.
3. Citizens' experience with OGD moderates the influence of social factors on behavioral intention, while citizens' education levels moderate the effect of technical factors on behavioral intention.
4. The relationships between the factors and behavioral intentions to engage with OGD are not statistically different across the two types of OGD engagement, namely citizen-led engagement, and government-led engagement.

6.1.4. Research objective: OGD citizen engagement model (OGD-CEM)

The outcome of this research is a model that explains the factors that influence citizen engagement with OGD, hereafter named the OGD Citizen Engagement Model (OGD-CEM) (see Figure 5.11). The model was built based on the multivariate analysis using a PLS-SEM approach. OGD-CEM model explained that (both extrinsic and intrinsic) motivations toward the engagement and perceived political factors toward OGD and its provider determine citizens' behavioral intention to engage with OGD. Notably, in the extrinsic motivation context, the more citizens perceive that engaging with OGD will give them an advantage and provide the opportunity to broaden their social networks, the more inclined they will be to engage with OGD. In the intrinsic motivation context, the more citizens perceive that they can engage with OGD easily, that engaging with OGD is enjoyable, and that OGD engagement challenges them intellectually, the more likely they will engage with it. Furthermore, the more citizens perceive that their engagement with OGD will influence public policy, and the higher citizens' trust in OGD and the governmental organizations that provide it, the more they will be inclined to engage with OGD.

Researchers have widely investigated the effect of motivations on citizens' intention to engage with OGD in the open data domain, and OGD-CEM reinforced the findings of these studies. On the other hand, researchers scarcely study the influence of political factors on citizens' intentions in the domain. OGD-CEM was among the first that contribute to explaining this relationship. OGD-CEM contradicts previous research findings; the influence of citizens' social relationships and their perceived OGD quality (i.e., data quality, system quality, and service quality) do not affect the intention to engage with OGD. However, the PLS-SEM results indicated a plausible explanation: social factors indirectly affect intention via motivations and political factors. In contrast, technical factors have direct and indirect effects (partial effect) on intention to engage with OGD via political factors.

6.2. Contributions

This section discusses the contributions of this research toward the science and practice of OGD. The scientific contributions of this research are mainly related to the adoption, acceptance, and usage of OGD from the citizens' perspective (Section 6.2.1). At the same time, the practical implications of this research specifically concern the use of the OGD-CEM model by public servants responsible for delivering OGD programs (Section 6.2.2).

6.2.1. Scientific contributions

This research scientifically contributes to the existing open data literature concerning the following areas:

1. This study is among the first to provide an integrated overview of the profiles of citizens who engage with OGD and the drivers and inhibitors of citizen engagement with OGD. In their review, Hossain et al. (2016) urge the importance of a comprehensive open data adoption model, which did not exist yet at the start of this study. This research contributes to the development of such a model. It offers an integrated theoretical framework of factors influencing OGD citizen engagement built on the existing empirical open data studies and theories on technology adoption/acceptance applied in open data.
   Furthermore, Susha, Grönlund, et al. (2015) propose research that focuses on understanding the OGD users. This research describes the profiles of citizens who engage with OGD. It also describes the relationships between the profiles and their intentions to engage with OGD.
2. This study mainly develops a theoretical model of OGD adoption by citizens, namely OGD-CEM, based on a systematic literature review. It extends and integrates theories and theoretical models of technology acceptance/adoption rooted in the individual's behavioral intention. This study also posits that OGD combines data, technology (e.g., portals that

provide access to OGD), and services (e.g., user support). It proposes a new classification of factors influencing citizens' intentions to engage with OGD from similar and sometimes overlapping dimensions formulated in the theories and theoretical models of technology acceptance/adoption. Such theories and theoretical models include the Theory of Planned Behavior (Ajzen, 1991), Technology Acceptance Model (Davis, 1989), Information Systems Success Model (DeLone & McLean, 2003), Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003). This study also integrates political participation and trust theories into the model because engaging with OGD can yield outcomes with high political values.

3. This study reinforces previous research focusing on citizens as the direct users of open data. It offers insights into citizens' roles in real-life OGD engagement cases (e.g., Hivon & Titah, 2017). Furthermore, this study complements previous research that views citizens solely as OGD beneficiaries who typically do not engage with OGD (e.g., Harrison et al., 2012; Parycek et al., 2014). It is among the first to provide empirical evidence about the citizen-led OGD engagement. In this type of OGD engagement, citizens entirely independently engage with OGD without explicit encouragement from the OGD provider, i.e., the government.
4. In their OGD literature review, Safarov et al. (2017) call for open data research that assesses the relationship between factors and the intention to engage with OGD. This study contributes to such an empirical assessment by evaluating the proposed OGD-CEM model using a quantitative research approach, namely multivariate analysis. It is among the first to empirically evaluate the unified theoretical model of factors influencing citizens' intention to engage with OGD. It is also among the first to empirically test assumptions held in the open data research regarding the effects of OGD quality, political participation, and trust on the intention to engage with OGD.

6.2.2. Practical implications

This research offers practical contributions to different OGD practitioners and stakeholders. Prominent practitioners from governmental organizations include civil servants responsible for designing and improving OGD programs to engage citizens in solving societal problems. At the same time, this research also offers insights to non-government OGD practitioners to help design and improve citizen-led OGD engagement initiatives. External auditors represent an important stakeholder of OGD programs who can also reap benefits from this research. The following sections explain these practical contributions.

Governmental organizations can adopt the theoretical framework derived from a systematic review of OGD literature as a basis for OGD program evaluation. Civil servants responsible for maintaining a particular program can adopt the framework and utilize the factors' indicators to develop an instrument to understand what type of citizens engage with OGD and identify their engagement motivations. Insights into citizens' demographic background and motivations can help civil servants design an engagement initiative that attracts a particular type of citizen. For example, they can add a physical game activity as a side event of an OGD hackathon to add more fun and enjoyment for intrinsically motivated participants. As another example, civil servants can invite more companies or universities to participate in the hackathon to recruit more participants who are likely motivated by engaging with OGD to increase their job performance.

Governmental organizations can also use the framework indicators to understand better the quality of the OGD program delivered according to its users' perceptions. Civil servants can develop a program quality evaluation by adopting the theoretical framework's three dimensions of OGD quality (i.e., OGD quality, OGD system quality, and OGD service quality). Based on this evaluation, they can prioritize the improvements of each quality dimension, which will help improve citizens' trust in their OGD. For example, civil servants can improve the interoperability of a particular opened data set by publishing other relevant data sets typically stored in their relational database so that the hackathon participants can generate more meaningful information from the OGD. Another example concerns improving the OGD portal functionality by updating broken navigation links to particular data sets. Civil servants can also increase the quality of the OGD related services by providing, for example, online help through social media channels and being more responsive towards OGD users' requests.

Citizens who initiate an independent OGD engagement without support or interventions from the government can use the insights from this research to improve the engage-ability of their initiatives in several ways. First, as reported in the case study, citizens who are likely to engage independently with OGD are typically activists having similar expectations toward benefitting society. Therefore, the independent OGD engagement initiators can recruit like-minded citizens using their social influence through social networks such as social media and emphasizing the importance of their peers' involvement in solving societal problems. Second, by bringing urgent attention to relevant societal problems, these citizens, who typically have high political efficacy, will likely be motivated to engage with OGD and use the engagement outcomes to exercise their influence on public policy.

As reported in the research model assessment, a potential improvement exists for the citizens' perception of enjoyment of studying OGD, intellectual challenge, and trust in OGD and OGD providers. Using these insights, citizens who initiate independent OGD engagement, for instance, can form groups of participants in which champions (i.e., members who have a higher level of data exploitation skills) can mentor other members. On the other hand, civil servants responsible for OGD programs should focus on increasing citizens' trust by improving the quality dimensions of OGD.

This study also offers practical recommendations for governmental organizations that have the mandate for evaluating OGD programs such as external auditors (e.g., National Audit Office) in the performance audit framework. Performance auditing is "an independent, objective and reliable examination of whether government undertakings, systems, operations, programs, activities or organizations are operating following the principles of economy, efficiency and effectiveness" (INTOSAI, 2013, p. 2). *Efficiency* refers to getting the most from the available resources and concerns the relationship between resources used and outputs delivered in quantity, quality, and timing dimensions. At the same time, *effectiveness* refers to meeting the objectives set and attaining the intended results. The use of OGD by citizens is one of the OGD program's objectives at local and national levels (Zuiderwijk, Shinde, & Janssen, 2019); the efficiency and effectiveness of OGD programs can be assessed using a performance audit framework. In performance auditing, auditors are often involved in developing or selecting the audit criteria. Particularly for the effectiveness auditing of OGD programs, auditors can adapt the measurements or indicators of the OGD-CEM model as their audit criteria. Auditors can use a survey instrument built on the indicators such as OGD quality to collect data about the user perceptions toward particular OGD program. Furthermore, a simple descriptive analysis can be employed to interpret the collected data to support the auditors' decisions to judge the effectiveness of the OGD program.

6.3. Limitations

This study has attained its aims and objectives and answered the research questions formulated in Section 1.5. However, it is essential to note that readers should interpret the findings of this study in light of its limitations. In total, six limitations were identified in the study. These limitations particularly concern using the theoretical framework in the selected cases, potential biases from applying the participant observation approach in the case study, and the generalization of the findings from the selected cases. The limitations are also related to using a non-probability sampling approach in the survey study, the generalization of the findings from the multivariate analysis, and missing factors excluded from the research model.

6.3.1. The role of the theoretical framework in the case study

The theoretical framework developed from the systematic literature review mainly guides the conduct of the multiple case studies in this research. The semi-structured interview instruments were developed based on the framework; most interview questions were derived from the drivers and inhibitors identified in the framework. Although the researcher also asked open questions to identify other factors that might influence the respondents and understand how they engage with OGD, the former question was asked at the end of the interview. More than twenty questions related to the factors derived from the theoretical framework were asked earlier in the interview (see Table 4.4 for a complete overview of the interview questions). Therefore, some interviewees might think that those factors are already complete and become too exhausted to think of other factors beyond those asked. As a result, the researcher might not identify additional relevant factors.

6.3.2. Potential bias of the participant observation approach

When conducting the multiple case study, different techniques were employed, including the participant observation approach to elicit relevant qualitative data; the researcher participated in both selected cases. In the Hack de Valse Start case, the researcher participated in the hackathon and was a member of one of the teams under investigation. In the Kawal Pemilu case, he was one of the volunteers who digitized election results data. Referring to Schwartz and Schwartz's (1955) categorization of participant observation, the researcher's role was an *active* participant-observer in both cases. In this role, the researcher "maximizes his participation with the observed in order to gather data and attempts to integrate his role with other roles in the social situation" (Schwartz & Schwartz, 1955, p. 349). In the Hack de Valse Start case, the researcher did not conceal his purpose to other hackathon team members, while in the Kawal Pemilu case, he was not yet researching at that time. The role of the researcher in these situations is named *complete participant*, whose true identity and purpose are entirely unknown to the observed (Gold, 1958).

Obtaining access to the case study as a participant-observer offers an exceptional opportunity to comprehend the OGD engagement processes from an insider's view (Yin, 2014). Nevertheless, the researcher is cautious that the participant-observation strategy may lead to potential biases. These biases stem from the researcher's influence on the study's context as a participant and the context's influence on the researcher as an observer. As an active participant, the researcher can influence the observed, i.e., the members of the hackathon team and Kawal Pemilu, by, for example, interacting sympathetically to gain more support when he interviewed them afterward. On the other hand, the observed respondents can also influence the researcher, making biased research decisions based on this interaction. For example, the researcher might exclude one of the respondents from the interview because, during their interaction, the observed respondent shows unfavorable behavior related to the studied factors. However, in the Hack de Valse Start case, the researcher was able to play a complete participant role during the hackathon and did not take the leading role in the team. He revealed his true identity and purpose of participating in the hackathon only after the event concluded. In a similar vein, in the Kawal Pemilu case, the researcher was not a volunteer group coordinator who could control other volunteers' job performance. Moreover, most activities involving many participants were performed virtually. At that time, the researcher has not yet started this research.

6.3.3. Generalization of the case study findings

Two cases selected in this research were examined to determine whether the factors influencing citizen engagement with OGD derived from the systematic literature review exist and identify new factors missing from the literature. These cases were selected based on OGD use by digitally literate or technologically skilled citizens who did not hold any government positions and aimed to solve societal issues. As a result, this study targeted a particular group of OGD users in the case studies, i.e., people who engaged with OGD and created OGD-based artifacts. This type of OGD use is substantially different from other types of OGD use. For example, social science researchers use OGD to develop scientific articles as part of their jobs. Another example concerns the use of OGD by parents to evaluate the quality of schools. These parents used OGD as part of their decision-making when looking for a school for their children. This type of OGD also significantly differs from the use of OGD by employees of a company or entrepreneurs to develop a potential commercial application.

The Hack de Valse Start case involved Dutch governmental organizations that provided open education inspection data, and citizens engaged with the data in a hackathon. On the other hand, the Kawal Pemilu case involved Indonesian governmental organizations that published open election data, and citizens engaged with the data in a community initiative. Both types of engagement share similar voluntariness characteristics; the engagement is voluntary. However, this type of OGD engagement may differ per country because it may be influenced by, for instance, the democratic culture and economic conditions of a particular country (Purwanto et al., 2020b). Moreover, two types of OGD engagement were central in the cases: government-led OGD engagement and citizen-led OGD engagement. The case selection enabled the researcher to examine whether the factors identified from the cases confirmed each other and compare the findings. For example, the financial benefit was identified only in the Hack de Valse Start case, while system reliability was found only in the Kawal Pemilu case. However, these findings can only be generalized to the contexts bound to both cases. More cases from different contexts such as different countries, different types of OGD, different types of OGD engagement may result in different outcomes.

6.3.4. The use of non-probability sampling approach

The quantitative study recruited survey respondents using a non-probability sampling approach instead of probability (random) sampling, the standard of quantitative-oriented studies. Probability sampling requires the researcher to define the study population and randomly select its members to compute the probability of each unit's inclusion (Bryman, 2012; Sue & Ritter, 2007). However, the researcher's decision is justified because the population of open data users is unknown; no database about the community of OGD users exists (Beno et al., 2017). Additionally, determining the individuals and their numbers in the population is barely possible. Although open data user communities exist in many parts of the world (Kuk & Davies, 2011), the membership of the open data user community cannot be clearly defined because the openness nature of OGD leads to use by unanticipated actors (Martin, 2014).

Moreover, non-probability sampling can be justified when the purpose of the research is exploratory (Bryman, 2012; Lehdonvirta et al., 2020) or modeling the relationships between variables (Baker et al., 2013). Since this research phase aims to test and assess a model that predicts the determinants of citizens' intention to engage with OGD, the use of a non-probability survey can be warranted. As a consequence of employing this approach, the minimum sample size in the quantitative study was determined using Hair et al.'s (2017) guidelines and Faul et al.'s (2007) software G*Power built on Cohen's (1988) work.

6.3.5. Generalization of the findings from the multivariate analysis

The first limitation of the quantitative study concerns the representativeness of the samples. Non-probability sampling approach (i.e., convenience sampling) was employed because currently, no database of OGD users exists, and the population of OGD users is unknown. Therefore, it is barely possible to make inferences whether the sample's demographics are representative of the population. However, the demographic representation of the samples is relatively similar to Jurisch et al.'s (2015) international samples: the majority of respondents' age ranges from twenty to fifty years old.

The second limitation regarding the generalization of the quantitative study results concerns the rate of usable responses, which was deemed relatively low compared to the entire collected responses (26.48%; n=166). The total number of responses collected during the study was 627. Among these responses, 471 respondents stated that they have engaged with OGD; 165 responses were dropped. 264 among 471 responses had complete data on the primary constructs tested in the study; 207 responses were removed. The researcher assumed that these incomplete responses might be due to the online survey interface's low level of user-friendliness. In addition, the Collector tool used for designing and publishing the survey was outdated; it cannot generate survey sites that accommodate the recent HTML technology featuring responsive and mobile-first websites. In the end, only 166 among 264 responses completed the survey without "not applicable" answers; 98 responses were excluded from the analysis. Including responses containing "not applicable" answers, treating them as missing data, and handling those responses using mean replacement may lead to biased inferences.

The third limitation concerning the generalization of the multivariate analysis findings is missing data on the respondents' profiles (i.e., age, gender, and education level). The researcher hypothesized that these profiles moderate the relationships between the factors and citizens' intention to engage with OGD. Therefore, a particular section of the survey that aims to elicit the profiles of the respondents was developed. However, due to General Data Protection Regulation (GDPR), the researcher provided an option for respondents not to answer these questions; only respondents who gave consent to provide privacy-related data can answer the section entirely. Consequently, missing values on the respondents' profiles are inevitable—only 134 of 166 responses have completed profile data while 32 contained missing values. The pairwise deletion approach was used to handle these particular responses. This approach aims to retain as much information as possible (Allison, 2002; Hair et al., 2017). In each analysis, the pairwise deletion approach only removes cases containing missing values in each pair of constructs. If missing values occur in unused constructs in an analysis, the analysis will use those cases for estimation purposes. Therefore, it is difficult to precisely understand the moderating effects of citizens' profiles on the relationships between the factors and citizens' intention to engage with OGD.

The fourth limitation concerns the demographic composition of the sample predominantly of Indonesian nationality (62.69%; n=84). Kruskal-Wallis H test was used to determine whether there are significant differences between groups of nationality on the determinants and behavioral intention constructs asked in the survey. The results showed no statistically significant differences in the respondents' attitudes and intentions between the Indonesian and non-Indonesian nationality groups. However, the survey responses from different nationalities cannot be compared because such comparison requires the minimum sample size of nationality responses to be met (Hair et al., 2017). For instance, responses collected from the American nationality were ten while responses from Belgians and French were both one. This limitation is due to the non-random probability approach used for collecting survey responses.

The fifth limitation is related to the types of relationships between the factors investigated in the study. This research examines whether the factors, i.e., (extrinsic and intrinsic) motivations, social factors, technical factors, and political factors, significantly influence citizens' intention to engage with OGD. Assessing the intra-relationships among the factors was outside the scope of this study. The impacts of the factors on each other were examined in this study. For example, the possible effects of the citizens' perceived technical factors (i.e., OGD quality) on the political factors (i.e., trust in OGD) were not taken into account. Similarly, the impact of the social factors (i.e., social influence) on the political factors was not considered either. However, the results of the multivariate analysis showed that the social factors highly correlate with motivations and political factors, while technical factors highly correlate with political factors.

The fifth limitation concerning the focus of this study was on the factors that influence citizens' intention to engage with OGD and not the actual engagement with OGD. This study did not focus on how the intention leads to actual engagement or other antecedents influence it. However, the researcher believes this study is the first important step before investigating the determinants related to the actual OGD engagement.

6.3.6. Missing factors excluded in the research model

Although the research model developed and tested in this study was built on a theoretical framework that stemmed from an extensive overview of open data literature, researchers can extend the factors that influence OGD citizen engagement. The results of the multivariate analysis show that the factors examined in the research model can explain 38.51% of the variance in citizens' intention to engage with OGD (see Section 5.3.3). This finding suggests that many factors unidentified and thus, untested in the model may also influence intention. One of the plausible factors that this research might have missed from the literature is the roles of technology used by citizens in an OGD engagement because engaging with OGD requires different technological skills (Janssen et al., 2012).

In this study, it is assumed that the OGD citizens engage with is technology-agnostic; citizens can use whichever technologies they can download and use OGD and create something out of it. However, citizen engagement with OGD heavily depends on these particular technologies. Thus, the perceived capabilities of the technologies might drive citizens to engage with OGD or inhibit them from engaging with OGD. Even if citizens have the necessary skills and knowledge on utilizing particular technology to exploit OGD, the technology features might restrict them from engaging with OGD continually to achieve specific objectives. These technologies have limitations. For example, citizens can use free tools such as R to conduct statistical and prediction analysis, but it might require more computer memory space over which they cannot afford to process big OGD. At the same time, proprietary technologies such as Excel cannot download a particular type of data formatted in JSON and store records that exceed a particular amount (i.e., slightly more than one million rows). Therefore, the use of these technologies may impact citizens' intention to engage with OGD. Nevertheless, the researcher does not take this assumption into account in this study.

6.4. Future research

Previous sections in this chapter reported what the researcher has achieved in this study and revealed various limitations that have restricted this study. As a result, researchers need to investigate several research themes further; this condition opens up a new avenue of OGD research. Therefore, in this section, several directions for future research agendas were proposed.

6.4.1. Evaluating the OGD-CEM model in different contexts

The OGD-CEM model was evaluated based on quantitative data collected from an international non-random sampling approach. This assessment did not focus on a specific OGD, governmental organization, or a specific OGD engagement. Governmental organizations at different administrative levels, i.e., federal or national, state or regional, and local, run OGD initiatives with different objectives (Zuiderwijk et al., 2019). Moreover, these initiatives differ from one another. Future research that focuses on examining whether the OGD-CEM model also applies to other contexts is recommended. For example, evaluating the model to a particular OGD, a particular government organization from a specific administrative level, in a particular country on different citizens (e.g., civil servants). This type of research is expected to provide insights into factors that stimulate OGD engagement in multiple domains. The research may also help better understand the dimension of OGD programs that need improvement to increase the programs' public engagement.

The case studies focused on a specific type of OGD used by citizens who did not hold any governmental positions to create artifacts such as applications that contribute to solving societal problems. The researcher studied only a relatively limited number of cases (i.e., two cases) and focused only on the determinants of citizens' intention to engage with OGD. Some of the identified factors may be typical for government-led or citizen-led engagement, yet they may not apply to other types of OGD engagement or the same type of engagement with other types of OGD. Future research that focuses on examining to which extent the identified factors also apply to other contexts is recommended. For example, cases that involve other types of OGD, other types of OGD engagement, in other countries, in other cultures, or on the local government level. Future research that replicates this study's citizen-led OGD engagement case, which is rarely found in recent literature, is also recommended. This type of research is expected to provide insights into the members of the OGD user population, their profiles, and their motivation for engaging with a particular OGD. The research may also help us to better understand the way citizens engage with OGD and create an intervention to increase the value creation of the engagement.

6.4.2. Taking intermediating factors into account

The way citizens can engage with OGD and create value-adding artifacts out of it heavily depends on the resources they possess, such as technology, skills/knowledge, time, and money spent exploring and exploiting OGD. From the technological resource perspective, technologies used to explore and exploit OGD are still developed and have limitations that may hinder citizens from engaging with OGD. The citizens' perception of the technology's capabilities in OGD exploration and exploitation might influence the citizens' intention to engage with OGD, either in a strengthening or diminishing way. From the capability perspective, the level of a citizen's skills and knowledge on exploring and exploiting OGD to create something out of it may also strengthen or decrease her or his intention to engage with OGD. Citizens have to spend their time and money to engage with OGD. However, this study did not focus on the effects of these intermediating factors on citizen engagement with OGD. Future research evaluating the OGD-CEM model to take intermediating factors into account is recommended. This type of research is expected to provide insights into the situational factors that may strengthen or reduce the OGD user motivation in engaging with OGD. The research may also help create relevant interventions to reduce the negative effects of these factors to facilitate OGD engagement.

6.4.3. Analyzing the intra-relationships among factors

This study focused only on particular relationships between the investigated factors, i.e., (extrinsic and intrinsic) motivations, social factors, technical factors, political factors, and citizens' intention to engage with OGD. The researcher did not study the intra-relationships among the factors; the effects on each other were not the focus of this study. However, the multivariate analysis results indicated that there is the possibility of intra-relationships among certain factors. For example, there may be relationships between the social factors and both motivations and political factors and those of the technical and political factors. Motivations and political factors fully mediate the relationship between social factors and citizens' intention to engage with OGD, while political factors partially mediate the technical factors – intention relationship. Future research that takes these intra-relationships among factors into account is recommended. This type of research is expected to provide insights into the interactions among the factors.

6.4.4. Investigating the actual OGD engagement

This research focused only on identifying, exploring, and assessing factors determining citizens' intention to engage with OGD. The researcher studied two cases of actual OGD engagement and assessed the OGD-CEM model using a quantitative study. In the latter study, the respondents were asked whether they had experience engaging with OGD and what type of OGD-based artifacts they created in the OGD engagement. It was assumed that these questions could indicate that the respondents truly have engaged with OGD. However, the outcomes of the respondents' engagement with OGD cannot be verified. There exists the possibility that the respondents did not actually engage with OGD or did engage with OGD but did not create OGD-based artifacts.

Furthermore, the researcher did not focus on investigating the link between the citizen's intention to engage with OGD and her or his actual OGD engagement in the latter study. The researcher neither focuses on the antecedents that influence the actual OGD engagement. It was assumed that the higher the respondents' intention would highly likely predict future OGD engagement; yet, the researcher did not empirically examine this assumption in this research. Moreover, intentions may not always lead to the expected behaviors or may do so in an inconsistent way (Bhattacherjee & Sanford, 2009). Therefore, future research that investigates the relationship between the intention to engage with OGD and actual OGD engagement using a longitudinal field survey strategy is recommended. The strategy would include an initial study focusing on the intention and its determinants, followed by another study investigating the actual OGD engagement, its link to the intention, and other determinants that may influence it. Another strategy can also be used: the inclusion of respondents based on their verifiable actual outcomes of OGD engagement. This type of research is expected to provide insights into the characteristics of citizens who have the intention to engage and actually engage with OGD. This research may also help tailor OGD programs to the specific needs of OGD user groups and carry out separate intervention plans for each group.

6.5. Reflection on this research

The reflection is divided into three themes: altruism-based OGD engagement, alternative model of OGD-CEM, and the future of OGD engagement.

6.5.1. Sense of urgency-motivated OGD engagement

The case study results show that economic factors such as monetary/financial rewards, economic motives, and potential gains did not influence citizens who engaged with OGD in the Kawal Pemilu case (see Section 4.4). On the contrary, the Kawal Pemilu's volunteers donated their money to support the initiative to, for example, buy and maintain its website domain and hosting to make it available for the public. This finding indicates that the Kawal Pemilu case is a distinctively unique/special case that contrasts with the typical OGD engagement. Generally, OGD engagement participants are extrinsically motivated to gain financial rewards. Culturally, initiatives such as Kawal Pemilu are believed to be rooted in the Indonesian "gotong royong" tradition of working together mutually and reciprocally (Purwanto et al., 2018a). In the Kawal Pemilu case, political polarization during the 2014 presidential election threatened social relationships, notably when competing candidates claimed their victories. Citizens who volunteered in Kawal Pemilu sensed the situation's urgency and worked together to provide evidence-based election results to the public. Although such initiative is rarely found in the literature, one previous open data research had a similar assumption in a different context (Khayyat & Bannister, 2017). Khayyat and Bannister (2017) revealed that the "meitheal" culture, which has a similar meaning to "gotong royong," motivates OGD co-creation in Ireland. The researcher believes that these similar cultures can be found in almost every community worldwide; this similarity opens up an avenue for future OGD research. It is crucial to investigate the relationships between cultures and sense of urgency and their impacts on citizens' motivation in engaging with OGD and understand whether particular cultures are more stimulating for citizen-led engagement.

6.5.2. An alternative model of OGD-CEM

The OGD-CEM model shows that extrinsic and intrinsic motivations toward the engagement and perceived political factors toward OGD and its provider determine citizens' behavioral intention to engage with OGD (see Section 5.4). Contrary to previous research, the model does not postulate that social influence and three dimensions of IS quality (i.e., data quality, system quality, and service quality) affect citizens' behavioral intention. However, the statistical analysis results indicate a plausible explanation: technical factors have both direct and indirect effects (partial effect) on intention via political factors. In contrast, political factors have a partial effect on intention via motivations. Since this study aims to examine whether the factors significantly influence citizens' intention to engage with OGD, assessing the intra-relationships among the factors was outside the scope of this study (see Section 6.3.5). Based on the results of the multivariate analysis using PLS-SEM, Figure 6.1 best describes such a plausible alternative model that takes the intra-relationships among factors into account.

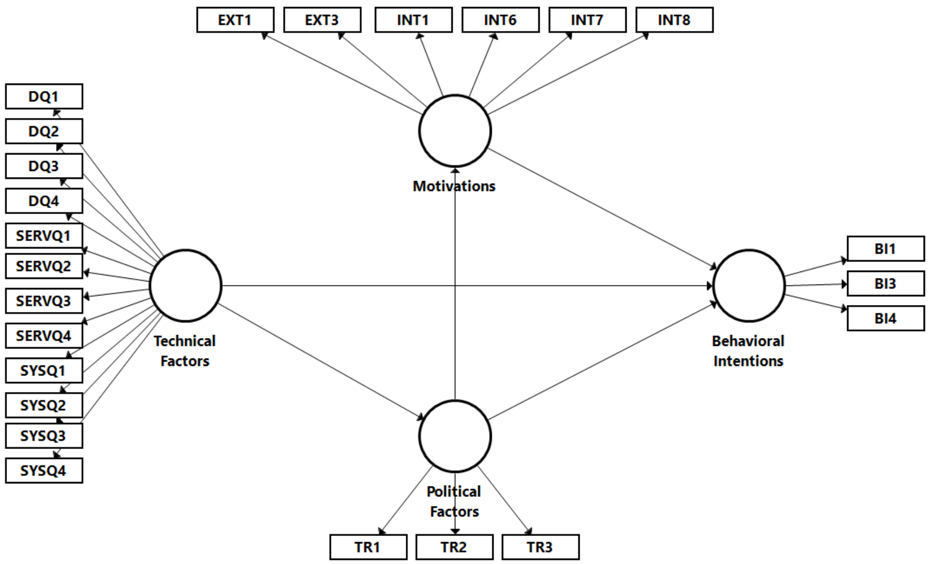


Figure 6.1. The alternative model of OGD-CEM.

The alternative model of OGD-CEM shows that citizens' behavioral intention to engage with OGD is determined by their motivations toward the engagement, perceived political factors toward OGD and its provider, and perceived technical factors of OGD. It also postulates that political factors partially mediate the relationship between technical factors and citizens' intention to engage with OGD while motivations partially mediate the political factors – intention relationship. Future research that plans to employ the model can consider these intra-relationships among factors when investigating the determinants of citizens' intention.

6.5.3. Future of OGD engagement

This study was initiated with an assumption that more OGD engagement is better for society, and consequently, opening more public data that stimulates engagement is also better for the future. On the one hand, the outcomes of OGD engagement can be misleading when users misinterpret data (Janssen et al., 2012). On the other hand, the researcher believes that OGD engagement outcomes will naturally be improved when made available to the public. A rational conversation about the outcomes will occur, and corrected outcomes will prevail, particularly when this assumption is reflected in the current global situation with the Covid-19 pandemic. More engagement with OGD has led to more publicly available outcomes that offer insights to society and governments on tackling the pandemic (Kim, 2021). When the Covid-19 virus started to spread worldwide, and the United Nation's World Health Organization announced its outbreak as a pandemic, citizens and scientists from different countries demanded their governments to make the pandemic situation available to the public. Various initiatives taken almost globally by citizens from different countries have witnessed this situation; they built portals on top of the opened pandemic data sets, enabling experts in epidemiology to analyze and provide insights to the public regarding the pandemic. As a result, citizen-led OGD engagement initiatives have sprung up. For instance, at the international level, Worldometers' Covid-19 Data manually aggregates data from the official open Covid-19 data sources worldwide and provides live statistics for a global audience. Another example concerns KawalCOVID19 from Indonesia initiated by the same founder of Kawal Pemilu in the case study. KawalCOVID19 scrapes data from the official portal published by the COVID-19 Task Force and the local government portals and compares the situation. It revealed that the Covid-19 data published at the national level 19 lags behind those published at the regional and local levels.

The researcher believes that citizens will proactively demand the opening of public data in the future and contribute to solving societal problems using OGD. More advanced technologies such as artificial intelligence will likely be applied in citizen-led OGD engagement initiatives. More tools for developing such technology (e.g., R language, python) are becoming free to use by anyone. These initiatives will be more advanced and complex and generate more meaningful information that can impact public policy. On the other hand, government-led OGD engagement such as hackathons will likely be permanently stopped and hypothetically continued in different forms. Hackathons typically involve the physical gathering of participants in one location for a particular period (e.g., one day, 24 hours, or more). As governments substantially focus on preventing the pandemic from spreading through the congregation of people, hackathons involving physical gatherings might not be ideal for promoting the use of OGD. It is plausible that hackathons will likely be held online, though supporting empirical evidence is currently scarce.

Reference List

ACE Electoral Knowledge Network. (2014). 2014 Indonesian Elections Case Study: Uploading millions of C1 result forms and achieving a crowdsourced outcome. Retrieved from http://aceproject.org/ace-en/topics/et/ety/copy_of_ety06/mobile_browsing

Achen, C. H. (1982). *Interpreting and Using Regression*. Newbury Park, California: Sage Publications, Inc.

Afful-Dadzie, E., & Afful-Dadzie, A. (2017). Open Government Data in Africa: A preference elicitation analysis of media practitioners. *Government Information Quarterly, 34*(2), 244-255.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2005). *Attitudes, Personality and Behavior* (2 ed.). Berkshire, England: Open University Press.

Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*: Prentice-Hall.

Al-Lozi, E., & Papazafeiropoulou, A. (2012). Intention-Based Models: The Theory of Planned Behavior Within the Context of IS. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information Systems Theory: Explaining and Predicting Our Digital Society* (Vol. 2). New York, USA: Springer.

Alathur, S., Ilavarasan, P. V., & Gupta, M. P. (2014). Determinants of citizens' electronic participation: Insights from India. *Transforming Government: People, Process and Policy, 8*(3), 447-472. doi:10.1108/TG-09-2013-0034

Alathur, S., Ilavarasan, P. V., & Gupta, M. P. (2016). Determinants of e-participation in the citizens and the government initiatives: Insights from India. *Socio-Economic Planning Sciences, 55*, 25-35. doi:10.1016/j.seps.2016.04.005

Alexopoulos, C., Zuiderwijk, A., Charalabidis, Y., Loukis, E., & Janssen, M. (2014, 27-30 October). *Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media.* Paper presented at the International Conference on Electronic Government 2014, Guimarães, Portugal.

Allison, P. D. (2002). *Missing Data*. Thousand Oaks, CA: Sage Publications, Inc.

Allison, P. D. (2003). Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology, 112*(4), 545–557. doi:10.1037/0021-843X.112.4.545

Alvesson, M. (2003). Beyond Neopositivists, Romantics, and Localists: A Reflexive Approach to Interviews in Organizational Research. *The Academy of Management Review, 28*(1), 13-33 doi:10.2307/30040687

Armstrong, J. S., & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research, 14*(3), 396-402.

Arnstein, S. R. (1969). A Ladder Of Citizen Participation. *Journal of the American Institute of Planners, 35*(4), 216-224.

Atkinson, R., & Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update, 33*(1), 1-4.

Ayele, W., Juell-Skielse, G., Hjalmarsson, A., & Johannesson, P. (2015, 6-9 July). *Evaluating Open Data Innovation: A Measurement Model for Digital Innovation Contests.* Paper presented at the 19th Pacific Asia Conference on Information Systems (PACIS), Singapore.

Bailey, J. E., & Pearson, S. W. (1983). Development of a Tool for Measuring and Analyzing Computer User Satisfaction. *Management Science, 29*(5), 530-545.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., . . . Tourangeau, R. (2013). Summary Report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology, 1*(2), 90–143. doi:10.1093/jssam/smt008

Bandura, A. (1982). Self-Efficacy Mechanism in Human Agency. *Amercan Psychologist, 37*(2), 122-147.

Barki, H., & Hartwick, J. (1989). Rethinking the Concept of User Involvement. *MIS quarterly, 13*(1), 53-63.

Becker, J.-M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats. *MIS quarterly, 37*(3), 665-694.

Bélanger, F., & Carter, L. (2008). Trust and risk in e-government adoption. *Journal of Strategic Information Systems, 17*(2), 165-176.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS quarterly, 11*(3), 369-386.
Benitez-Paez, F., Degbelo, A., Trilles, S., & Huerta, J. (2018). Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain A Data User's Perspective. *International Journal of Geo-Information, 7*(1).
Beno, M., Figl, K., Umbrich, J., & Polleres, A. (2017). *Open Data Hopes and Fears: Determining the Barriers of Open Data*. Paper presented at the 2017 Conference for E-Democracy and Open Government (CeDEM), Krems, Austria.
Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351 - 368. doi:10.1093/pan/mpr057
Bhattacherjee, A., & Sanford, C. (2009). The intention–behaviour gap in technology usage: the moderating role of attitude strength. *BEHAVIOUR & INFORMATION TECHNOLOGY, 28*(4), 389-401.
Biernacki, P., & Waldorf, D. (1981). Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research, 10*(2), 141-163.
Boudreau, K. J., & Lakhani, K. R. (2009). How to manage outside innovation. *MIT Sloan Management Review, 50*(4), 69-76.
Brady, H. E., Verba, S., & Schlozman, K. L. (1995). Beyond Ses: A Resource Model of Political Participation *The American Political Science Review, 89*(2), 271-294.
Brajawidagda, U., & Chatfield, A. T. (2014). *Roles of Social Media in Open Data Environments: A Case Study of the 2014 Indonesian Presidential Election Voting Results.* Paper presented at the 25th Australasian Conference on Information Systems, Auckland, New Zealand.
Briscoe, G., & Mulligan, C. (2014). *Digital Innovation: The Hackathon Phenomenon*. Retrieved from London, UK:
Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2 ed.). New York, USA: The Guilford Press.
Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford, UK: Oxford University Press.
Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life* (1 ed.). London, UK: Heinemann Educational Books Ltd.
Callegaro, M., & DiSogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly, 72*(5), 1008–1032. doi:10.1093/poq/nfn065
Canares, M. P. (2014, 27-30 October). *Opening the Local: Full Disclosure Policy and its Impact on Local Governments in the Philippines.* Paper presented at the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV), Guimaraes, Portugal.
Carter, L., & Bélanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors. *Information Systems Journal, 15*(1), 5-25. doi:10.1111/j.1365-2575.2005.00183.x
Cavaye, A. L. M. (1996). Case study research: a multi-faceted research approach for IS. *Information Systems Journal, 6*(3), 227-242. doi:10.1111/j.1365-2575.1996.tb00015.x
Charalabidis, Y., Loukis, E., & Alexopoulos, C. (2014). *Evaluating Second Generation Open Government Data Infrastructures Using Value Models*. Paper presented at the 47th Hawaii International Conference on System Sciences (Hicss).
Charalabidis, Y., Loukis, E., Alexopoulos, C., & Lachana, Z. (2019). *The Three Generations of Electronic Government: From Service Provision to Open Data and to Policy Analytics*. Paper presented at the 18th IFIP WG 8.5 International Conference (EGOV 2019), San Benedetto Del Tronto, Italy.
Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers* (2 ed.). Hoboken, NJ: John Wiley & Sons, Inc.
Chin, W. W. (1998a). Issues and Opinion on Structural Equation Modeling. *MIS quarterly, 22*(1), vii-xvi.
Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–358). Mahwah, NJ: Lawrence Erlbaum Associates.
Chin, W. W. (2010). How to Write Up and Report PLS Analyses. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 655-690). Berlin, Germany: Springer-Verlag.

Chin, W. W., & Dibbern, J. (2010). An Introduction to a Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services Between Germany and the USA. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 171-194). Berlin, Germany: Springer-Verlag.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research, 14*(2), 189-217. doi:10.1287/isre.14.2.189.16018

Choi, J., & Tausczik, Y. (2017). *Characteristics of Collaboration in the Emerging Practice of Open Data Analysis*. Paper presented at the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2017), Portland,USA.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2015). *Research Methods, Design, and Analysis* (12 ed.). Essex, UK: Pearson Education Limited.

Chua, W. F. (1986). Radical Developments in Accounting Thought. *The Accounting Review, 61*(4), 601-632.

Cogan, A., & Sharpe, S. (1986). The theory of citizen involvement. In *Planning analysis: The theory of citizen participation*: University of Oregon.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.): Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Concilio, G., Molinari, F., & Morelli, N. (2017). *Empowering Citizens with Open Data by Urban Hackathons*. Paper presented at the 7th International Conference for E-Democracy and Open Government (CeDEM), Krems, Austria.

Coppock, A., & McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics, 6*(1), 1-14. doi:10.1177/2053168018822174

Cranefield, J., Robertson, O., & Oliver, G. (2014). *Value in the MASH: Exploring the Benefits, Barriers and Enablers of Open Data Apps*. Paper presented at the 22nd European Conference on Information Systems (ECIS 2014), Tel Aviv, Israel.

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5 ed.). Los Angeles, USA: SAGE Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3 ed.). Los Angeles, USA: SAGE Publications, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Crusoe, J., Simonofski, A., Clarinval, A., & Gebka, E. (2019). *The Impact of Impediments on Open Government Data Use: Insights from Users*. Paper presented at the 13th International Conference on Research Challenges in Information Science (RCIS), Brussels, Belgium.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety: The Experience of Play in Work and Games*. San Francisco: Jossey-Bass, Inc.

Cunningham, J. V. (1972). Citizen Participation in Public Affairs. *Public Administration Review, 32*, 589-602.

da Silva Craveiro, G., & Albano, C. (2017). Open data intermediaries: coproduction in budget transparency. *Transforming Government: People, Process and Policy, 11*(1), 119-131.

Danziger, J. N., & Andersen, K. V. (2002). The Impacts of Information Technology on Public Administration: An Analysis of Empirical Research from the "Golden Age" of Transformation. *International Journal of Public Administration, 25*(5), 591-627. doi:10.1081/PAD-120003292

Davies, T. (2010). *Open data, democracy and public sector reform: A look at open government data use from data.gov.uk.* (Master). University of Oxford, Oxford, UK. Retrieved from www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS quarterly, 13*(3), 319-340.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. New York, NY: Cambridge University Press.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly, 33*(1), 15-27. doi:10.1016/j.giq.2016.01.003

Day, D. (1987). Citizen Participation in the Planning Process: An Essentially Contested Concept? *Journal of Planning Literature, 11*(3), 421-434.

de Bruijne, M., & Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review, 31*(4), 482-504. doi:10.1177/0894439313483976

de Deus Ferreira, G., & Farias, J. S. (2018). The Motivation to Participate in Citizen-Sourcing and Hackathons in the Public Sector. *Brazilian Administration Review, 15*.

de Kool, D., & Bekkers, V. (2015). The Perceived Impact of Open Inspection Data on the Quality of Education in Dutch Primary Schools: A Parent Perspective. *Social Science Computer Review, 33*(5), 645-659. doi:10.1177/0894439314560853

de Kool, D., & Bekkers, V. (2016). The perceived value-relevance of open data in the parents' choice of Dutch primary schools. *International Journal of Public Sector Management, 29*(3), 271-287. doi:10.1108/Ijpsm-02-2016-0022

Deci, E. L. (2004). Intrinsic Motivation. In W. E. Craighead & C. B. Nemeroff (Eds.), *The Concise Corsini Encyclopedia of Psychology and Behavioral Science* (3 ed., pp. 494-495). New Jersey, USA: John Wiley & Sons, Inc.

DeLone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Journal of management information systems, 3*(1), 60-95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems, 19*(4), 9-30. doi:10.1080/07421222.2003.11045748

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science, 40*, 434–449. doi:10.1007/s11747-011-0300-3

Dietrich, D. (2015). The Role of Civic Tech Communities in PSI Reuse and Open Data Policies. In *European Public Sector Information Platform Topic Report 05/2015*: European Union.

Dijkstra, T. K., & Henseler, J. (2015). Consistent Partial Least Squares Path Modeling. *MIS quarterly, 39*(2), 297-316.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail and Mixed-mode Surveys: The Tailored Design Method* (4 ed.). New Jersey, USA: JohnWiley & Sons, Inc.

Dittus, M., Quattrone, G., & Capra, L. (2016). *Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale*. Paper presented at the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2016), San Francisco,USA.

Doll, W. J., & Torkzadeh, G. (1988). The Measurement of End-User Computing Satisfaction. *MIS quarterly, 12*(2), 259-274.

dos Santos Brito, K., dos Santos Neto, M., da Silva Costa, M. A., Garcia, V. C., & de Lemos Meira, S. R. (2014). *Using Parliamentary Brazilian Open Data to Improve Transparency and Public Participation in Brazil*. Paper presented at the 15th Annual International Conference on Digital Government Research, Aguascalientes, Mexico.

Dubé, L., & Paré, G. (2003). Rigor in information systems positivist case research: current practices, trends, and recommendations. *MIS quarterly, 27*(4), 597-636.

Edelenbos, J., van Meerkerk, I., & Schenk, T. (2018). The Evolution of Community Self-Organization in Interaction With Government Institutions: Cross-Case Insights From Three Countries. *American Review of Public Administration, 48*(1), 52-66. doi:10.1177/0275074016651142

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science, 1*(1), 54-77.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review, 14*(4), 532-550. doi:10.2307/258557

Enders, C. K. (2013). Analyzing Structural Equation Models with Missing Data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2 ed., pp. 493-520). USA: Information Age Publishing Inc.

Enticott, G., Boyne, G. A., & Walker, R. M. (2009). The Use of Multiple Informants in Public Administration Research: Data Aggregation Using Organizational Echelons. *Journal of Public Administration Research and Theory, 19*(2), 229-253.
Evans, J. R., & Mathur, A. (2018). The value of online surveys: a look back and a look ahead. *Internet Research, 28*(4), 854-887. doi:10.1108/IntR-03-2018-0089
Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research, 69*(10), 4565-4582. doi:10.1016/j.jbusres.2016.03.050
Eysenbach, G. (2004). Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research, 6*(3). doi:10.2196/jmir.6.3.e34
Eysenbach, G. (2012). Correction: Improving the Quality of Web Surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research, 14*(1). doi:10.2196/jmir.2042
Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. New York: Oxford University Press, Inc.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:10.3758/BF03193146
Field, A. (2009). *Discovering Statistics Using SPSS* (3 ed.): SAGE Publications Ltd.
Finney, S. J., & DiStefano, C. (2013). Nonnormal and Categorical Data in Structural Equation Modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2 ed., pp. 439-492). USA: Information Age Publishing Inc.
Fishbein, M., & Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York, USA: Taylor and Francis Group, LLC.
Fitriani, W. R., Hidayanto, A. N., Sandhyaduhita, P. I., Purwandari, B., & Kosandi, M. (2019). Determinants of Continuance Intention to Use Open Data Website: An Insight from Indonesia. *Pacific Asia Journal of the Association for Information Systems, 11*(2), 96-120. doi:10.17705/1pais.11205
Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry, 12*(2), 219-245. doi:10.1177/1077800405284363
Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*(1), 39-50. doi:10.1177/002224378101800104
Friese, S. (2012). *Qualitative Data Analysis with ATLAS.ti*. London, UK: SAGE Publications Ltd.
Gable, G. G., Sedera, D., & Chan, T. (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the Association of Information Systems, 9*(7), 377-408. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-57149105774&partnerID=40&md5=0673337a322d4a677af637416ae64a15
Gama, K. (2017). *Crowdsourced Software Development in Civic Apps - Motivations of Civic Hackathons Participants*. Paper presented at the 19th International Conference on Enterprise Information Systems, Porto, Portugal.
Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models*. Asheboro, NC: Statistical Publishing Associates.
Gascó-Hernández, M., Martin, E. G., Reggi, L., Pyo, S., & Luna-Reyes, L. F. (2018). Promoting the use of open government data: Cases of training and engagement. *Government Information Quarterly, 35*(2), 233-242.
Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega, 28*(6), 725-737. doi:10.1016/S0305-0483(00)00021-9
Gefen, D., Rigdon, E. E., & Straub, D. (2011). An Update and Extension to SEM Guidelines for Administrative and Social Science Research. *MIS quarterly, 35*(2), iii-xiv.
Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems, 4*(7).
Geisser, S. (1974). A predictive approach to the random effect model *Biometrika, 61*(1), 101–107. doi:10.1093/biomet/61.1.101
Gerbing, D. W., & Anderson, J. C. (1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. *Journal of Marketing Research, 25*(2), 186-192. doi:10.1177/002224378802500207

Gold, R. L. (1958). Roles in Sociological Field Observations. *Social Forces, 36*(3), 217-223.
Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead...long live the paradigm: the legacy of Burrell and Morgan. *Omega, 28*(3), 249-268. doi:10.1016/S0305-0483(99)00042-0
Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly, 32*(4), 441-452. doi:dx.doi.org/10.1016/j.giq.2015.09.001
Graft, A., Verhulst, S., & Young, A. (2016). Indonesia's Kawal Pemilu. Elections: Free, Fair, and Open Data. In *The Global Impact of Open Data: Key Findings from Detailed Case Studies Around the World* (1 ed., pp. 97-112): O'Reilly Media Inc.
Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2 ed.). New Jersey, USA: John Wiley & Sons, Inc.
Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 105-117). Thousand Oaks, CA: SAGE.
Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday, 16*(2). doi:10.5210/fm.v16i2.3316.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis: A Global Perspective* (7 ed.). Essex, England: Pearson Education Limited.
Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (1 ed.). Los Angeles: SAGE Publications, Inc.
Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2 ed.). Los Angeles: SAGE Publications, Inc.
Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-152. doi:10.2753/MTP1069-6679190202
Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2-24. doi:10.1108/EBR-11-2018-0203
Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Los Angeles, USA: SAGE Publications, Inc.
Harrington, D. (2009). *Confirmatory Factor Analysis*. New York, USA: Oxford University Press, Inc.
Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet, 4*(4), 900-928.
Hars, A., & Ou, S. (2002). Working for Free? Motivations for Participating in Open-Source Projects. *International Journal of Electronic Commerce, 6*(3), 25-39. doi:10.1080/10864415.2002.11044241
Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination*: SAGE.
Hartmann, S., Mainka, A., & Stock, W. G. (2016). Opportunities and Challenges for Civic Engagement: A Global Investigation of Innovation Competitions. *International Journal of Knowledge Society Research, 7*(3), 1-15. doi:10.4018/IJKSR.2016070101
Hellberg, A.-S., & Hedström, K. (2015). The story of the sixth myth of open data and open government. *Transforming Government: People, Process and Policy, 9*(1), 35 - 51.
Henseler, J., & Chin, W. W. (2010). A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Path Modeling. *Structural Equation Modeling, 17*(1), 82-109. doi:10.1080/10705510903439003
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*, 115–135. doi:10.1007/s11747-014-0403-8
Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405-431. doi:10.1108/IMR-09-2014-0304
Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In R. R. Sinkovics & P. N. Ghauri (Eds.), *New Challenges to International Marketing* (Vol. Advances in International Marketing, pp. 277-319). Bingley, UK: Emerald Group Publishing Limited.
Herriott, R. E., & Firestone, W. A. (1983). Multisite Qualitative Policy Research: Optimizing Description and Generalizability. *Educational Researcher, 12*(2), 14-19.

Hill, R. (1998). What Sample Size is "Enough" in Internet Survey Research? *Interpersonal Computing and Technology: An Electronic Journal for the21st Century, 6*(3-4).
Hirschheim, R., & Klein, H. K. (2003). Crisis in the IS Field? A Critical Reflection on the State of the Discipline. *Journal of the Association for Information Systems, 4*(10), 237-293. doi:10.17705/1jais.00037
Hivon, J., & Titah, R. (2017). Conceptualizing citizen participation in open data use at the city level. *Transforming Government: People, Process and Policy, 11*(1), 99-118. doi:10.1108/TG-12-2015-0053
Hjalmarsson, A., Johannesson, P., Juell-Skielse, G., & Rudmark, D. (2014). *Beyond innovation contests: A framework of barriers to open innovation of digital services*. Paper presented at the 22nd European Conference on Information Systems (ECIS 2014), Tel Aviv, Israel.
Höck, C., Ringle, C. M., & Sarstedt, M. (2010). Management of multi-purpose stadiums: importance and performance measurement of service interfaces. *International Journal of Services Technology and Management, 14*(2-3), 188–207. doi:10.1504/IJSTM.2010.034327
Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival*: The McGraw-Hill Companies, Inc.
Hofstede Insights. (2018a). Country Comparison - Indonesia. Retrieved from https://www.hofstede-insights.com/country-comparison/indonesia/
Hofstede Insights. (2018b). Country Comparison - The Netherlands. Retrieved from https://www.hofstede-insights.com/country-comparison/the-netherlands/
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*(1-2), 14-40.
Hox, J. J., & Bechger, T. M. (1998). An Introduction to Structural Equation Modeling. *Family Science Review, 11*(4), 354-373.
Hoyle, R. H. (1995). The Structural Equation Modeling Approach: Basic Concepts and Fundamental Issues. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 1-15). California, USA: Sage Publications, Inc.
Hoyle, R. H. (2012). Introduction and Overview. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 3-16). New York, USA: The Guilford Press.
Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice, 12*(1), 4-16.
Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal, 20*(2), 195-204.
Hutter, K., Füller, J., & Koch, G. (2011). *Why citizens engage in open government platforms?* Paper presented at the INFORMATIK 2011 Jahrestagung der Gesellschaft für Informatik, Berlin, Germany.
Hwang, M. I., & Thorn, R. G. (1999). The effect of user engagement on system success: a meta-analytical integration of research findings. *Information & Management, 35*(4), 229-236.
INTOSAI. (2013). *ISSAI 300 - Fundamental Principles of Performance Auditing*. Copenhagen: INTOSAI Professional Standards Committee
Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal, 39*, 2255–2276. doi:10.1002/smj.2904
Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education, 38*(12), 1217-1218. doi:10.1111/j.1365-2929.2004.02012.x
Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly, 28*(4), 446–456. doi:10.1016/j.giq.2011.01.004
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. doi:10.1080/10580530.2012.716740
Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review, 32*(5), 694-711. doi:dx.doi.org/10.1177/0894439314525902
Jarke, J. (2019). Open government for all? Co-creating digital public services for older adults through data walks. *Online Information Review, 43*(6), 1003-1020. doi:10.1108/OIR-02-2018-0059

Johnson, P. A., & Robinson, P. J. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement? *Review of Policy Research, 31*(4), 349-357. doi:10.1111/ropr.12074

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26. doi:10.3102/0013189X033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research, 1*(2), 112-133. doi:10.1177/1558689806298224

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409–426. doi:10.1007/BF02291366

Juell-Skielse, G., Hjalmarsson, A., Johannesson, P., & Rudmark, D. (2014). *Is the Public Motivated to Engage in Open Data Innovation?* Paper presented at the 13th International Conference on Electronic Government Dublin, Ireland.

Jurisch, M. C., Kautz, M., Wolf, P., & Krcmar, H. (2015). *An International Survey of the Factors Influencing the Intention to use Open Government*. Paper presented at the 48th Hawaii International Conference on System Sciences (HICSS), Kauai, Hawaii.

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The Effect of Invitation Design on Web Survey Response Rates. *Social Science Computer Review, 30*(3), 339-349. doi:10.1177/0894439311419084

Kappelman, L. A., & McLean, E. R. (1992). Promoting information system success: The respective roles of user participation and user involvement. *Journal of Information Technology Management, 3*(1), 1-12.

Kassen, M. (2018). Open data and its intermediaries: a cross-country perspective on participatory movement among independent developers. *Knowledge Management Research & Practice, 16*(3), 327-342.

Kawal Pemilu. (2014). The Design and Architecture of kawalpemilu.org. Retrieved from https://github.com/kawalpemilu/kawalpemilu2014/blob/master/doc/20140809-DesignDoc.pdf

Kettinger, W. J., & Lee, C. C. (1997). Pragmatic Perspectives on the Measurement of Information Systems Service Quality. *MIS quarterly, 21*(2), 223-240.

Khan, S., & VanWynsberghe, R. (2008). Cultivating the Under-Mined: Cross-Case Analysis as Knowledge Mobilization. *Forum: Qualitative Social Research, 9*(1). doi:10.17169/fqs-9.1.334

Khayyat, M., & Bannister, F. (2017). Towards a model for facilitating and enabling co-creation using open government data. *Information Polity, 22*(4), 211-231. doi:10.3233/IP-170406

Khurshid, M. M., Zakaria, N. H., Rashid, A., & Shafique, M. N. (2018). Examining the Factors of Open Government Data Usability From Academician's Perspective. *International Journal of Information Technology Project Management, 9*(3), 72-85.

Kim, H. (2021). Lesson Learned from the Power of Open Data: Resolving the Mask Shortage Problem Caused by COVID-19 in South Korea. *Sustainability, 13*(1). doi:10.3390/su13010278

King, C., & Cruickshank, M. (2012). Building capacity to engage: community engagement or government engagement? *Community Development Journal, 47*(1), 5-28. doi:10.1093/cdj/bsq018

King, G. (1986). How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science, 30*(3), 666-687.

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology, 25*(4), 615-624. doi:10.1002/acp.1730

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology, 51*(1), 7-15.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. In *EBSE Technical Report EBSE-2007-01* (pp. 65). Keele, UK: Keele University.

Klein, H. K., & Myers, M. D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS quarterly, 23*(1), 67-93. doi:10.2307/249410

Kline, R. B. (2016). Regression Fundamentals. In *Principles and Practice of Structural Equation Modeling* (4 ed.). New York, USA: The Guilford Press.

Komisi Pemilihan Umum. (2014a). *Informasi PPWP 2014*. Jakarta: Komisi Pemilihan Umum
Komisi Pemilihan Umum. (2014b). *Peraturan Komisi Pemilihan Umum Nomor 21 Tahun 2014 tentang Rekapitulasi Hasil Penghitungan Perolehan Suara dan Penetapan Hasil Pemilihan Umum Presiden dan Wakil Presiden serta Penetapan Pasangan Calon Presiden dan Wakil Presiden Terpilih Tahun 2014*. (21). Jakarta: Komisi Pemilihan Umum
Kuk, G., & Davies, T. (2011). *The Roles of Agency and Artifacts in Assembling Open Data Complementarities*. Paper presented at the International Conference on Information Systems 2011, Shanghai, China.
Lassinantti, J., Bergvall-Kåreborn, B., & Ståhlbröst, A. (2014). Shaping Local Open Data Initiatives: Politics and Implications. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 17-33. doi:10.4067/S0718-18762014000200003
LeCompte, M. D., & Goetz, J. P. (1982). Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research, 52*(1), 31-60. doi:10.3102/00346543052001031
Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS quarterly, 13*(1), 33-50.
Lee, A. S. (1999). Rigor and Relevance in MIS Research: Beyond the Approach of Positivism Alone. *MIS quarterly, 23*(1), 29-33. doi:10.2307/249407
Lee, A. S., & Baskerville, R. L. (2003). Generalizing Generalizability in Information Systems Research. *Information Systems Research, 14*(3), 221-243. doi:10.1287/isre.14.3.221.16560
Lee, G., Benoit-Bryan, J., & Johnson, T. P. (2011). Survey Research in Public Administration: Assessing Mainstream Journals with a Total Survey Error Framework. *Public Administration Review, 72*(1), 87-97. doi: 10.111/j.1540-6210.2011.02482.x
Lehdonvirta, V., Oksanen, A., Räsänen, P., & Blank, G. (2020). Social Media, Web, and Panel Surveys: Using Non-Probability Samples in Social and Policy Research. *Policy & Internet*. doi:10.1002/poi3.238
Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 5-55.
Lim, M. (2014). Is it a good day for (good) stats? #pilpres2014 #kawalsuara #quickcount #citizensourcedcount. In.
Lim, M. (2017). Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. *Critical Asian Studies, 49*(3), 411-427. doi:10.1080/14672715.2017.1341188
Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly, 29*(4), 446-454. Retrieved from <Go to ISI>://WOS:000309308300002
Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association, 83*(404), 1198-1202.
Little, R. J. A., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data* (3 ed.). Hoboken, NJ: John Wiley & Sons, Inc.
Liu, M., & Wronski, L. (2018). Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. *Social Science Computer Review, 36*(1), 116-124. doi:10.1177/0894439317695581
Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly, 32*(3), 323-332. doi:10.1016/j.giq.2015.05.006
Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. *Journal of Data and Information Quality, 1*(1), 2-22. doi:10.1145/1515693.1516680
Martin, C. (2014). Barriers to the Open Government Data Agenda Taking a Multi-Level Perspective. *Policy & Internet, 6*(3), 217-240.
Maruyama, M., Douglas, S., & Robertson, S. (2013). *Design Teams as Change Agents: Diplomatic Design in the Open Data Movement*. Paper presented at the 46th Hawaii International Conference on System Sciences (Hicss).
McDermott, P. (2010). Building open government. *Government Information Quarterly, 27*(4), 401-413. doi:10.1016/j.giq.2010.07.002

Mellahi, K., & Harris, L. C. (2016). Response Rates in Business and Management Research: An Overview of Current Practice and Suggestions for Future Direction. *British Journal of Management, 27*(2), 426–437. doi:10.1111/1467-8551.12154

Mercado-Lara, E., & Ramon Gil-Garcia, J. (2014). *Open government and data intermediaries: The case of aiddata.* Paper presented at the ACM International Conference Proceeding Series.

Mingers, J. (2004). Real-izing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems. *Information and Organization, 14*(2), 87-103. doi:10.1016/j.infoandorg.2003.06.001

Molla, A., & Licker, P. S. (2001). E-Commerce Systems Success: An Attempt to Extend and Respecify the Delone and McLean Model of IS Success. *Journal of Electronic Commerce Research, 2*(4), 131-141.

Nam, T. (2012). Suggesting frameworks of citizen-sourcing via Government 2.0. *Government Information Quarterly, 29*(1), 12-20. doi:10.1016/j.giq.2011.07.005

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 263-280. doi:10.1002/ejsp.2420150303

Nugroho, R. P., Zuiderwijk, A., Janssen, M., & Jong, M. d. (2015). A comparison of national open data policies lessons learned. *Transforming Government: People, Process and Policy, 9*(3), 286-308. doi:dx.doi.org/10.1108/TG-03-2014-0008

Obama, B. (2009). Transparency and Open Government. Retrieved from https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

Ojo, A., Porwol, L., Waqar, M., Stasiewicz, A., Osagie, E., Hogan, M., . . . Zeleti, F. A. (2016). *Realizing the Innovation Potentials from Open Data: Stakeholders' Perspectives on the Desired Affordances of Open Data Environment*. Paper presented at the PRO-VE 2016: Collaboration in a Hyperconnected World, Porto, Portugal.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems, 10*(26). Retrieved from http://sprouts.aisnet.org/10-26

Onderwijs Inspectie. (2015). *Dutch Inspectorate of Education: Background information and working method of the inspectorate*. Utrecht, Netherlands: Inspectie van het Onderwijs Retrieved from https://www.sici-inspectorates.eu/getattachment/ab42ac74-fe55-45af-9ecc-85a757003fc6/Profiel-Inspectie-van-het-Onderwijs_13-10-2015.pdf;.jpg;.aspx

Onderwijs Inspectie. (2019). Onderwijsdata | Trends en ontwikkelingen | Inspectie van het onderwijs. Retrieved from https://www.onderwijsinspectie.nl/trends-en-ontwikkelingen/onderwijsdata

Open State Foundation. (2018a). Challenges en Data - Hack de Valse Start. Retrieved from https://hackdevalsestart.nl/challenges/

Open State Foundation. (2018b). 'Next STEP' wint hackathon over kansenongelijkheid - Hack de Valse Start. Retrieved from https://hackdevalsestart.nl/2018/03/03/next-step-wint-hackathon-over-kansenongelijkheid/

Orbell, S., Hodgkins, S., & Sheeran, P. (1997). Implementation Intentions and the Theory of Planned Behavior. *Personality and Social Psychology Bulletin, 23*(9), 945-954. doi:10.1177/0146167297239004

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations Research Approaches and Assumptions. *Information Systems Research, 2*(1), 1-28.

Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L., & Ojo, A. (2017). *Usability Evaluation of an Open Data Platform*. Paper presented at the 18th Annual International Conference on Digital Government Research, Staten Island, USA.

Osborne, J. W. (2013). *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data*. California, USA: SAGE Publications, Inc.

Osborne, S. P. (2006). The New Public Governance. *Public Management Review, 8*(3), 377-387. doi:10.1080/14719030600853022

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing, 49*(4), 41-50.

Parycek, P., Höchtl, J., & Ginner, M. (2014). Open Government Data Implementation Evaluation. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 80-99. doi:10.4067/S0718-18762014000200007

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of management information systems, 24*(3), 45-77. doi:10.2753/MIS0742-1222240302

Peixoto, T. (2013). The Uncertain Relationship Between Open Data and Accountability: A Response to Yu and Robinson's The New Ambiguity of "Open Government". *Ucla Law Review, 60*(200). Retrieved from http://ssrn.com/abstract=2264369

Pervan, G., & Maimbo, H. (2005). *Designing a case study protocol for application in IS research*. Paper presented at the 9th Pacific Conference on Information Systems, Bangkok, Thailand.

Pinsonneault, A., & Kraemer, K. (1993). Survey Research Methodology in Management Information Systems: An Assessment. *Journal of management information systems, 10*(2), 75-105. doi:10.1080/07421222.1993.11518001

Porwol, L., Ojo, A., & Breslin, J. (2013). *On The Duality of E-Participation – Towards a foundation for Citizen-Led Participation*. Paper presented at the The 2nd Joint International Conference on Electronic Government and the Information Systems Perspective and International Conference on Electronic Democracy, Prague, Czech Republic.

Purbo, O. W., Deograt, G., Satriyanto, R., Ferdinand, A., Gesang, P., Kesuma, R., & Silaen, K. E. (2019). KawalPilpres2019: a highly secured real count voting escort architecture. *TELKOMNIKA, 17*(6), 2834-2841. doi:10.12928/TELKOMNIKA.v17i6.13241

Purwanto, A., Janssen, M., & Zuiderwijk, A. (2017). *Towards an Open Government Data Success Model: A Case Study from Indonesia*. Paper presented at the 17th European Conference on Digital Government, Lisbon, Portugal.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2018a). *Citizen engagement in an open election data initiative: A case study of Indonesian's "Kawal Pemilu"*. Paper presented at the 19th Annual International Conference on Digital Government Research, Delft, The Netherlands.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2018b, 3-5 September). *Group Development Stages in Open Government Data Engagement Initiatives: A Comparative Case Studies Analysis.* Paper presented at the Electronic Government 2018, Krems, Austria.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2019, 2-4 September). *Citizens' Motivations for Engaging in Open Data Hackathons.* Paper presented at the Electronic Participation 2019, San Benedetto Del Tronto, Italy.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020a). Citizen Engagement With Open Government Data: A Systematic Literature Review of Drivers and Inhibitors. *International Journal of Electronic Government Research, 16*(3), 1-32.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020b). Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. *Transforming Government: People, Process and Policy, 14*(1), 1-30. doi:10.1108/TG-06-2019-0051

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020c, 15-17 June). *Citizens' Trust in Open Government Data: A Quantitative Study about the Effects of Data Quality, System Quality and Service Quality.* Paper presented at the 21st Annual International Conference on Digital Government Research, Seoul, South Korea.

Quintelier, E., & Vissers, S. (2008). The Effect of Internet Use on Political Participation: An Analysis of Survey Results for16-Year-Olds in Belgium. *Social Science Computer Review, 26*(4), 411-427. doi:10.1177/0894439307312631

Raman, B. (2012). The Rhetoric and Reality of Transparency: Transparent Information, Opaque City Spaces and the Empowerment Question. *The Journal of Community Informatics, 8*(2).

Ramo, D. E., Hall, S. M., & Prochaska, J. J. (2010). Reaching young adult smokers through the Internet: Comparison of three recruitment mechanisms. *Nicotine & Tobacco Research, 12*(7), 768–775. doi:10.1093/ntr/ntq086

Renaud, K., & van Biljon, J. (2008). *Predicting Technology Acceptance and Adoption by the Elderly: A Qualitative study*. Paper presented at the the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries, Wilderness, South Africa.

Republik Indonesia. (1946). *Undang-undang Republik Indonesia Nomor 12 Tahun 1946 tentang Pembaharuan Komite Nasional Pusat*. Jakarta: Presiden Republik Indonesia

Republik Indonesia. (2011). *Undang-undang Republik Indonesia Nomor 15 Tahun 2011 tentang Penyelenggara Pemilihan Umum*. Jakarta: Presiden Republik Indonesia

Rigdon, E. E. (2012). Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. *Long Range Planning, 45*(5–6), 341-358. doi:10.1016/j.lrp.2012.09.010

Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results: The importance-performance map analysis. *Industrial Management & Data Systems, 116*(9), 1865-1886. doi:10.1108/IMDS-10-2015-0449

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. Boenningstedt: SmartPLS GmbH. Retrieved from http://www.smartpls.com

Roberts, N. (2004). Public Deliberation in an Age of Direct Citizen Participation. *The American Review of Public Administration, 34*(4), 315-353.

Rogers, E. M. (1983). *Diffusion of Innovations* (3 ed.). New York: The Free Press.

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

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. doi:10.1093/biomet/63.3.581

Rudmark, D., Arnestrand, E., & Avital, M. (2012). *Crowdpushing: The flip Side of Crowdsourcing*. Paper presented at the 20th European Conference on Information Systems (ECIS'12), Barcelona, Spain.

Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A., & Meijer, A. (2017). Connecting societal issues, users and data. Scenario-based design of open data platforms. *Government Information Quarterly, 34*(3), 470-480.

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering, 14*, 131–164. doi:10.1007/s10664-008-9102-8

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity, 22*(1), 1-24. doi:10.3233/IP-160012

Sarstedt, M., & Mooi, E. (2019). *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics* (3 ed.). Berlin, Germany: Springer-Verlag GmbH Germany.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial Least Squares Structural Equation Modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of Market Research*. Heidelberg: Springer International Publishing AG.

Saxena, S., & Janssen, M. (2017). Examining open government data (OGD) usage in India through UTAUT framework. *Foresight, 19*(4), 421-436. doi:doi.org/10.1108/FS-02-2017-0003

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014). *A framework for benchmarking open government data efforts.* Paper presented at the 2014 47th Hawaii International Conference on System Sciences.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7*(2), 147–177. doi:10.1037/1082-989X.7.2.147

Schlägel, C., & Sarstedt, M. (2016). Assessing the measurement invariance of the four-dimensional cultural intelligence scale across countries: A composite model approach. *European Management Journal, 34*(6), 633-649. doi:10.1016/j.emj.2016.06.002

Schmidthuber, L., Hilgers, D., Gegenhuber, T., & Etzelstorfer, S. (2017). The emergence of local open government: Determinants of citizen participation in online service reporting. *Government Information Quarterly, 34*(3), 457-469. doi:10.1016/j.giq.2017.07.001

Schmidthuber, L., Piller, F., Bogers, M., & Hilgers, D. (2019). Citizen participation in public administration: investigating open government for social innovation. *R&D Management, 49*(3), 343-355. doi:10.1111/radm.12365

Schwartz, M. S., & Schwartz, C. G. (1955). Problems in Participant Observation. *American Journal of Sociology, 60*(4), 343-353.

Sekaran, U., & Bougie, R. (2016). *Research Methods for Business: A Skill-Building Approach* (7 ed.). West Sussex, UK: John Wiley & Sons Ltd.

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS quarterly, 35*(3), 553-572. doi:10.2307/23042796

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552-4564. doi:10.1016/j.jbusres.2016.03.049

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322-2347. doi:10.1108/EJM-02-2019-0189

Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An Assessment of the Use of Structural Equation Modeling in Strategic Management Research. *Strategic Management Journal, 25*(4), 397-404.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly, 32*(3), 308-315. doi:10.1016/j.giq.2015.05.003

Small, M. L. (2011). How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature. *Annual Review of Sociology, 37*, 57-86. doi:10.1146/annurev.soc.012809.102657

Smith, G., Ofe, H. A., & Sandberg, J. (2016). Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace. *2016 49th Hawaii International Conference on System Sciences (Hicss)*, 1277-1286. doi:10.1109/Hicss.2016.162

Smith, G., & Sandberg, J. (2018). Barriers to innovating with open government data: Exploring experiences across service phases and user types. *Information Polity, 23*(3), 249–265. doi:10.3233/IP-170045

Organisatie- en mandaatbesluit OCW 2005, (2005).

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B (Methodological), 36*(2), 111-133. doi:10.1111/j.2517-6161.1974.tb00994.x

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems, 13*, 380-427. doi:10.17705/1CAIS.01324

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the Acm, 40(5)*. Retrieved from http://dl.acm.org/citation.cfm?id=253804

Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. California, USA: Sage Publications, Inc.

Susha, I., Grönlund, Å., & Janssen, M. (2015). Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy, 9*(2), 181 - 206. doi:10.1108/TG-05-2014-0016

Susha, I., Zuiderwijk, A., Janssen, M., & Grönlund, Å. (2015). Benchmarks for Evaluating the Progress of Open Data Adoption Usage, Limitations, and Lessons Learned. *Social Science Computer Review, 33*(5), 613-630.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7 ed.). Boston, MA: Pearson Education. Inc.

Talukder, M. S., Shen, L., Talukder, M. F. H., & Bao, Y. (2019). Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. *Technology in Society, 56*, 147-156. doi:10.1016/j.techsoc.2018.09.013

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The New Era of Mixed Methods. *Journal of Mixed Methods Research, 1*(1), 3-7. doi:10.1177/2345678906293042

Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. California, USA: SAGE Publications, Inc.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis, 48*(1), 159-205. doi:10.1016/j.csda.2004.03.005

Teo, T. S. H., Srivastava, S. C., & Jiang, L. (2009). Trust and electronic government success: An empirical study. *Journal of management information systems, 25*(3), 99-131. doi:10.2753/MIS0742-1222250303

Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65. doi:10.1146/annurev.clinpsy.1.102803.144239

Toots, M., McBride, K., Kalvet, T., & Krimmer, R. (2017). *Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers*. Paper presented at the 2017 Conference for E-Democracy and Open Government (CeDEM), Krems, Austria.

Ubaldi, B. (2013). *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. Retrieved from Paris, France:

Ullman, J. B. (2006). Structural Equation Modeling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment, 87*(1), 35-50.
Urbach, N., & Ahlemann, F. (2010). Structural Equation Modeling in Information Systems Research Using Partial Least Squares. *Journal of Information Technology Theory and Application, 11*(2), 5-40.
Van de Ven, A. H. (2007). *Engaged Scholarship: A Guide for Organizational and Social Research*. New York, US: Oxford University Press Inc.
Van Essen, A. (1997). English in Mainland Europe – A Dutch Perspective. *World Englishes, 16*(1), 95-103. doi:10.1111/1467-971X.00050
van Loenen, B., Ubacht, J., Labots, W., & Zuiderwijk, A. (2017, 12-13 June). *Log File Analytics for Gaining Insight into Actual Use of Open Data.* Paper presented at the 17th European Conference on Digital Government, Lisbon, Portugal.
van Twist, M., van der Steen, M., Kleiboer, M., Scherpenisse, J., & Theisens, H. (2013). *Coping with very weak primary schools: Towards 'smart intervention' in Dutch education policy*. Retrieved from
Veeckman, C., & van der Graaf, S. (2015). The City as Living Laboratory: Empowering Citizens with the Citadel Toolkit. *Technology Innovation Management Review*, 6-17.
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS quarterly, 37*(1), 21-54.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS quarterly, 27*(3), 425-478.
Wakefield, R. L., Stocks, M. H., & Wilder, W. M. (2004). The Role of Web Site Characteristics in Initial Trust Formation. *Journal of Computer Information Systems, 45*(1), 94-103. doi:10.1080/08874417.2004.11645821
Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems, 4*(2), 74-81.
Wang, D., Richards, D., & Chen, C. (2018, August 28–29). *An Analysis of Interaction Between Users and Open Government Data Portals in Data Acquisition Process.* Paper presented at the 15th Pacific Rim Knowledge Acquisition Workshop, PKAW 2018, Nanjing, China.
Wang, D., Richards, D., & Chen, C. (2019). *Connecting Users, Data and Utilization: A Demand-Side Analysis of Open Government Data*. Paper presented at the Information in Contemporary Society. iConference 2019, Washington, DC, USA.
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. doi:10.1080/07421222.1996.11518099
Wang, Y.-S., & Liao, Y.-W. (2008). Assessing eGovernment systems success: A validation of the DeLone and McLean model of information systems success. *Government Information Quarterly, 25*(4), 717-733. doi:10.1016/j.giq.2007.06.002
Wangpipatwong, S., Chutimaskul, W., & Papasratorn, B. (2009). Quality enhancing the continued use of e-government web sites: Evidence from e-citizens of Thailand. *International Journal of Electronic Government Research, 5*(1), 19-35.
Warkentin, M., Gefen, D., Pavlou, P. A., & Rose, G. M. (2002). Encouraging Citizen Adoption of e-Government by Building Trust. *Electronic Markets, 12*(3), 157-162. doi:10.1080/101967802320245929
Weber, R. (2004). Editor's Comments: The Rhetoric of Positivism versus Interpretivism: A Personal View. *MIS quarterly, 28*(1), iii-xii. doi:10.2307/25148621
Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly, 26*(2), xiii-xxiii.
Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontiers, 19*(2), 285–300. doi:10.1007/s10796-016-9679-1
Weerakkody, V., Kapoor, K., Balta, M. E., Irani, Z., & Dwivedi, Y. K. (2017). Factors influencing user acceptance of public sector big open data. *Production Planning & Control, 28*(11-12), 891–905. doi:10.1080/09537287.2017.1336802
Whitmore, A. (2014). Using open government data to predict war: A case study of data and systems challenges. *Government Information Quarterly, 31*(4), 622-630. doi:10.1016/j.giq.2014.04.003

Wicks, A. C., & Freeman, R. E. (1998). Organization Studies and the New Pragmatism: Positivism, Anti-positivism, and the Search for Ethics. *Organization Science, 9*(2), 123-251. doi:10.1287/orsc.9.2.123
Wijnhoven, F., Ehrenhard, M., & Kuhn, J. (2015). Open government objectives and participation motivations. *Government Information Quarterly, 32*(1), 30-42. doi:10.1016/j.giq.2014.10.002
Wirtz, B. W., Weyerer, J. C., & Rösch, M. (2018). Citizen and Open Government: An Empirical Analysis of Antecedents of Open Government Data. *International Journal of Public Administration, 41*(4), 308-320. doi:10.1080/01900692.2016.1263659
Wirtz, B. W., Weyerer, J. C., & Rösch, M. (2019). Open government and citizen participation: an empirical analysis of citizen expectancy towards open government data. *International Review of Administrative Sciences, 85*(3), 566-586. doi:10.1177/0020852317719996
Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS quarterly, 36*(3), 787-810.
Yin, R. K. (1981). The Case Study Crisis: Some Answers. *Administrative Science Quarterly, 26*(1), 58-65.
Yin, R. K. (2014). *Case Study Research: Design and Methods* (5th ed.). California, USA: SAGE Publications, Inc.
Young, A., Sangokoya, D., & Verhulst, S. (2016). Singapore's Dengue Cluster Map: Open Data for Public Health. In *The Global Impact of Open Data: Key Findings from Detailed Case Studies Around the World* (1 ed., pp. 381-396): O'Reilly Media Inc.
Young, A., & Verhulst, S. (2016). *The Global Impact of Open Data: Key Findings from Detailed Case Studies Around the World* (S. Cutt Ed. 1 ed.): O'Reilly Media Inc.
Young Creators. (2018). Over Young Creators. Retrieved from https://youngcreators.co/p/over
Zuiderwijk, A. (2015). *Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use.* (PhD). TU Delft, Delft.
Zuiderwijk, A., & Cligge, M. (2016). The Acceptance and Use of Open Data Infrastructures - Drawing upon UTAUT and ECT. In H. J. Scholl, O. Glassey, M. Janssen, B. Klievink, I. Lindgren, P. Parycek, E. Tambouris, M. Wimmer, T. Janowski, & D. S. Soares (Eds.), *Electronic Government and Electronic Participation* (pp. 91-98): IOS Press.
Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), i-xiii. doi:10.4067/S0718-18762014000200001
Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of eGovernment, 10*(2), 156-172.
Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity, 19*(1, 2), 17–33. doi:doi.org/10.3233/IP-140329
Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly, 32*(4), 429-440. doi:10.1016/j.giq.2015.09.005
Zuiderwijk, A., Janssen, M., & Susha, I. (2016). Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), 116-146.
Zuiderwijk, A., Shinde, R., & Janssen, M. (2019). Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results? *International Review of Administrative Sciences, 85*(4), 645-672. doi:10.1177/0020852317739115
Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management, 49*, 228-241. doi:10.1016/j.ijinfomgt.2019.05.024
Zuiderwijk, A., Susha, I., Charalabidis, Y., Parycek, P., & Janssen, M. (2015). *Open data disclosure and use: Critical factors from a case study.* Paper presented at the CeDEM15: Conference for E-Democracy and Open Government.
Zurada, J., & Karwowski, W. (2011). Knowledge Discovery Through Experiential Learning From Business and Other Contemporary Data Sources: A Review and Reappraisal. *Information Systems Management, 28*(3), 258-274. doi:10.1080/10580530.2010.493846

Summary

Problem statement

This research investigates the factors influencing the citizens' intention to engage with Open Government Data (OGD). Governmental organizations are increasingly making non-privacy-restricted and non-confidential data publicly available on the internet that anyone can freely use, reuse, and distribute without any restrictions. The opening up the government data enables citizens to engage with OGD, i.e., to collaboratively convert OGD into valuable artifacts that benefit society. The artifacts are expected to be used to solve societal problems (e.g., detection and prevention of corruption). However, citizen engagement with OGD is contingent upon many factors, and researchers often overlook these factors, and insight is needed to stimulate engagement.

Moreover, the current insights into the factors influencing citizens' intentions to engage with OGD are fragmented. This fragmentation refers to the situation in which previous research has examined different factors based on different theories and theoretical models. However, researchers have not investigated these factors in an integrated model. Therefore, this study aims *to develop a model for understanding factors contributing to citizen engagement with OGD*. This research focuses on digitally literate or technologically skilled citizens who are not government officials. Outside the scope of this research are governmental organizations, OGD users from the public (e.g., civil society organizations or non-governmental organizations) and private sectors (e.g., companies), and OGD beneficiaries (end-users).

Research method, questions, and contribution

A mixed-methods research approach that combines qualitative and quantitative methods is used to achieve the aim of this research. For this study, an exploratory sequential mixed methods research design was used, combining three phases of research with two qualitative approaches and one quantitative approach. This study answered the following three research questions, which also correspond to the three research phases:

1. *What drivers and inhibitors for citizen engagement with OGD have been identified in previous research?* The first phase of this research aims at better understanding citizen engagement with OGD by systematically analyzing the current literature and identifying factors that drive and inhibit citizen engagement. Many factors influence individual citizen engagement in a positive (driving) or negative (inhibiting) way. However, no comprehensive overview of these factors exists. Various factors were identified and organized into seven clusters: citizen profiles, intrinsic

motivations, extrinsic motivations, economic factors, social factors, technical factors, and political factors. A theoretical framework has been derived from these findings to study factors influencing citizen engagement with OGD in practice. This research is among the first that provides an integrated overview of the driving and inhibiting factors of citizen engagement with OGD.

2. *Why do citizens engage with OGD in existing government-led and citizen-led OGD initiatives?* The second research phase aims to identify factors derived from the literature and explore missing factors in real-life OGD engagement cases. A multiple case study approach was used to investigate real-life cases of OGD engagement to answer the second research question. Two cases were selected: one government-led engagement case that dealt with primary education and one citizen-led engagement case dealing with the presidential election. The theoretical framework developed in the first research phase was used to investigate these two case studies. Fifteen factors were identified and grouped in intrinsic and extrinsic motivations, social, technical, and political categories, which played an influential role in the government-led and citizen-led OGD engagement cases. As the outcome of the second research phase, 25 propositions were developed concerning the effects of factors grouped in five clusters on the intention to engage with OGD and the moderation roles of the citizen profiles on these effects. This research is among the first to provide empirical evidence of citizen-led OGD engagement. In this type of OGD engagement, citizens independently engage with OGD without explicit encouragement from the OGD provider, i.e., the government.
3. *What model explains citizens' intention to engage with OGD?* The third research phase aimed to assess and validate the model that explains the factors influencing citizens' intention to engage with OGD. Ten hypotheses were formulated, and a research model was developed based on the propositions derived in the second research phase. A partial least square structural equation modeling (PLS-SEM) approach was used to answer the third research question. An online questionnaire has been developed and distributed to various channels to collect data. The questionnaire was completed by citizens from different backgrounds who are not employed by the government. The findings are as follows:
   - Both extrinsic and intrinsic motivations and political factors are the most important groups of factors influencing citizens' intention to engage with OGD. However, motivations have a more significant effect on intention than political factors.

- At the individual factor measurement level, relative advantage, the desire to get to know new people, self-efficacy, and enjoyment are important motivations that have the largest effects on the intention to engage with OGD. At the same time, the desire to benefit society is an important political factor that has the largest effect on the intention to engage with OGD.
- Citizens' experience with OGD moderates the influence of social factors on behavioral intention, while citizens' education level moderates the effect of technical factors on behavioral intention. Respondents who have a long experience with OGD engagement will likely need social influence to keep them engaging with OGD because they may have had bad experiences. Respondents who have a higher level of education are more likely to engage with a higher quality of OGD.
- The relationships between the factors and behavioral intentions to engage with OGD show no statistical difference across the two types of OGD engagement: citizen-led engagement and government-led engagement.

This research is among the first to empirically evaluate the unified theoretical model of factors influencing citizens' intention to engage with OGD. This research is also among the first to empirically test assumptions held in the open data research regarding the effects of OGD quality, political participation, and trust on the intention to engage with OGD.

Conclusions

This study developed a model that explains the factors influencing citizen engagement with OGD through three research phases: an exploratory sequential mixed methods design combining qualitative and quantitative approaches. The model was built based on the multivariate analysis results using a PLS-SEM approach named the OGD Citizen Engagement Model (OGD-CEM). The OGD-CEM model explains that two categories of factors determine citizens' behavioral intention to engage with OGD. The first category is the citizen's (both extrinsic and intrinsic) motivations toward the OGD engagement. The second category concerns the citizen's perceived political factors toward OGD and its provider (i.e., governmental organizations). The model results show that the more citizens perceive that engaging with OGD will give them an advantage and allow them to broaden their social networks, the more inclined they will be to engage with OGD. This research also found that the more citizens perceive that they can engage with OGD easily, that engaging with OGD is enjoyable, and that OGD engagement challenges them intellectually, the more they will likely engage with OGD. Furthermore, the model shows that the more citizens perceive that their engagement with OGD will influence public policy, and the higher citizens' trust in OGD and the governmental organizations that provide the data, the more they will be inclined to engage with OGD.

The OGD-CEM model reinforces the current insights of OGD literature on the effect of motivations on the citizens' intentions to engage with OGD. At the same time, the model is among the first that reveals the effect of political factors on the citizens' intentions to engage with OGD. However, the model contradicts the current beliefs that social influence and three dimensions of IS quality (i.e., data quality, system quality, and service quality) affect citizens' behavioral intention to engage with OGD.

Further research recommendations

Four directions of future OGD research concerning citizen engagement with OGD were recommended. The first recommendation concerns the evaluation of the OGD-CEM model in different contexts. For example, evaluating the model quantitatively to a particular OGD, a particular government organization from a specific administrative level, in a particular country on a different type of citizens (e.g., civil servants). Future research that focuses on examining the extent to which the identified factors of the model also apply to other contexts is also recommended, for instance, in situations involving other types of OGD, other types of OGD engagement, in other countries, in other cultures, or at the local government level.

The second recommendation is concerned with taking intermediating factors into account. The effect of citizens' resources, such as technology, skills/knowledge, time, and money, on their intentions to engage with OGD is rarely investigated. Future research on the evaluation of the OOD-CEM model taking into account such intermediate factors is recommended.

The third recommendation is related to analyzing the intra-relationships among factors. The multivariate analysis results indicated that there is the possibility of intra-relationships among certain factors. For example, there may be relationships between the social factors and both motivations and political factors and those of the technical and political factors.

The fourth recommendation concerns investigating the actual OGD engagement. The citizens' intention to engage with OGD may not always lead to the expected behaviors (i.e., the actual OGD engagement) or may do so in an inconsistent way. Future research investigating the relationship between the intention to engage with OGD and actual OGD engagement using a longitudinal field survey strategy is recommended.

Samenvatting

Probleemstelling

Dit onderzoek bestudeert de factoren die van invloed zijn op de intentie van burgers om met Open Overheidsdata (OOD) om te gaan. Overheidsorganisaties maken in toenemende mate niet-privacy-beperkte en niet-vertrouwelijke gegevens openbaar op het Internet die iedereen vrij kan gebruiken, hergebruiken en verspreiden. Het ontsluiten van de overheidsgegevens stelt burgers in staat om met OOD in contact te komen, d.w.z. om samen OOD om te zetten in waardevolle artefacten die de samenleving ten goede komen. De artefacten zullen naar verwachting worden gebruikt om maatschappelijke problemen op te lossen (bijvoorbeeld detectie en preventie van corruptie). De betrokkenheid van burgers bij OOD is echter afhankelijk van veel factoren, en onderzoekers zien deze factoren vaak over het hoofd.

De huidige inzichten in de factoren die van invloed zijn op de intenties van burgers om met OOD om te gaan zijn versnipperd. Verschillende factoren zijn onderzocht op basis van verschillende theorieën en theoretische modellen, maar onderzoekers hebben deze factoren niet onderzocht in een geïntegreerd model. Daarom heeft deze studie tot doel *een model te ontwikkelen voor het begrijpen van factoren die bijdragen aan de betrokkenheid van burgers bij OOD*. Dit onderzoek richt zich op digitaal geletterde of technologisch vaardige burgers die geen overheidsfunctionaris zijn. Buiten de reikwijdte van dit onderzoek vallen overheidsorganisaties, OOD-gebruikers uit het publiek (bijvoorbeeld maatschappelijke organisaties of niet-gouvernementele organisaties) en particuliere sectoren (bijvoorbeeld bedrijven), en OOD-begunstigden (eindgebruikers).

Onderzoeksmethode, vragen en bijdrage

Om het doel van dit onderzoek te bereiken, wordt gebruik een *mixed methods onderzoeksaanpak* die kwalitatieve en kwantitatieve methoden combineert. Voor deze studie werd een *exploratory sequential mixed methods* onderzoeksontwerp gebruikt, bestaande uit drie onderzoeksfasen werden. In dit ontwerp werden twee kwalitatieve benaderingen en één kwantitatieve benadering gecombineerd. Dit onderzoek beantwoordde de volgende drie onderzoeksvragen, die overeenkomen met de drie onderzoeksfasen:

1. *Welke drijfveren en barrières voor burgerbetrokkenheid bij OOD zijn in eerder onderzoek geïdentificeerd?* De eerste fase van dit onderzoek is bedoeld om de betrokkenheid van burgers bij OOD beter te begrijpen door de huidige literatuur systematisch te analyseren. Drijvende en remmende

factoren zijn geïdentificerd die de betrokkenheid van burgers beïnvloedenen om zo een theoretisch raamwerk van factoren te ontwikkelen. Een veelheid aan factoren beïnvloeden de individuele burgerbetrokkenheid op een positieve (drijvende) of negatieve (remmende) manier. Er bestaat echter geen volledig overzicht van deze factoren. Verschillende factoren werden geïdentificeerd en georganiseerd in zeven clusters: burgerprofielen, intrinsieke motivaties, extrinsieke motivaties, economische factoren, sociale factoren, technische factoren en politieke factoren. Uit deze bevindingen is een theoretisch kader afgeleid om factoren te bestuderen die de betrokkenheid van burgers bij OOD beïnvloeden in de praktijk. Dit onderzoek is een van de eerste onderzoeken die een geïntegreerd overzicht geeft van de drijvende en remmende factoren van burgerbetrokkenheid bij OGD.

2. *Waardoor raken burgers betrokken bij OOD-initiatieven in bestaande door de overheidsgestuurde en door burgersgestuurde OOD-initiatieven?* De tweede onderzoeksfase is gericht op het onderzoeken van factoren in de praktijk. Er werd een benadering met meerdere casestudy's gebruikt om praktijkgevallen van OOD-betrokkenheid te onderzoeken om de tweede onderzoeksvraag te beantwoorden. Er werden twee casestudies geselecteerd; één overheidsgestuurde betrokkenheid case in het onderwijs domein en één burgergestuurde case in het verkiezingsdomein. Het theoretische kader dat in de eerste onderzoeksfase werd ontwikkeld, werd gebruikt als raamwerk om deze twee case studies te onderzoeken. Er werden vijftien factoren geïdentificeerd, gegroepeerd in intrinsieke motivatie, extrinsieke motivatie, sociale, technische en politieke categorieen, die een invloedrijke rol speelden in de door de overheidgestuurd en door burgersgestuurde OOD-betrokkenheidsinitiatieven. Als uitkomst van de tweede onderzoeksfase zijn 25 stellingen ontwikkeld over de effecten van factoren gegroepeerd in vijf clusters op de intentie om met OOD om te gaan en de moderatierollen van de burgerprofielen op deze effecten. Dit is één van de eerste onderzoeken naar burgergestuurde OOD-betrokkenheid. Bij dit type OOD-betrokkenheid werken burgers volledig onafhankelijk van de overheid met OOD zonder expliciete aanmoediging vanuit de de OOD-aanbieder, d.w.z. de overheid.
3. *Welk model verklaart de intentie van burgers om met OOD om te gaan?* De derde onderzoeksfase richte zich op het beoordelen en valideren van het onderzoeksmodel dat de factoren verklaart die van invloed zijn op de intentie van burgers om met OOD betrokken te raken. Op basis van de stellingen die in de tweede onderzoeksfase zijn afgeleid, zijn tien hypothesen geformuleerd en is een onderzoeksmodel ontwikkeld. Een

PLS-SEM-techniek (Partial Least Square-Structural Equation Modeling) werd gebruikt om de derde onderzoeksvraag te beantwoorden. Om data te verzamelen is een online vragenlijst ontwikkeld en uitgezet via verschillende kanalen. De vragenlijst is ingevuld door burgers met verschillende achtergronden die niet werkzaam zijn voor de overheid. De bevindingen zijn als volgt:

- Zowel extrinsieke als intrinsieke motivaties en politieke factoren zijn de belangrijkste groepen van factoren die van invloed zijn op de intentie van burgers om met OOD betrokken te raken. Motivaties hebben echter een groter effect op intentie dan politieke factoren.
- Op het niveau van individuele factormeting hebben ‘relatief voordeel’, ‘de wens om nieuwe mensen te leren kennen’, ‘zelfeffectiviteit’ en ‘plezier’ de grootste effecten op de intentie om bij OOD betrokken te raken. Tegelijkertijd is ‘de wens om de samenleving ten goede te komen’ een belangrijke politieke factor die het grootste effect heeft op de intentie om met OOD aan de slag te gaan.
- De ervaring van burgers met OOD modereert de invloed van sociale factoren op gedragsintentie, terwijl het opleidingsniveau van burgers het effect van technische factoren op gedragsintentie modereert. Respondenten die veel ervaring hebben met OOD, hebben sociale invloed waarschijnlijk nodig om hen betrokken te houden bij OOD, omdat ze mogelijk ook slechte ervaringen hebben. Respondenten met een hoger opleidingsniveau hebben meer kans om deel te nemen aan een hogere kwaliteit van OOD.
- De relaties tussen de factoren en gedragsintenties om met OOD in contact te komen laten geen statistisch verschil zien tussen de twee soorten OOD- betrokkenheid: door burgers geleide betrokkenheid en door de overheid geleide betrokkenheid.

Als één van de eerste op dit gebied evalueert dit empirisch onderzoek een uitgebreid theoretisch model van factoren die van invloed zijn op de intentie van burgers om met OOD in contact te komen. Dit onderzoek is ook één van de eerste onderzoeken die veronderstellingen in het open data-onderzoek empirisch toetst met betrekking tot de effecten van OOD-kwaliteit, politieke participatie en vertrouwen op de intentie om met OOD in contact te komen.

Conclusies

In deze studie is een model ontwikkeld dat de factoren verklaart die van invloed zijn op de betrokkenheid van burgers bij OOD door middel van drie onderzoeksfasen. Het model is gebouwd op basis van de multivariate analyseresultaten met behulp van een PLS-SEM-benadering, het *OOD Citizen Engagement Model* (OOD-CEM) genoemd. Het OOD-CEM-model legt uit dat de gedragsintentie van burgers om met OOD om te gaan, wordt bepaald door twee groepen van factoren. De eerste groep is de (zowel extrinsieke als intrinsieke) motivatie van de burger ten aanzien van de OOD-betrokkenheid. De tweede groep betreft de door de burger waargenomen politieke factoren ten opzichte van OOD en zijn aanbieder (d.w.z. overheidsorganisaties). Het model gaat ervan uit dat hoe meer burgers ervaren dat betrokkenheid bij OOD hen voordeel geeft en de mogelijkheid biedt om hun sociale netwerken te verbreden, hoe meer ze geneigd zullen zijn om zich met OOD bezig te houden. Het model laat zien dat hoe meer burgers ervaren dat ze gemakkelijk met OOD kunnen omgaan, dat het leuk is om met OOD bezig te zijn en dat OOD-betrokkenheid hen intellectueel uitdaagt, hoe meer ze zich waarschijnlijk met OOD zullen bezighouden. Bovendien laat het model ook zien dat hoe meer burgers zien dat hun betrokkenheid bij OOD het openbare beleid beïnvloeden, en hoe meer burgers vertrouwen hebben in OOD en de overheidsorganisaties die de gegevens verstrekken, hoe meer ze geneigd zullen zijn om zich met OOD bezig te houden.

Het OOD-CEM-model versterkt de huidige inzichten in de OOD-literatuur over het effect van motivaties op de intenties van burgers om met OOD betrokken te zijn. Tegelijkertijd is het model een van de eerste die het effect van politieke factoren op de intenties van burgers om zich met OOD bezig te houden, laat zien. Het model is echter in tegenspraak met de huidige opvattingen dat sociale invloed en drie dimensies van IS-kwaliteit (d.w.z. gegevenskwaliteit, systeemkwaliteit en servicekwaliteit) van invloed zijn op de gedragsintentie van burgers om met OOD om te gaan.

Verdere onderzoeksaanbevelingen

Vier suggesties voor toekomstig onderzoek met betrekking tot burgerbetrokkenheid bij OOD worden gesuggereerd. De eerste aanbeveling betreft de evaluatie van het OOD-CEM-model in verschillende contexten. Bijvoorbeeld omm het model kwantitatief te evalueren voor een bepaald type OOD, een bepaalde overheidsorganisatie van een bepaald bestuurlijk niveau, en in een bepaald land op een ander type burgers (bijvoorbeeld ambtenaren). Toekomstig onderzoek dat zich richt op het onderzoeken van in hoeverre de geïdentificeerde factoren van het model ook van toepassing zijn op andere contexten, wordt ook aanbevolen. Bijvoorbeeld in situaties waarin sprake is van andere soorten OOD, andere soorten OOD-betrokkenheid, in andere landen, in andere culturen of op het niveau van de lokale overheid.

De tweede aanbeveling betreft het in aanmerking nemen van intermediaire factoren. Het effect van de middelen waarover burgers beschikken, zoals technologie, vaardigheden/kennis, tijd en geld, op hun intenties om zich met

OOD bezig te houden, is nauwelijks onderzocht. Toekomstig onderzoek naar de evaluatie van het OOD-CEM-model, rekening houdend met dergelijke intermediaire factoren, wordt aanbevolen

De derde aanbeveling heeft betrekking op het analyseren van de onderlinge relaties tussen de factoren. De multivariate analyseresultaten gaven aan dat er intra-relaties tussen bepaalde factoren mogelijk zijn. Zo kunnen er verbanden zijn tussen de sociale factoren en zowel drijfveren en politieke factoren als die van de technische factoren en politieke factoren.

De vierde aanbeveling betreft het onderzoeken van de daadwerkelijke betrokkenheid van burgers bij OOD. De intentie van de burger om met OOD om te gaan leidt mogelijk niet altijd tot het verwachte gedrag (d.w.z. de daadwerkelijke OOD-betrokkenheid) of kan dit op een inconsistente manier doen. Toekomstig onderzoek aan om de relatie tussen de intentie om met OOD in contact te komen en daadwerkelijke OOD-betrokkenheid te onderzoeken met behulp van een longitudinale veldstudie, wordt aanbevolen.

Appendices

Appendix A: List of the reviewed papers

Table A.1. List of the reviewed papers in the systematic literature review (see Chapter 3).

Authors	Context	Period	Domain	Number of Respondents	Method	Unit of Analysis
Afful-Dadzie and Afful-Dadzie (2017)	Africa (Ghana, Kenya, Sierra Leone, South Africa, Tanzania)	2016	NA	198	Survey	Media practitioners
Benitez-Paez et al. (2018)	Colombia, Spain	2016-2017	Geographic	195 (survey), 155 (workshop)	Mixed Methods	Geographical data users
Beno et al. (2017)	Austria	2015-2016	NA	110	Survey	Multiple types of actors
Canares (2014)	Philippines	2011	Spending	NA	Mixed Methods	Civil society organization
Charalabidis et al. (2014)	Europe (Greece, Netherlands)	NA	Research	42	Survey	Post-graduate students
Choi and Tausczik (2017)	United States (USA), South Korea, Singapore	NA	NA	18 (interview), 22 (survey)	Mixed Methods	Hackathon participants
Cranefield et al. (2014)	New Zealand	NA	Geospatial	17	Case study	Multiple types of actors
Crusoe et al. (2019)	Namur, Belgium	2018	NA	30 (survey), 9 (interview)	Mixed Methods	Students
de Deus Ferreira and Farias (2018)	Brazil	2016	NA	308	Survey	Hackathon participants
de Kool and Bekkers (2015)	Netherlands	NA	Education inspection	245 (survey), 35 (interview)	Mixed Methods	Parents of primary school students
de Kool and Bekkers (2016)	Netherlands	NA	Education inspection	245 (survey), 35 (interview)	Mixed Methods	Parents of primary school students
Dittus et al. (2016)	Philippines, Guinea, international	2013-2014	Humanitarian	1570	Secondary Data Analysis	Mapping contributors
dos Santos Brito et al. (2014)	Brazil	2013	Election	NA	Case study	The authors
Fitriani et al. (2019)	Indonesia	2016	NA	513	Survey	Citizens
Gama (2017)	Australia, New Zealand, Brazil	2016-2017	NA	123	Survey	Developers participating in hackathons
Hellberg and Hedström (2015)	Sweden	2012-2013	NA	9	Case study	Multiple types of actors
Hivon and Titah (2017)	Montreal, Canada	NA	NA	14	Case study	Hackathon participants
Hjalmarsson et al. (2014)	Stockholm, Sweden	2013	Transportation	249	Survey	Application developers
Hutter et al. (2011)	Bavarian State, German	2010	NA	437	Survey	Open government project participants

Authors	Context	Period	Domain	Number of Respondents	Method	Unit of Analysis
Jarke (2019)	Bremen Hemelingen, Germany	2017	Public facility	46	Action Research	Older adults in open data walk
Juell-Skielse et al. (2014)	Stockholm, Sweden	2013	Transportation	39	Case study	Hackathon participants
Jurisch et al. (2015)	German, Switzerland, Austria, USA, UK, Sweden	2013	NA	6051	Survey	Citizens
Khayyat and Bannister (2017)	Dublin, Ireland	NA	Housing, public service, transportation, budget	48	Case study	Volunteers, application developers, academics
Khurshid et al. (2018)	Pakistan	NA	NA	141	Survey	Academicians
Kuk and Davies (2011)	UK	2010	Transportation, education, spending	8	Case study	Hackathon participants
Martin (2014)	UK	2012-2013	NA	135	Survey	Citizens
Maruyama et al. (2013)	USA	2012	Public service	25	Case study	Fellowship participants
Ojo et al. (2016)	Dublin, Ireland	2015	NA	10	Mixed Methods	Application developers
Osagie et al. (2017)	Dublin, Ireland	NA	NA	19	Case study	Citizens
Purwanto et al. (2018a)	Indonesia	2017-2018	Election	16	Case study	Developers, volunteers
Purwanto et al. (2019)	Netherlands	2017-2019	Agriculture	161	Case study	Hackathon participants
Rudmark et al. (2012)	Stockholm, Sweden	NA	Transportation	13	Case study	Application developers
Ruijer et al. (2017)	Groningen, Netherlands	2015	NA	16	Case study	Citizens
Saxena and Janssen (2017)	India	2014	NA	244	Survey	Citizens
Schmidthuber et al. (2019)	Linz, Austria	2016	NA	73	Survey	Open government platform participants
Smith et al. (2016)	Sweden	2015	Transportation	19	Case study	Open data marketplace users
Smith and Sandberg (2018)	Sweden	2015	Transportation	19	Case study	Open data marketplace users
Talukder et al. (2019)	Dhaka, Bangladesh	2017	NA	285	Survey	Citizens
Toots et al. (2017)	Belgium, Estonia, Greece, Ireland, Lithuania, UK	NA	NA	63	Survey	Citizens
Veeckman and van der Graaf (2015)	Europe (Belgium, France, UK, Greece)	2012	Transportation, tourism	25	Case study	Living lab participants
Wang et al. (2018)	China	2017	NA	208	Survey	Citizens
Wang et al. (2019)	China	2017	NA	208	Survey	Citizens

Authors	Context	Period	Domain	Number of Respondents	Method	Unit of Analysis
Weerakkody, Irani, et al. (2017)	UK	NA	NA	516	Survey	Citizens
Weerakkody, Kapoor, et al. (2017)	UK	NA	NA	350	Survey	Citizens
Whitmore (2014)	USA	2012	Defense	NA	Mixed Methods	The author
Wijnhoven et al. (2015)	German	2013	NA	161	Survey	Citizens
Wirtz et al. (2018)	German	NA	NA	210	Survey	Citizens
Wirtz et al. (2019)	German	NA	NA	210	Survey	Citizens
Zuiderwijk et al. (2012)	International	2011-2012	NA	77	Case study	Multiple types of actors
Zuiderwijk, Janssen, et al. (2015)	International	2012	NA	111	Survey	Researchers
Zuiderwijk, Susha, et al. (2015)	International	NA	NA	23	Mixed Methods	Researchers
Zuiderwijk et al. (2016)	Netherlands	2014	Research	127	Experiment	Students, professional open data users

Appendix B: List of links to the case study database records

Table B.1. List of the qualitative data collection instruments (semi-structured interviews) for the case studies.

Roles of interviewee	**Language**	**4TU link**
Government-led OGD citizen engagement (Case study 1: Hack de Valse Start)		
Hackathon participants	English	https://doi.org/10.4121/15082512
OGD providers	English	https://doi.org/10.4121/15082512
Hackathon organizers	English	https://doi.org/10.4121/15082512
Citizen-led OGD citizen engagement (Case study 2: Kawal Pemilu)		
App developers	Indonesia	https://doi.org/10.4121/15082512
Data volunteers	Indonesia	https://doi.org/10.4121/15082512
OGD providers	Indonesia	https://doi.org/10.4121/15082512

Appendix C: List of links to the survey instruments

Table C.1. List of the survey instruments for the quantitative study.

Language version	**4TU link**
English	https://doi.org/10.4121/16902787
Indonesia	https://doi.org/10.4121/16902787

Appendix D: The list of survey sections of the quantitative study and their questions

Table D.1. The questions on respondents' experience in engaging with OGD (Section A of the survey).

Question No.	Label	Topic	Scale(s)	Measures
Q1	EXP	Experience in OGD engagement	Nominal	"Yes" and "No"
Q2	TIME	Last experience	Ordinal	"Less than 1 year", "1-2 years ago", "2-5 years ago", and "More than 5 years ago"
Q3	TYPE	Type of engagement	Nominal	"I individually self-organized", "II individually organization-sponsored", "III collectively self-organized" and "IV collectively organization-sponsored"
Q4	DOM	Domain of OGD	Nominal (multiple) and open-ended	"Agriculture", "Care and health", "Climate", "Business, economy, and finance", "Defense", "Ecosystems, nature, and environment", "Education, science, and research", "Energy", "Government and management", "Housing", "Industry and manufacturing", "Infrastructure, space, and transportation", "Maritime and ocean", "Public order and safety", "Society and social", "Tourism", and "Other (please specify):"
Q5	OUT	Output of engagement	Nominal (multiple) and open-ended	"Application (e.g., mobile apps, computer application, web-based application)", "Map (e.g., geographical information-based map)", "Visualization (e.g., statistical chart, infographics)", "Article (e.g., research or conference paper, essay)", "News (e.g., investigative journalism, citizen journalism)", "New database", "Other (please specify):", and "No product or service created"
Q6	SOL	Relation of engagement purpose with	Ordinal (Likert)	"Strongly disagree", "Disagree", "Neither disagree nor agree", "Agree", "Strongly agree", and "Not applicable (don't know)"

Table D.2. The questions on respondents' intrinsic and extrinsic motivations (Section B of the survey).

Label	Construct operationalization	Reference from OGD literature	Source of question
Intrinsic motivations (INT)			
INT1	I clearly understand how to engage with open government data	Jurisch et al. (2015), Wirtz et al. (2018)	Venkatesh et al. (2003)
INT2	My engagement with open government data is due to my dissatisfaction with the government	Jurisch et al. (2015)	Rogers (1983), Alathur, Ilavarasan, and Gupta (2016)

Label	**Construct operationalization**	**Reference from OGD literature**	**Source of question**
Intrinsic motivations (INT)			
INT3	My engagement with open government data is on issues of my personal concern	Jurisch et al. (2015)	Rogers (1983), Alathur et al. (2016)
INT4	My engagement with open government data is on issues of my non-personal concerns	Jurisch et al. (2015)	Rogers (1983), Alathur et al. (2016)
INT5	I engage with open government data because the ongoing discussions are against my social or religious interest	Jurisch et al. (2015)	Rogers (1983), Alathur et al. (2016)
INT6	I enjoyed studying open government data	Juell-Skielse et al. (2014), Schmidthuber et al. (2019)	Boudreau and Lakhani (2009)
INT7	I enjoyed my experience in using open government data	Juell-Skielse et al. (2014), Schmidthuber et al. (2019)	Boudreau and Lakhani (2009)
INT8	I am intellectually challenged when engaging with open government data	Juell-Skielse et al. (2014), Wirtz et al. (2018)	Boudreau and Lakhani (2009)
Extrinsic motivations (EXT)			
EXT1	Engaging with open government data is of benefit to me	Juell-Skielse et al. (2014), Weerakkody, Irani, et al. (2017), Wirtz et al. (2018)	Rogers (1983), Venkatesh et al. (2003)
EXT2	My activities do not require me to engage with open government data	Zuiderwijk, Janssen, et al. (2015)	Venkatesh et al. (2003)
EXT3	I want to get to know new people by engaging with open government data	Hutter et al. (2011)	
EXT4	I want to secure my future career by engaging with open government data		Boudreau and Lakhani (2009)

Table D.3. The questions on respondents' evaluation on OGD quality, OGD system quality and OGD service quality (Section C of the survey).

Label	**Construct operationalization**	**Reference from OGD literature**	**Source of question**
Data quality (DQ)			
DQ1	The open government data I engaged with are free from errors	Zuiderwijk et al. (2012), Osagie et al. (2017)	Bailey and Pearson (1983), Doll and Torkzadeh (1988), DeLone and McLean (1992), Wang and Strong (1996), Strong, Lee, and Wang (1997), Wangpipatwong, Chutimaskul, and Papasratorn (2009)
DQ2	The open government data I engaged with are complete (i.e., cover all attributes needed, no missing value)	Zuiderwijk et al. (2012), Beno et al. (2017), Osagie et al. (2017)	DeLone and McLean (1992), Wang and Strong (1996), Strong et al. (1997), Wang and Liao (2008), Wangpipatwong et al. (2009)
DQ3	The open government data I engaged with are well-formatted	Whitmore (2014), Ojo et al. (2016), Smith and Sandberg (2018)	Doll and Torkzadeh (1988), DeLone and McLean (1992), Gable, Sedera, and Chan (2008)

Label	Construct operationalization	Reference from OGD literature	Source of question
Data quality (DQ)			
DQ4	It is easy to link or combine a data set to/with other open government data	Zuiderwijk et al. (2012), Beno et al. (2017), Crusoe et al. (2019)	
System quality (SYSQ)			
SYSQ1	The open government data portal that I engaged with is available at all times		Bailey and Pearson (1983), DeLone and McLean (1992), Wangpipatwong et al. (2009)
SYSQ2	The open government data systems that I engaged with responds at an acceptable speed	Smith et al. (2016)	Molla and Licker (2001)
SYSQ3	The open government data systems that I engaged with provides functionalities needed (e.g., data visualization, feedback mechanism, quality rating)	Zuiderwijk et al. (2012)	DeLone and McLean (1992), Wangpipatwong et al. (2009)
SYSQ4	The open government data systems that I engaged with provided guidance and documentation to download and interpret the data	Beno et al. (2017), Smith and Sandberg (2018)	Parasuraman et al. (1985), Kettinger and Lee (1997), DeLone and McLean (2003)
Service quality (SERVQ)			
SERVQ1	The open government data provider responds sufficiently timely	Hivon and Titah (2017), Smith and Sandberg (2018)	Parasuraman et al. (1985), Kettinger and Lee (1997), DeLone and McLean (2003)
SERVQ2	The open government data provider follows up on a user's report	Smith and Sandberg (2018)	Parasuraman et al. (1985), DeLone and McLean (2003)
SERVQ3	The open government data provider has adequate knowledge to answer a user's request	Zuiderwijk et al. (2012)	Kettinger and Lee (1997), DeLone and McLean (2003)
SERVQ4	The open government data provider prioritizes the user's needs		Kettinger and Lee (1997), DeLone and McLean (2003)

Table D.4. The questions on respondents' evaluation on the influence of social relationships (Section D of the survey).

Label	Construct operationalization	Reference from OGD literature	Source of question
SOC1	People who are important to me (e.g., family, friends) think that I should engage with open government data	Zuiderwijk, Janssen, et al. (2015), Weerakkody, Irani, et al. (2017)	Venkatesh et al. (2003)
SOC2	People who are important to me in my daily job (e.g., colleagues) think that I should engage with open government data	Zuiderwijk, Janssen, et al. (2015), Weerakkody, Irani, et al. (2017)	Venkatesh et al. (2003)

Label	Construct operationalization	Reference from OGD literature	Source of question
SOC3	A community that is important to me (e.g., neighbors, communities, society) think that I should engage with open government data	Zuiderwijk, Janssen, et al. (2015), Weerakkody, Irani, et al. (2017)	Venkatesh et al. (2003)
SOC4	I believe that my engagement with open government data would create benefits for society	Khayyat and Bannister (2017), Hivon and Titah (2017)	

Table D.5. The questions on respondents' evaluation on their trust in OGD and political participation (Section E of the survey).

Label	Construct operationalization	Reference from OGD literature	Source of question
Trust in OGD			
TR1	Open government data providers can be trusted		Wakefield, Stocks, and Wilder (2004), Carter and Bélanger (2005), Bélanger and Carter (2008), Teo et al. (2009)
TR2	The open government data that I engaged with seemed truthful to me		Wakefield et al. (2004), Carter and Bélanger (2005), Teo et al. (2009)
TR3	The open government data I engaged with can be trusted		Wakefield et al. (2004), Carter and Bélanger (2005), Bélanger and Carter (2008), Teo et al. (2009)
Political participation			
POL1	My engagement with open government data can influence public policy	Hutter et al. (2011), Wijnhoven et al. (2015)	Quintelier and Vissers (2008)
POL2	I am interested in politics	Hutter et al. (2011), Jurisch et al. (2015)	Brady et al. (1995), Quintelier and Vissers (2008)
POL3	I am active in political activities (e.g., vote in the election, lobby policy makers, organize demonstration)	Hutter et al. (2011), Jurisch et al. (2015)	Brady et al. (1995), Quintelier and Vissers (2008)

Table D.6. The questions on respondents' intentions to engage with OGD (Section F of the survey).

Label	Construct operationalization	Reference from OGD literature	Source of question
BI1	I predict that I will engage with open government data in the future	Zuiderwijk, Janssen, et al. (2015)	Venkatesh et al. (2003), Fishbein and Ajzen (2010)
BI2	I intend to engage with open government data in the future	Zuiderwijk, Janssen, et al. (2015)	Venkatesh et al. (2003), Fishbein and Ajzen (2010)
BI3	I expect that I will engage with open government data	Zuiderwijk, Janssen, et al. (2015)	Venkatesh et al. (2003), Fishbein and Ajzen (2010)
BI4	I plan to engage with open government data in the future	Zuiderwijk, Janssen, et al. (2015)	Venkatesh et al. (2003), Fishbein and Ajzen (2010)

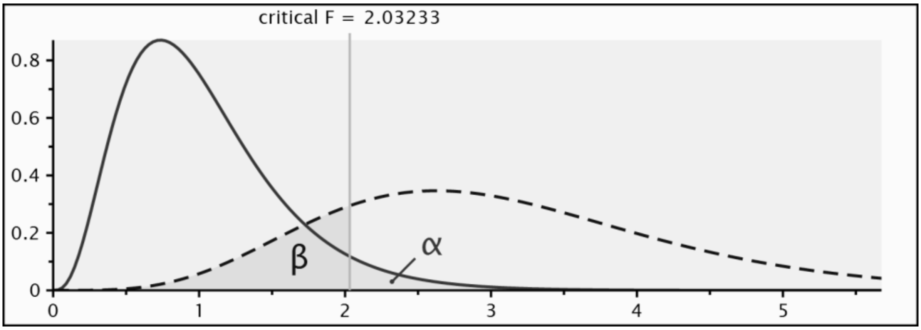
Table D.7. The questions on respondents' demographic information (Section G of the survey).

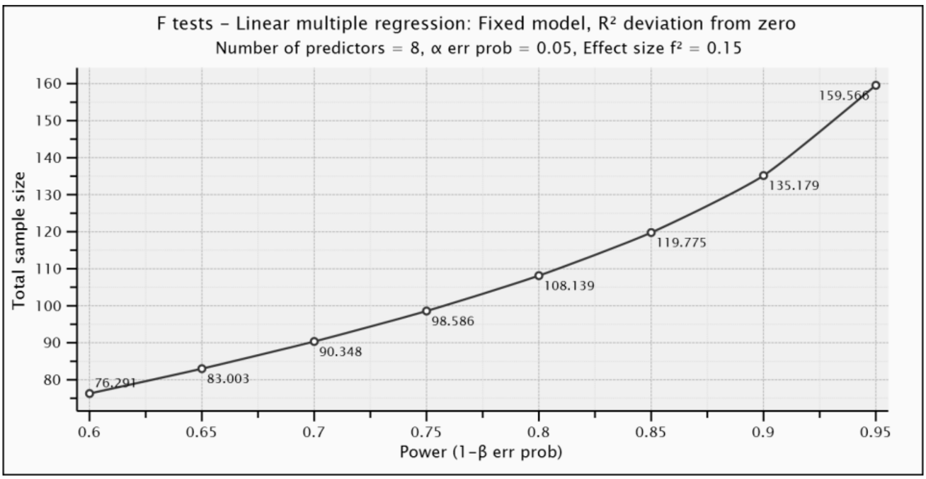
Question No.	**Label**	**Topic**	**Scale(s)**	**Measures**
Q46	GEN	Gender	Nominal	"Female", "Male", "Non-binary/third gender", "Prefer to self-describe", and "Prefer not to answer"
Q47	AGE	Age	Interval (open-ended)	
Q48	NAT	Nationality	Nominal and open-ended	113 options
Q49	EDU	Education level	Nominal	"No formal education", "High school diploma", "College degree", "Vocational training", "Bachelor's degree", "Master's degree", "Professional degree", "Doctorate degree", and "Other"
Q50	WORK	Work status	Nominal	"Employed", "Self-employed/Freelance", "Interning", "Unemployed – Looking for work", "Unemployed – Not looking for work", "Homemaker", "Studying", "Military/Forces", "Retired", "Not able to work", and "Other"
Q51	JOB	Current job	Open-ended	

Appendix E: The list of nationalities and their codes provided in the demographic section of the survey

Code	Name	Code	Name	Code	Name	Code	Name
1	Afghan	31	Fijian	61	Maltese	91	Spanish
2	Albanian	32	Finnish	62	Mexican	92	Sri Lankan
3	Algerian	33	French	63	Mongolian	93	Sudanese
4	Argentinian	34	German	64	Moroccan	94	Swedish
5	Australian	35	Ghanaian	65	Mozambican	95	Swiss
6	Austrian	36	Greek	66	Namibian	96	Syrian
7	Bangladeshi	37	Guatemalan	67	Nepalese	97	Taiwanese
8	Belgian	38	Haitian	68	Dutch	98	Tajikistani
9	Bolivian	39	Honduran	69	New Zealand	99	Thai
10	Batswana	40	Hungarian	70	Nicaraguan	100	Tongan
11	Brazilian	41	Icelandic	71	Nigerian	101	Tunisian
12	Bulgarian	42	Indian	72	Norwegian	102	Turkish
13	Cambodian	43	Indonesian	73	Pakistani	103	Ukrainian
14	Cameroonian	44	Iranian	74	Panamanian	104	Emirati
15	Canadian	45	Iraqi	75	Paraguayan	105	British
16	Chilean	46	Irish	76	Peruvian	106	American
17	Chinese	47	Israeli	77	Philippine	107	Uruguayan
18	Colombian	48	Italian	78	Polish	108	Venezuelan
19	Costa Rican	49	Jamaican	79	Portuguese	109	Vietnamese
20	Croatian	50	Japanese	80	Romanian	110	Welsh
21	Cuban	51	Jordanian	81	Russian	111	Zambian
22	Czech	52	Kenyan	82	Saudi	112	Zimbabwean
23	Danish	53	Kuwaiti	83	Scottish	113	Azerbaijan
24	Dominican	54	Lao	84	Senegalese	114	Tanzanian
25	Ecuadorian	55	Latvian	85	Serbian	115	Kyrgyzstan
26	Egyptian	56	Lebanese	86	Singaporean	116	Burkina Faso
27	Salvadorian	57	Libyan	87	Slovak	117	Other
28	English	58	Lithuanian	88	South African		
29	Estonian	59	Malaysian	89	North Korean		
30	Ethiopian	60	Malian	90	South Korean		

Appendix F: Sample size computation using a priori power analyses of G*Power





Appendix G: Publications by the author

Journal articles

- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020a). Citizen Engagement with Open Government Data: A Systematic Literature Review of Drivers and Inhibitors. International Journal of Electronic Government Research, 16(3), 1-32.
- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020b). Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. Transforming Government: People, Process and Policy, 14(1), 1-30. doi:10.1108/TG-06-2019-0051

Conference papers

- Purwanto, A., Janssen, M., & Zuiderwijk, A. (2017). Towards an Open Government Data Success Model: A Case Study from Indonesia. Paper presented at the 17th European Conference on Digital Government, Lisbon, Portugal.
- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2018a). Citizen engagement in an open election data initiative: A case study of Indonesian's "Kawal Pemilu". Paper presented at the 19th Annual International Conference on Digital Government Research, Delft, The Netherlands.
- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2018b, 3-5 September). Group Development Stages in Open Government Data Engagement Initiatives: A Comparative Case Studies Analysis. Paper presented at the Electronic Government 2018, Krems, Austria.
- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2019, 2-4 September). Citizens' Motivations for Engaging in Open Data Hackathons. Paper presented at the Electronic Participation 2019, San Benedetto Del Tronto, Italy.
- Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020c, 15-17 June). Citizens' Trust in Open Government Data: A Quantitative Study about the Effects of Data Quality, System Quality and Service Quality. Paper presented at the 21st Annual International Conference on Digital Government Research, Seoul, South Korea.

Curriculum Vitae

Arie Purwanto was born in Wonosobo in Indonesia on June 9, 1978. In2002, Arie completed his study in electrical engineering with information systems specialization and earned a bachelor's degree from the University of Gadjah Mada (UGM), one of the oldest universities in Indonesia.

After graduating, Arie started to work as a programmer and course instructor for a company named SalatigaCamp, owned by the Satya Wacana Christian University, which develops tailor made software and provides computer related courses. In April 2004, Arie joined the Information Technology (IT) Bureau of Indonesian Supreme Audit Court (Indonesia: *Badan Pemeriksa Keuangan*; *BPK*). His main responsibilities include developing tailor-made application for internal use and providing information systems or computer-assisted audit services.

In 2006, Arie received a master's degree scholarship from the World Bank to study accounting science at UGM's Economic and Business Faculty. His specialization was auditing. During the study, Arie became interested in electronic government (e-government). In his thesis, he investigated the effectiveness and efficiency of e-government applications in the local government of Sragen Regency in Indonesia. He graduated from the faculty in 2008.

In 2011, Arie received a scholarship from the Studeren in Nederland (StuNED) to attend a tailor-made short course about business process management for public sector at the Maastricht School of Management. He led a team of programmers and successfully developed an integrated and interoperable audit information system, which is still used until today. He also contributed a lot in designing BPK's information system architecture. In 2012, he was appointed as the Bureau's exemplary employee of the year. Arie was promoted as the head of the IT and General Affairs Subdivision of BPK's South Kalimantan office in 2014. In this role, he contributed to establishing corruption-free public services and obtained the government procurement certification. During this work, he became interested in open government data.

In July 2016, Arie received a doctoral scholarship from the Education Endowment Fund (Indonesian: *Lembaga Pengelola Dana Pendidikan*; LPDP) to research citizen engagement with open government data in the Information and Communication Technology research group at the Technology Policy and Management Faculty at the Delft University of Technology. During the doctoral program, Arie has jointly supervised a master thesis that was based on his research topic, namely open budget data, and authored a number of journal articles and conference papers listed in Appendix H. He was also involved in reviewing manuscripts submitted to several journal and conference outlets such as Government Information Quarterly, Heliyon, Information Polity, Transforming Government: People, Process and Policy, Annual International Conference on Digital Government Research (dg.o), European Conference on Information Systems, and International Conference on Information Systems.

CITIZEN ENGAGEMENT WITH OPEN GOVERNMENT DATA

A MODEL FOR ANALYZING FACTORS INFLUENCING CITIZEN ENGAGEMENT

Governmental organizations are increasingly making non-privacy-restricted and non-confidential data publicly available on the internet that anyone can freely use, reuse, and distribute without any restrictions. The opening up the government data enables citizens to engage with OGD, i.e., to collaboratively convert OGD into valuable artifacts that benefit society. However, citizen engagement with OGD is contingent upon many factors, and researchers often overlook these factors, and insight is needed to stimulate engagement. This study develops the OGD Citizen Engagement Model (OGD-CEM) that can be used to understand factors contributing to citizen engagement with OGD.

The OGD-CEM model hypothesizes that (both extrinsic and intrinsic) motivations toward the engagement and perceived political factors toward OGD and government determine citizens' intention to engage with OGD. Notably, the more citizens perceive that engaging with OGD will give them an advantage and provide the opportunity to broaden their social networks, the more inclined they will be to engage with OGD. Furthermore, the more citizens perceive that they can engage with OGD easily, that engaging with OGD is enjoyable, and that OGD engagement challenges them intellectually, the more likely they will engage with it. Finally, the more citizens perceive that their engagement with OGD will influence public policy, and the higher citizens' trust in OGD and the governmental organizations that provide it, the more they will be inclined to engage with OGD.

This study is among the first to provide an integrated overview of the profiles of citizens who engage with OGD and the drivers and inhibitors of citizen engagement with OGD. This dissertation also contributes to science by integrating political participation and trust theories into the model because engaging with OGD can yield outcomes with high political values. Moreover, this study is among the first to provide empirical evidence about the citizen-led OGD engagement. Finally, this study is also among the first to empirically test assumptions held in the open data research regarding the effects of OGD quality, political participation, and trust on the intention to engage with OGD.