Connecting supply and demand through portals: A study of open government data in China

By

Di Wang

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
in the
Department of Computing
Faculty of Science and Engineering
Macquarie University

2019

Statement of Candidate

I certify that this thesis entitled ***Connecting supply and demand through portals: A study of open government data in China*** is being submitted to Macquarie University and Wuhan University in accordance with the Cotutelle agreement dated 12/07/2017. To the best of my knowledge and belief, the thesis has not been previously submitted for a degree nor has it been submitted as part requirement for a degree to any university or institution other than Macquarie University and Wuhan University.

I also certify that the thesis is an original piece of research and it has been written by me (Di Wang). Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis was approved by the Human Research Ethics Committee, Macquarie University with reference number: 5201800056. A copy of the final ethics approval letter can be found in Appendix A.

Di Wang

This work of mine is dedicated to

my dear mom & dad

for your non-stop love and support!

Acknowledgments

First and foremost, I would like to express my great gratitude to my supervisor, Prof. Deborah Richards. Your immense knowledge and research experiences guided me through my whole research. All her in-depth feedback and penetrating views helped me a lot both for my study and for my personal growth in carrying out scientific research. Thanks to your supervision, I have spent two years of fulfilling study time in Macquarie University, which enlightened my thoughts, broadened my views and elevated my critical thinking. I couldn't have completed the thesis without your help and advice.

I would also like to express my sincere appreciation to my associate supervisor, Prof. Ayse Bilgin, and my adjunct supervisor, Prof. Chuanfu Chen, for your support, motivation and suggestions during all the time of research and writing of this thesis.

Besides my supervisors, I would like to thank the people who have offered generous assistance with my studies, particularly Ms Meredith Taylor, who helped me with building the virtual agent for the experiment, and Ms Xu Chen, who helped me with setting OGD portals for the experiment.

I extend my thanks to Macquarie University, which provided the MQ Research Training Program (MQRTP), and the China Scholarship Council, which has provided financial assistance during this research. I would also like to express my appreciation for the kind support of the Department of Computing that has allowed me to complete this research. My acknowledgement also goes to all the science IT staff on duty and department office staff for their cooperation.

Professional accredited editor Mary-Jo O'Rourke AE provided copyediting and proofreading services according to the national university-endorsed 'Guidelines for the editing of research theses' (IPEd 2019).

Finally, I would like to thank everybody who was important to the successful realization of my thesis, as well as expressing my apology that I could not mention each personally one by one.

Abstract

Open government data (OGD) has developed rapidly due to various benefits that can be derived through transparency and public access. However, researchers have emphasized the lack of use instead of the lack of disclosure of OGD as a key problem in the present development of OGD. To find ways for better utilization of OGD by connecting the supply and demand sides through OGD portals, the author designed four connected studies. The first study built an evaluation framework for understanding the development of the supply-side of OGD by evaluating existing Chinese province-level OGD portals. Secondly, with the primary users on the demand-side and the major beneficiaries of OGD, the second study focused on a survey conducted to analyze citizens' awareness and utilization of OGD portals. A third study compared the supply and demand sides of OGD by using the data collected in the previous two parts based on Diffusion of Innovation (DOI) theory. Lastly, a final study tested the proposed usability criteria for building an OGD portal in helping users to use the data on the portal by carrying out a between-subjects experiment. All the case studies in these four parts were carried out in China.

This research recognizes that Chinese province-level OGD portals are in an early stage of development. Citizens have limited awareness of OGD and OGD portals. Significant relationships are recognized among citizens, their demands of OGD and their utilization of OGD. Significant conflicts lie on the supply and demand sides of OGD in relation to relative advantage, compatibility and complexity. Our experiment shows that following these usability criteria for building an OGD portal could improve acceptance of the portal. Improving the help functions would also help users to find the data they need. Based on the analysis results from this study, future directions for developing OGD and OGD portals are identified, including: (1) identifying different user types; (2) increasing the online visibility of OGD; (3) improving the help functions of OGD portals such as providing online smart agents; (4) strengthening user interaction with the portals; and (5) reinforcing OGD characteristics.

Publications

Results of the research work leading to this PhD thesis have been published in various journals and conference proceedings. Presentations have been given at international conferences:

1	Title	An Analysis of Interaction Between Users and Open Government Data Portals in Data Acquisition Process
	Publication details	K. Yoshida and M. Lee (Eds.): PKAW 2018, LNAI 11016, pp.184-200, 2018. Springer, Cham
	Authors	Di Wang, Deborah Richards, Chuanfu Chen
2	Title	A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments
	Publication details	Government Information Quarterly, Volume 35, Issue 4, 2018, Pages 644-656, ISSN 0740-624X.
	Authors	Di Wang, Chuanfu Chen, Deborah Richards
3	Title	Connecting Users, Data and Utilization: A Demand-Side Analysis of Open Government Data
	Publication details	N. G. Taylor et al. (Eds.): iConference 2019, LNCS 11420, pp. 488-500, 2019. Springer, Cham.
	Authors	Di Wang, Deborah Richards, Chuanfu Chen

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Term
AHP	Analytic Hierarchy Process
API	Application programming interface
CR	Consistency ratio
DOI	Diffusion of Innovation theory
FAQ	Frequently asked questions
GDP	Gross domestic product
HCI	Human–computer interaction
ICL	Intention – complexity
ICP	Intention – compatibility
ICT	Information and communication technology
IOB	Intention – observability
IRA	Intention – relative advantage
ITR	Intention – trialability
MM	Motivational model
OGD	Open government data
PBC	Perceived behavioural control
PEOU	Perceived ease of use
PSI	Public sector information
PU	Perceived usefulness
RQ	Research question
TAM	Technology acceptance model
TIPI	Ten-item personality inventory
TPB	Theory of planned behavior
TRA	Theory of reasoned action
TTS	Text to speech
UTAUT	Unified theory of acceptance and use of technology

1 CHAPTER 1: INTRODUCTION

The impressive development of information and communication technologies (ICTs) has had a profound effect on the world and has created a so-called "knowledge-based society" (David & Foray, 2003). It is assumed by scholars that the concept of information plays a more and more central role in the competitiveness of today's knowledge economy (Lundvall & Johnson, 1994). On the other hand, every interaction we have with any digital device is programmed to generate data, which results in the emergence of great amounts of data at any given moment (Ribeiro, 2017). The data flow in our modern society, together with the internet, "Offer unprecedented practical means to access, process, share, combine, organize and reuse vast amounts of information" (Iemma, 2012, p. 3).

The surge in data generation can also be noted in the realm of government. In fact, different levels of public departments produce or collect a wide range of different types of data during the process of performing their tasks, from maps, traffic and weather to company registers (Aichholzer & Burkert, 2004; M. Janssen, Charalabidis, & Zuiderwijk, 2012; Ubaldi, 2013). At the same time, by the influence of digital media and Web 2.0 technologies (Ribeiro, 2017), great amounts of data are generated, collected and stored because more and more governments nowadays are engaged in providing online public services for citizens (Kassen, 2013).

Traditionally, this vast amount of data was kept internal to the government (M. Janssen et al., 2012; Sieber & Johnson, 2015). Since citizens' right to public access to government data serves as a fundamental tenet of government transparency and the right to know (Conradie & Choenni, 2014; Jaeger, 2005; Jaeger & Bertot, 2010; Jaeger & Burnett, 2005; Zuiderwijk & Janssen, 2014), many civil society movements have campaigned in many countries around the world for greater openness of the information, documents and datasets held by governments (Ubaldi, 2013). Moreover, the internet has enabled the use of ICTs to satisfy society's desire for information and has motivated governments to make their governance more open and transparent (Corrêa, Paula, Corrêa, & Silva, 2017; Feeney & Brown, 2017; Sandoval-Almazan & Gil-Garcia, 2012). In this context, open data policies have been established in different countries, e.g., the United States' Open Government Initiative (Obama, 2009), the EU Commission's decision on the reuse of Commission documents (Barroso, 2011) and P.R. China's policy for the promotion of the development of big data (StateCouncil, 2016). Consequently, many open data initiatives have been proposed at different levels of government around the world (Conradie & Choenni, 2014), which has led to the birth of the core concept of this thesis, *open government data* (OGD).

1.1 Research context

The definition by Open Knowledge International makes precise the meaning of "open" in the term "open data". According to its definition, open data can be freely used, modified and shared by anyone for any purpose (OpenKnowledgeFoundation). OGD is commonly treated as a subset of open data and is simply government-related data that is made open to the public (Attard, Orlandi, Scerri, & Auer, 2015; Kučera, Chlapek, & Nečaský, 2013) with a possibility of redistribution in any form without any copyright restrictions (Chatfield & Reddick, 2017; Kassen, 2013). Due to its relationship with governments, OGD is also treated as a subset of public sector information (PSI) (Ubaldi, 2013), which is broadly defined as "information, including information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for a government or public institution" (OECD, 2008, p. 4). Thus, we could conclude that OGD is the intersection of open data and PSI, as shown in Figure 1.1. OGD not only contains datasets directly produced by the government, including those relating to budgets and spending, population and census, and also data which is collected and indirectly owned by public institutions, such as data relating to climate, public transportation, education, etc.

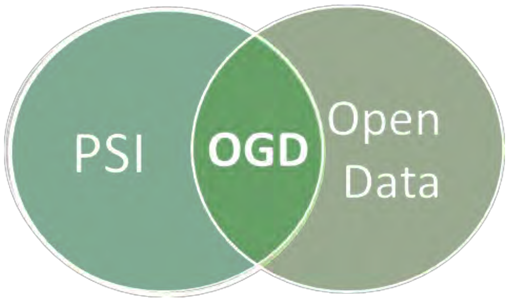


Figure 1-1 Relationship between open data, OGD and PSI

Over the whole life cycle of OGD, the opened data itself could be treated as a kind of product being created, published and consumed (Attard et al., 2015). Therefore, OGD by nature have stakeholders of supplier and consumer. The supply-side of OGD indicated procedures and stakeholders relating to making more and better quality government data publicly available (Tim Davies, 2011; Graves-Fuenzalida, 2013; Ohemeng & Ofosu-Adarkwa, 2015). While the demand-side of OGD refers to utilizations and users of the published data (Ohemeng & Ofosu-Adarkwa, 2015).

By releasing government data to the public, governments are able to "trigger profound changes in the relationship between governmental agencies and their stakeholders" (Ribeiro, 2017, p. 3). The implications of the opening up and utilization of OGD could contribute to public administration (Attard et al., 2015; Bertot, Jaeger, & Grimes, 2010; Dawes, 2010; Giorgi, Jones, & Asrar, 2009) including greater transparency, enhancing the accountability of government, as well as bringing economic benefits (Florini, 2008; Willinsky, 2005). Governments are also expecting an increase in collaboration and citizen participation from the opening up of government data (S. Martin, Foulonneau, Turki, & Ihadjadene, 2013). Overall, due to its potential impact on society, politics and the economy (Jetzek, Avital, & Bjorn-Andersen, 2012), OGD "has been hailed as one of the most important public policies of our time" (Halonen, 2012, p. 6).

Great benefits that are expected to be derived from OGD have led to the rapid development of OGD initiatives around the world (Jetzek, Avital, & Bjorn-Andersen, 2014). In these OGD initiatives, scholars have found publishing and consumption processes (Edelmann, Höchtl, & Sachs, 2012) to be the most essential ones in the whole OGD life cycle (Attard et al., 2015). Being the most commonly implemented approaches for publishing and consuming OGD (Attard et al., 2015), OGD portals have been treated as flagship initiatives (Lourenço, 2015) or even as OGD programs themselves (Kassen, 2013; Lourenço, 2013, 2015). Due to the functions of OGD portals and their ability to achieve the goals of OGD initiatives (Lourenço, 2013), they are also recognized as "services supporting the location of public sector information (PSI)" (Shadbolt et al., 2012, p. 17). As a result, many OGD portals have been launched by governments to make their data available to the public (Kassen, 2013; Lourenço, 2013), with examples like data.gov of the US government, data.gov.uk of the United Kingdom, data.gov.sg of Singapore, etc.

However, releasing OGD to the public through portals does not guarantee the successful promotion of transparency and accountability of local governments (Attard et al., 2015). On one hand, for the supply-side of OGD, many issues have emerged at local levels during the development process, such as technology diversity and lack of standardization (Armstrong, 2011), and the immaturity of local governments regarding how to correctly disclose data (Conradie & Choenni, 2014; Corrêa et al., 2017; Lourenço, Sá, Jorge, & Pattaro, 2013; Yavuz & Welch, 2014). On the other hand, although more and more data is provided through OGD portals (A. Meijer, de Hoog, van Twist, van der Steen, & Scherpenisse, 2014; Ruijer, Grimmelikhuijsen, Hogan, et al., 2017), researchers have noted a lack of use on the demand-side of OGD becoming a key problem and the most critical challenge in OGD development (Halonen, 2012; Ruijer, Grimmelikhuijsen, & Meijer, 2017; Safarov, Meijer, & Grimmelikhuijsen, 2017; Zuiderwijk & Janssen, 2013). Citizens' lack of knowledge concerning the existence of OGD and OGD portals (D. Wang, Richards, & Chen, 2019) also leads to possible failure in achieving the targeted aims of OGD programs (Attard et al., 2015; Heise & Naumann, 2012), such as stimulating innovation (M. Janssen et al., 2012) and creating economic benefits (Willinsky, 2005). The lack of use of OGD limits the social impact of OGD initiatives (OpenDataBarometer, 2018a). Therefore, at present, a disconnection could be noticed lying between the supply-side and the demand-side of OGD. For the supply-side, governments are trying to open up more data to the public through portals for the possible benefits that could derive from its utilization (Kassen, 2013). While for the demand-side, although citizens campaigned for more openness of the government information resources (Ubaldi, 2013), the utilization of the released data by citizens is limited and shows unobvious social impact (OpenDataBarometer, 2018a). This disconnection between the two sides of OGD restricted the benefits of OGD programs.

In addition, although the expectation of OGD benefits has drawn the attention and interest of the academic community in studying the topic of OGD (Zuiderwijk, Helbig, Gil-Garcia, & Janssen, 2014), reviews and evaluation by practitioners and researchers in the area of OGD, including successes and barriers, are still in their infancy (Attard et al., 2015; Zuiderwijk et al., 2014). Lots of issues, challenges and barriers can still be found in this field (Safarov et al., 2017; D. Wang et al., 2019).

1.2 Research objectives

The main aims of disclosing government data are for its use, reuse and distribution (Attard et al., 2015). However, in contrast to the continual development on OGD's supply-side, plenty of challenges can still be recognized when implementing OGD initiatives (Ribeiro, 2017), among which lack of use has been pointed out by scholars as a key problem for OGD (Ruijer, Grimmelikhuijsen, & Meijer, 2017; Safarov et al., 2017; Zuiderwijk & Janssen, 2013). As a result, inadequate evidence of real benefits from OGD initiatives can be found, particularly in relation to social impact (OpenDataBarometer, 2018a). With the development of OGD in the past few years, open data is driving more participation of citizens in public administration, such as the citizen participatory budgeting in South Korea that offers the chance to scrutinize government spending, and the government IT investments in Japan that can be monitored by the public. But reports have shown the stagnation of engagement between government and civil society (OpenDataBarometer, 2018a).

Since collaboration between the supply-side and demand-side of OGD is the premise for making OGD work for people, we set the ultimate goal of this study as to improve the utilization of OGD by balancing the supply-side and demand-side of OGD according to the demands of citizens through the development of OGD portals. In order to reach this goal, we set several objectives for this study:

O1. Evaluate the supply-side development of OGD.

O2. Understand the demands and motivations of citizens on the demand-side of OGD.

O3. Recognize the existing conflicts between the demand-side and supply-side of OGD.

O4. Find possible solutions for the conflicts and discuss future development directions of OGD portals to improve OGD utilization.

Our decision to carry out an evaluation of the supply-side of OGD is grounded in the notion that evaluations can provide a better understanding of the current stage of development, as well as ensuring the goals of initiatives are achieved

(Schellong, 2009). It can also support comparative study of the supply-side and demand-side of OGD.

Unlike the vast number of studies focusing on the supply-side development (Attard et al., 2015), the demand-side or the user-side of OGD remains a relatively understudied part in the whole OGD value-chain (Kitchin, 2013; C. Martin, 2014; A. Meijer et al., 2014). Jetzek et al. (2012) emphasized the need for a better understanding of the generation, capture and measurement of OGD. Users are the main actors on the demand-side due to their direct effect on the organization of OGD resources (Zuiderwijk & Janssen, 2013). Among different users, citizens are commonly identified as primary stakeholders who receive major benefits from OGD utilization (Parycek, Hochtl, & Ginner, 2014). But the acceptance and utilization of OGD from the citizen's perspective have not received sufficient attention in the current literature (Weerakkody, Irani, Kapoor, Sivarajah, & Dwivedi, 2017; Zuiderwijk, Janssen, & Dwivedi, 2015). We thus focus specifically on citizens of the demand-side due to their role in the life cycle of OGD.

The objectives of this thesis for balancing the supply-side and demand-side are not only due to the inadequacy of studies connecting the two sides of OGD (A. Meijer et al., 2014), but also well-aligned with the ideas of scholars for encouraging active engagement and collaboration between governments and different stakeholders (Dietrich, 2015; OpenDataBarometer, 2018a).

1.3 Research questions

Open data is just a kind of data, emphasizing the possibility and desirability of openness (Vassilakopoulou, Skorve, & Aanestad, 2018). To make government datasets publicly available, usually official portals are launched in OGD programs (Kassen, 2013; Kostovski, Jovanovik, & Trajanov, 2012; Lourenço, 2013, 2015). Being the most common approaches for publishing and consuming OGD (Attard et al., 2015), OGD portals are used by different governments to expose OGD on the web. Thus, they are not only treated as flagship initiatives of OGD programs (Lourenço, 2015), but also act as the bridge between the supply-side and demand-side of OGD. Since the most critical challenge in OGD development is the lack of use of OGD and the scarce extant research on this topic is mainly of an exploratory nature (Foulonneau, Martin, & Turki, 2014; Ribeiro, 2017), to achieve the proposed research objectives, we therefore raise the following generic research question for this thesis:

How to improve citizens' utilization of OGD by connecting the supply-side and demand-side through portals?

This generic question can be further divided into six more specific sub-questions:

RQ1. How to analyze the development of the supply-side of OGD?

RQ2. How can citizens' demand for and utilization of OGD be characterized?

RQ3. What is the relationship between the supply-side and demand-side of OGD?

RQ4. What are the decisive factors for citizens' utilization of OGD portals?

RQ5. How to connect the supply-side and demand-side of OGD through portals?

RQ6. What are the future directions for developing OGD portals?

The first research question (RQ1) focuses on the supply-side of OGD, aiming at understanding the present development of OGD through systematic evaluations of OGD portals, which reflected the first objective of this thesis (O1). Although many evaluations regarding country-level OGD portals have been carried out by different scholars and organizations (Attard et al., 2015; Lourenço, 2013; Thorsby, Stowers, Wolslegel, & Tumbuan, 2017), assessments of the potential and capabilities of OGD portals at local levels are limited (Chatfield & Reddick, 2017; Conradie & Choenni, 2014; Kassen, 2013; Petychakis, Vasileiou, Georgis, Mouzakitis, & Psarras, 2014). Thus, for the first research question, we choose to focus on analyzing the local development of OGD portals. We then divide RQ1 into four sub-questions:

RQ1.1 What framework can be used to assess local-level OGD portals?

RQ1.2 How to obtain priorities for elements of the framework?

RQ1.3 How to use the framework to analyze the present development of local OGD portals?

RQ1.4 How to use the framework to guide the future development of local OGD portals?

The second research question (RQ2) focuses on the demand-side of OGD, which aims at gaining a better understanding of the primary stakeholder of the demand-side, citizens, in relation to their demands and utilization habits of OGD, which reflected the second objective of this thesis (O2). We further divide RQ2 into three sub-questions:

RQ2.1 What are citizens' demands for OGD?

RQ2.2 What are citizens' demands for and utilization of OGD portal services?

RQ2.3 What are the relationships between citizens, their OGD demands and OGD utilization?

The third research question (RQ3) focuses on comparison of the supply-side and demand-side of OGD based on the previous studies of RQ1 and RQ2. Although related studies have been carried out on both the supply-side and demand-side, they failed to systematically analyze the relationship between both sides (A. Meijer et al., 2014). To figure out the possible relationships between these two sides of OGD, especially the existing conflicts as stated in the third objective of this thesis (O3), we thus raised RQ3 for the study. In order to find possible methods for analyzing the relationship between two sides, we further divide RQ3 into two sub-questions:

RQ3.1 What research model could be used for the comparison of the supply-side and demand-side of OGD?

RQ3.2 How to analyze the present relationship between the supply-side and demand-side of OGD by using the research model?

The fourth (RQ4), fifth (RQ5) and sixth (RQ6) research questions are aimed at finding possible solutions based on the results of RQ3 and figuring out future development directions for OGD portals to encourage the utilization of OGD, which reflect the fourth objective of this thesis (O4).

The fourth research question (RQ4) is based on comparison of the supply-side and demand-side of OGD. It focuses on the bridge between the two sides, that is, the OGD portal. From the aspect of the supply-side of OGD, in order for a portal to be accepted by its users, it must first be usable and possess the functionality that can satisfy the needs and requirements of its users. From the aspect of the demand-side of OGD, citizens' acceptance of the portal is the first step in using the portal to get access to the data. However, at present, the decisive factors in citizens' actual utilization of OGD portals are unknown (Rana, Dwivedi, & Williams, 2015). To understand the decisive factors in citizens' utilization of OGD, we further divide RQ4 into four sub-questions:

RQ4.1 How to evaluate the usability of current OGD portals?

RQ4.2 Can citizens use the OGD portal with the best usability from the usability evaluation?

RQ4.3 What are the factors affecting citizens' acceptance of the OGD portal with the best usability from the usability evaluation?

RQ4.4 What are the factors affecting citizens' actual utilization of the OGD portal with the best usability from the usability evaluation?

In response to the findings from Chapter 5, one further sub-question was added to focus on alternative help functions:

RQ4.5 What is the effect of using traditional help functions compared to using a virtual assistant?

The last two research questions (RQ5 & RQ6) focus on the future development of OGD and OGD portals. Because the main aims of disclosing government data are for its use, reuse and distribution, these two research questions concentrate on finding possible ways to develop OGD portals to stimulate the utilization of OGD.

The structure and the corresponding relationship of objectives and research questions of this thesis is shown in Table 1.1.

Table 1-1 Structure of objectives and research questions

Objectives	Research Questions	
	Level 1	Level 2
O1: Evaluate the supply-side development of OGD.	**RQ1:** How to analyze the development of the supply-side of OGD?	**RQ1.1:** What framework can be used to assess local-level OGD portals? **RQ1.2:** How to obtain priorities for elements of the framework? **RQ1.3:** How to use the framework to analyze the present development of local OGD portals? **RQ1.4:** How to use the framework to guide the future development of local OGD portals?
O2: Understand the demands and motivations of citizens on the demand-side of OGD.	**RQ2:** How can citizens' demand for and utilization of OGD be characterized?	**RQ2.1:** What are citizens' demands for OGDs? **RQ2.2:** What are citizens' demands for and utilization of OGD portal services? **RQ2.3:** What are the relationships between citizens, their OGD demands and OGD utilization?
O3: Recognize the existing conflicts between the demand-side and supply-side of OGD.	**RQ3:** What is the relationship between the supply-side and demand-side of OGD?	**RQ3.1:** What research model could be used for the comparison of the supply-side and demand-side of OGD? **RQ3.2:** How to analyze the present relationship between the supply-side and demand-side of OGD by using the research model?
O4: Find possible solutions for the conflicts and discuss future development directions of OGD portals to improve OGD utilization.	**RQ4:** What are the decisive factors for citizens' utilization of OGD portals?	**RQ4.1:** How to evaluate the usability of current OGD portals? **RQ4.2:** Can citizens use the OGD portal with the best usability from the usability evaluation? **RQ4.3:** What are the factors affecting citizens' acceptance of the OGD portal with the best usability from the usability evaluation? **RQ4.4:** What are the factors affecting citizens' actual utilization of the OGD portal with the best usability from the usability evaluation? **RQ4.5:** What is the effect of using traditional help functions compared to using a virtual assistant?
	RQ5: How to connect the supply-side and demand-side of OGD through portals?	
	RQ6: What are the future directions for developing OGD portals?	

1.4 Research design and roadmap

The whole research design of this thesis is based on the conflict paradigm (Babbie, 2015) and the theory of supply and demand by Karl Marx (Marx, 1844). The theory of supply and demand shows the tendency of the interaction between these two to reach a balance. Scholars have interpreted this balance using equilibrium theory (Sekine, 2004). In the scope of economy, the market behavior of individuals and groups were propelled by the imbalances between supply and demand (Marx, 1844). This imbalance between the supply and demand reflected possible conflicts between the two. Enlightened by the theory of supply and demand in the realm of economic analyses, we supposed that the development of the supply-side and demand-side of OGD may also be affected by the imbalances that exist between the two sides. These conflicts that now exist between the two sides could be possible ways for establishing new directions for OGD development, since the conflict motivates change and interactions between the two sides in order to reach a balance. The conflict paradigm, originally used in economic analyses, could also be appropriate in the context of struggles between different groups, including understanding relations among different departments, organizations, and student-faculty-administrative relations, etc. (Babbie, 2015). In our studies, we applied the conflict paradigm to analyzing the supply-side and demand-side of OGD, including understanding the interactions between users and OGD portals.

By applying the above paradigm and theory to this thesis as well as to achieve the objectives and answer the proposed research questions, we designed a four-stage study on the development of Chinese local OGD, as depicted in the roadmap of the research design in Figure 1.2.

We chose China to carry out each of the stages of the research study for three main reasons. Firstly, encouraged by various regulations launched by the Chinese government (Horsley, 2007; Piotrowski, Zhang, Lin, & Yu, 2009), recent years have seen the rapid development of OGD and emergence of OGD portals in China (R. Huang & Wang, 2016). Thus, the objects for our study already exist in China. Secondly, in contrast to the US, the UK and many other countries, the development of OGD portals in China started from local areas instead of building one nation-wide OGD portal. Until now, no national-level OGD portal in China has been founded. Moreover, the identified research gap in the literature for an up-to-date study of OGD development in China (R. Huang & Wang, 2016) also strengthened our intention to carry out the whole study in China. Lastly, as China is the largest developing country in the world, it includes multiple sub-areas and regions that represent different contexts, including population, economy, culture and policy. Thus, carrying out the study in China helped demonstrate the wide applicability of our research outcomes. Since the majority of countries around the world are developing countries, our study could also enlighten OGD developments in these countries.

In the first stage of the study, we designed a case study for local-level OGD portal evaluation in China. The first stage focused on the first research question about the supply-side of OGD. In the second stage of the study, we designed a survey-based research study to understand citizens' demands and utilization of OGD and OGD portals. This stage of study focused on the second research question about the demand-side of OGD. For the third stage of the study, we designed a comparative case study based on the first two stages of the supply-side and demand-side of OGD. The comparative study focused on RQ3 about finding the present relationships between the two sides of OGD. The last stage of the study was based on the results of Stage 3 about the conflicts existing between the supply-side and demand-side of OGD. It contained two parts: a usability evaluation and an experiment. The first part involved development of an evaluation framework for OGD portals and use of this framework to conduct a usability evaluation of OGD portals in China from the supply-side, focusing on RQ4.1. The output of this step was identification of the existing Chinese OGD portal with the highest usability. This provided an important foundation for designing a simulated portal for the experiment of the second part. The second part included an experimental study to answer RQ4.2 to 4.4 that aimed to discover the decisive factors affecting citizens' acceptance and utilization of OGD portals. By comparing the results of all four stages and related literature, we answer RQ5 and RQ6 in the discussion part of this thesis.

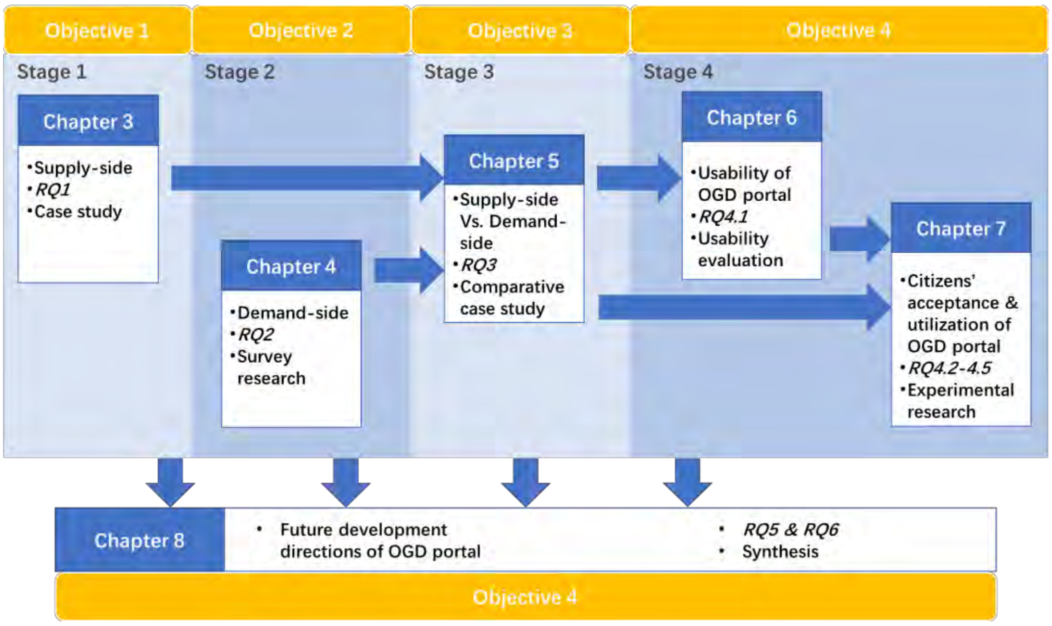


Figure 1-2 Research roadmap

1.5 Overview of the chosen methodology

According to the research design in the last section, a combination of both qualitative and quantitative research methods has been applied in different stages and studies. Generally, five research methods were used, which include: case study, survey research, evaluation, experimental research and comparative research. The overviews of each of these methods are as follows:

Case study is "a research strategy which focuses on understanding the dynamics present within single settings" (Eisenhardt, 1989, p. 534). It usually examines a single instance of some social phenomenon (Babbie, 2015). It is a popularly used qualitative research method in the field of information systems. A case study can investigate a phenomenon in complicated situations by using multiple data collection methods, thus leading to a better understanding of the topic (Recker, 2012). Besides simply describing and understanding phenomenon, the extended case method, which means entering "the field with full knowledge of existing theories" (Babbie, 2015, p. 298) and "laying out as coherently as possible" of what expected to find (Burawoy, Burton, Ferguson, & Fox, 1991, p. 9), has been found to be effective in discovering flaws in existing theories and providing modification. On the other hand, scholars have criticised the case study method for its limitation in generalizability of the observed outcome in one single instance of some phenomenon (Eisenhardt & Graebner, 2007). The challenge could be reduced, however, by applying multiple-case research to present relatively rich qualitative data (Babbie, 2015; Eisenhardt & Graebner, 2007). Usually, the comparative method is combined with the multiple-case research (Babbie, 2015). Therefore, in our study, we chose to carry out comparative multiple-case research in China by including different OGD portals of different sub-areas for theory testing, as well as for gaining insights into the OGD development in China from various aspects.

Survey research is a common method for gathering information about a group of selected units of observation. It is the best method available for describing a large population that is difficult for direct observation (Babbie, 2015). Use of standardized questionnaires also offers various strengths such as increased reliability, repeatability and comparability of measurement. The results of the collected data are usually analyzed with statistical techniques or other quantitative approaches (Recker, 2012). Therefore, survey research fits the needs of this study for understanding citizen's demands for OGD in China. However, besides its strengths in describing the characteristics of a large population, survey research has several weaknesses in flexibility and dealing with complex topics and the context of social life. In our thesis, we mainly used self-administered questionnaires for collecting information relating to the OGD demands and utilization characteristics of citizens in China. We tried to overcome the natural weaknesses of survey research by separating our research questions into small topics and designing questionnaires accordingly.

Evaluation study is a kind of measurement against a set of criteria (Symons, 1991), which plays a crucial role in different stages of information system development. Evaluation study could be divided into two main modes: the goal and system models (Etzioni, 1960; Symons, 1991). The goal model emphasises the achievement of predetermined outcomes as a measure of effectiveness

(Etzioni, 1960). The system model carries out rating according to a developed multi-dimensional range of criteria using scoring methods (Strasser, Eveland, Cummins, Deniston, & Romani, 1981). Weights are usually assigned to the selected criteria and aggregate scores are calculated after the evaluation (Symons, 1991). Strengths of the system-oriented evaluation lies in its ability to cope with the interdependencies and complexity of organizations due to its explicit monitoring of a whole array of variables (Symons, 1991). However, the selected criteria/indicators in the system-oriented evaluations are relatively straightforward (Strasser et al., 1981). The concentration on operational effectiveness of indicators can obscure the desired outcomes (Symons, 1991). In this thesis, we mainly used the system-oriented evaluation for understanding the present development of OGD portals and its usability in China because of the complexity of OGD development and the various stakeholders concerned in the development.

An experiment is “a mode of observation that enables researchers to probe causal relationships” (Babbie, 2015, p. 220). Experimental research has its advantage for gauging cause-and-effect relationships compared with observational research methods including surveys and case studies. By applying a certain treatment to one group of respondents but not to another group while maintaining stability of other potential factors between the two groups, experimental studies can show the effect(s) of the treatment (Recker, 2012). However, the greatest weakness of the experimental research method is due to its artificiality. What occurs in a laboratory setting might not necessarily occur in natural social settings (Babbie, 2015). In this thesis, we designed an experiment to test the effect of different help functions on a simulated version of an OGD portal on citizens’ ability to correctly use the portal to find the data they need. To overcome the possible weakness of the experimental research, we selected a target portal from existing OGD portals in China to imitate and build the experiment environment accordingly which could help simulating the true environment of OGD utilization for OGD users.

Comparative research belongs to unobtrusive research, which allows researchers to "study social life from afar, without influencing it in the process" (Babbie, 2015, p. 318). Usually, other research methods such as experiment (Van de Vijver & Leung, 1997), survey research and case study (Babbie, 2015) are combined with comparative research to provide sources of comparative data for analyzing. Van de Vijver and Leung (1997) have divided comparative research into four common types (Generalizability studies, theory-driven studies, psychological differences studies and external validation studies) depending on whether the orientation is exploratory or hypothesis testing, and on whether or not contextual factors are considered. In our study, we used two types of comparative research. Theory-driven comparative study, which emphasizes testing hypothesis about particular relationships between variables and outcomes (Van de Vijver & Leung, 1997), is used in analyzing survey results and experiment results. The most important strength of theory-driven comparative study lies in its explicit postulation of the relationship between variables and outcomes and the major weakness is its lack of attention for explanations of the observed differences (Van de Vijver & Leung, 1997). While psychological differences study, which applies a measurement instrument in at least two objectives to find the differences in averages, standard deviations, reliabilities or other properties of the instrument across different groups (Van de Vijver & Leung, 1997), is used in the evaluation studies and the comparison of the supply-side and demand-side of OGD. The strength of psychological differences study is its "open-mindedness" of the existing differences while the weakness of differences study lies in the interpretation of the observed differences (Van de Vijver & Leung, 1997).

In stage 1, comparative research method has been used in the comparison of different evaluation frameworks and principles in order to form an assessment framework for the case study in this stage. Case study together with the evaluation methods have been applied to understand the development of OGD in China. In stage 2, survey research has been carried out for investigating the demands of OGD users. Theory-driven comparative research method has been applied to the analysis of the data collected through the survey. In stage 3, a differences comparative study has been carried out for understanding the possible conflicts that lie between the supply-side and the demand-side of OGD. In stage 4, a system-oriented evaluation of the usability of OGD portals together with an experiment has been carried out together with the comparative study for comparing the data collected through the evaluation and the experiment. Details of the practical operation and procedure of each stages could be found in the methodology sections in Chapters 3 to 7.

1.6 Intended contributions

This thesis aims to encourage better utilization of OGD to achieve its potential effects on society by shedding light on how to develop OGD portals through matching the supply-side with the demand-side. The top contributions of this thesis can be summarized into two aspects.

For theoretical contributions, this thesis builds two evaluation frameworks: one for the supply-side to evaluate the present development of OGD and another to evaluate the usability of OGD portals. This thesis also builds models to reveal the relationships on the demand-side among citizens as the primary users of OGD and OGD portals (Parycek et al., 2014), their demands of OGD and their OGD utilization habits. A comparison framework is also built for analyzing the supply-side and demand-side of OGD for possible concords and conflicts between these two sides. Finally, theoretical models are derived for showing the factors affecting citizens' acceptance and utilization of OGD portals.

For practical implications, this thesis has carried out a series of case studies in China from both the supply-side and demand-side of OGD. Thus, it offers a systematic view of the present development of OGD and OGD portals in different areas of China. Also, based on the research results, this thesis proposes feasible suggestions for the future development of OGD portals. Last but not least, by applying the models built in this thesis, it is possible for stakeholders to recognize the factors in citizens' utilization of OGD portals and thus make changes to the OGD portals' functions accordingly, and education and training programs or help functions for OGD can be developed for citizens to increase the potential users of OGD.

In addition, this thesis also makes contributions to the present academic field by providing an updated review of related literature on the development of OGD and OGD portals, as well as the potential of using virtual assistants to provide portal support.

1.7 Thesis outline

The thesis is composed of nine chapters. The organization of this thesis is as follows:

Chapter 2 provides theoretical background information for the whole thesis and reviews related research about OGD and its utilization, usability evaluation and the virtual agent.

Chapters 3 to 7 comprise the empirical research part of this thesis. Chapter 3 is dedicated to the supply-side of OGD. In this chapter, an evaluation framework for analyzing and guiding the development of local OGD portals has been developed based on the comparison of related studies and principles. A case study of Chinese province-level OGD portals has then been carried out to test the capability of the framework, as well as to understand the present development of OGD in China.

Chapter 4 looks into the demand-side of OGD through a survey of Chinese citizens in relation to their demands and utilization of OGD and OGD portals. Analysis of the interaction among citizens, their demands of OGD and utilization of OGD has been carried out to reveal the relationships among these three.

Chapter 5 compares the results of Chapters 3 and 4 according to an extended model built on the Diffusion of Innovation (DOI) theory. The concords and conflicts between the supply-side and demand-side of OGD in China have been identified through the data analysis from four aspects: observability, relative advantage, compatibility and complexity.

Chapter 6 focuses on the usability of OGD portals. Firstly, a usability evaluation framework has been built to direct a heuristic evaluation of Chinese province-level OGD portals. Secondly, the heuristic evaluation has been carried out for selected Chinese province-level OGD portals to identify the best performing one to be imitated in development of the simulated experimental portal used in the following chapter.

Chapter 7 presents an experiment to determine how well the best-performing OGD portal could be used by citizens and also to evaluate alternative help functions. Help functions were chosen as the factor to manipulate based on the conflicts between the supply-side and demand-side recognized in Chapter 5. An experiment involving two versions of simulated portals was carried out on Chinese citizens. One group (the control) included all identified features in the usability evaluation framework, while the other (the experimental group) additionally included a virtual assistant to provide help for users. Analysis has been carried out to compare the experimental results from both groups. Models revealing the factors in citizens' acceptance and utilization of OGD portals are built through machine learning.

Chapter 8 includes a comparative analysis of the results presented in Chapters 3 to 6. Based on the results of the empirical research, we shed light on the future development of OGD portals to improve the utilization of OGD by matching the supply-side with the demand-side.

The thesis ends with a concluding chapter (Chapter 9) which revisits the research questions and summarizes the key findings and outcomes. The theoretical contributions and pragmatic implications, limitations and potential future work are also described in this chapter.

2 CHAPTER 2: LITERATURE REVIEW

To form the theoretical foundation of this thesis, in this chapter we describe research and topics related to our study from various aspects. Generally, the contents of the literature review follow the four stages designed for this thesis. This chapter starts with key concepts in Section 2.1 providing the general background for the whole thesis, which include the definition and understanding of OGD, the characteristics and impacts of OGD. We also introduce the concept of OGD ecosystem here. Based on the concept of OGD ecosystem, a review of the supply-side in Section 2.2 and the demand-side of OGD in Section 2.3 as well as the bridge between the supply-side and the demand-side in Section 2.4, are further developed, which offered support for the Stage 1 to 3 studies in this thesis. For the supply-side of OGD, summaries of the worldwide evaluation of OGD development have been presented to provide more contextual information and theoretical foundation for the Stage 1 study. For the demand-side of OGD, OGD users and OGD utilizations are further discussed to support the Stage 2 study. For OGD portals, more information about the development of OGD portals in local areas are also provided to support the research in Stage 1.

Concepts and models relating to the acceptance and usability of OGD are presented in Section 2.5 to provide theoretical foundation for Stage 4 of this thesis. After Section 2.5, three aspects are further discussed which are related to different stages of this thesis. Firstly, the OGD development in China including the political context of its development are thoroughly discussed in Section 2.6. The Analytic Hierarchy Process (AHP) is presented in Section 2.7 since it is an important theory of measurement being applied in Stage 1 of this thesis. Introductions and discussions of virtual agents are presented in Section 2.8 because it is the key element included in the experiment in Stage 4. We included the literature review of OGD development in China, AHP and virtual agents in this chapter to avoid lengthy explanations in later chapters which may reduce readability and disturb the logical flow and description of the related research process. This chapter ends with a synthesized conclusion about the discussed literature in Section 2.9. Figure 2.1 shows the structure and flow of the different concepts and topics in this chapter.

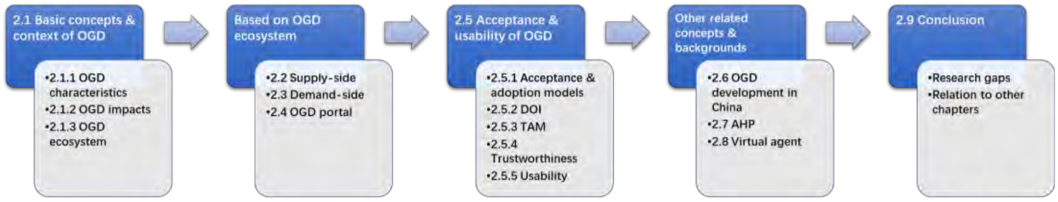


Figure 2-1 Structure and flow of the selected literature review topics

2.1 Open government data

The development of ICTs that grant citizens the ability to easily use, share and distribute information is leading to a culture of more openness (Van Veenstra & Van Den Broek, 2013), which has spreads to a wide range of areas in the contemporary world (Ribeiro, 2017) including governments (Kube, Hilgers, Fueller, & Koch, 2013). This phenomenon has given rise to the development of OGD. As discussed in Chapter 1, we accept the definition of OGD as the intersection of open data and PSI, indicating it to be the data, including products and services, "generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for a government or public institution" (OECD, 2008, p. 4) that can be freely used, modified and shared by anyone for any purpose (OpenKnowledgeFoundation).

Scholars have noted the lack of a common understanding of the concepts of OGD (Jetzek et al., 2012; Lindman, Rossi, & Tuunainen, 2013). Some focused on politically important disclosures, while others talked about data that is both open access and government-related but might or might not be politically important (Harlan Yu & Robinson, 2011). To reduce this possible ambiguity in the term OGD (Ribeiro, 2017), in our study we have looked into the issue of OGD from the second perspective by treating it as a type of open data related to the public sector and concerning more the technologies and attributes of open data rather than the politics of open government.

2.1.1 OGD characteristics

Since we treat OGD as a certain kind of open data, it possesses common characteristics that are shared in open data. In this section, we discuss the characteristics of OGD from three aspects: data quantity, data accessibility and data quality.

Data quantity for OGD means the volume of data available to the public (Kaisler, Armour, Espinosa, & Money, 2013). This has been treated as an indicator of the growth of an organization or service. Categories of datasets are often included as an indicator of whether governments have offered the kind of data people need (OpenDataBarometer, 2017b).

Data accessibility is a dimension reflecting the extent to which data is available (Pipino, Lee, & Wang, 2002), approachable (Hawker & Hawkins, 2001) and convenient to reach (Iwarsson & Ståhl, 2003). Because accessibility is the premise of data utilization, this concept is mentioned in almost all OGD principles, including the Eight Open Government Data Principles, the United Kingdom's Public Data Principles and Vivek Kundra's 10 principles for improving federal transparency. Given its primacy, we separate data accessibility from other forms of data quality.

Data quality "is most often characterized as simple accuracy" (Dawes, 2010, p. 379), but high quality indicates that data is "not only intrinsically correct, but also contextually appropriate for the task". Thus, data quality is commonly perceived to be fitness for use by data consumers (Attard et al., 2015; Dawes, 2010; R. Y. Wang & Strong, 1996). Quality attributes like accuracy, timeliness, precision, reliability, currency, completeness and relevance (R. Y. Wang & Strong, 1996) are often used in OGD principles.

2.1.2 OGD impacts

The constant release of large amounts of OGD by different governments has opened new opportunities to exploit the economic, social and political benefits of the data (Graves-Fuenzalida, 2013; A. Meijer et al., 2014). Many studies have confirmed the positive social and economic impacts (Ribeiro, 2017) of open data for citizens, researchers, companies and other stakeholders (Zuiderwijk et al., 2014). For social impacts, it is envisioned by many governments as promoting both efficiency and transparency at the same time through releasing government data to the public (Bertot et al., 2010). Case studies and statistical analyses indicated OGD's potential use in anti-corruption (Bertot et al., 2010) by "enhancing the effectiveness of internal and managerial control over corrupt behaviors" (Bertot et al., 2010, p. 265) and in promoting transparency and accountability (Florini, 2008). Through improving the reuse of government information and offering more efficient government data services (Graves-Fuenzalida, 2013), OGD development can also support public administration functions (Attard et al., 2015) and enhance citizen participation. For economic impacts, OGD itself is part of the emergence of "open innovation", which can endow stakeholders with the capacity to acquire knowledge from outside and thus result in the acceleration of internal innovation and expanded markets (Ribeiro, 2017; Willinsky, 2005). Therefore, generally, OGD development is treated as a beneficial initiative for the whole of society (Attard et al., 2015; Graves-Fuenzalida, 2013).

In contrast to the above suggested benefits, although governments have tried to prove the value generated by OGD initiatives (Horrigan, Rainie, & Page, 2015) and included the evaluation of OGD impacts as a central component in current public discourse (Ribeiro, 2017), researchers and organizations have criticized OGD initiatives for not yet delivering its promised positive impacts (Horrigan et al., 2015; OpenDataBarometer, 2018a). Inadequate evidence of real benefits from OGD initiatives, especially for social impact, circumstantially reflects the possible problems in the present development of OGD (OpenDataBarometer, 2018a).

2.1.3 OGD ecosystem

The situation for providing and making use of OGD has been found to be similar to a natural ecosystem (Graves-Fuenzalida, 2013), because there are participants providing data which is later consumed by others. Thus, T. M. Harrison, Pardo, and Cook (2012) firstly proposed the concept of an "open government data ecosystem", conveying "a sense of the interdependent social systems of actors, organizations, material infrastructures, and symbolic resources that can be created in technology-enabled, information-intensive social systems" (T. M. Harrison et al., 2012, p. 900). In such ecosystems, the core resource is OGD, which differentiates it from other ecosystem analogies (Heimstädt, Saunderson, & Heath, 2014). However, the success of this ecosystem relies on more than the dataset (Tim Davies, 2011). Another important feature of the environment is a set of systems and mechanisms (Tim Davies, 2011) that "allow participants to provide and consume Open Government Data" (Graves-Fuenzalida, 2013, p. 12). Considering the specific needs around OGD, Attard et al. (2015) proposed the life cycle of OGD as is shown in Figure 2.2, focusing on the essential processes of OGD through OGD portals from two aspects: publishing and consuming of open data. The publishing process includes two sections, preparing the data to be published (the four steps of the Pre-processing section) and maintaining the published data (Data curation of the Maintenance section). The consumption process includes the four steps of the Exploitation section: Data interlinking; Data discovery; Data exploration; and Data exploitation.

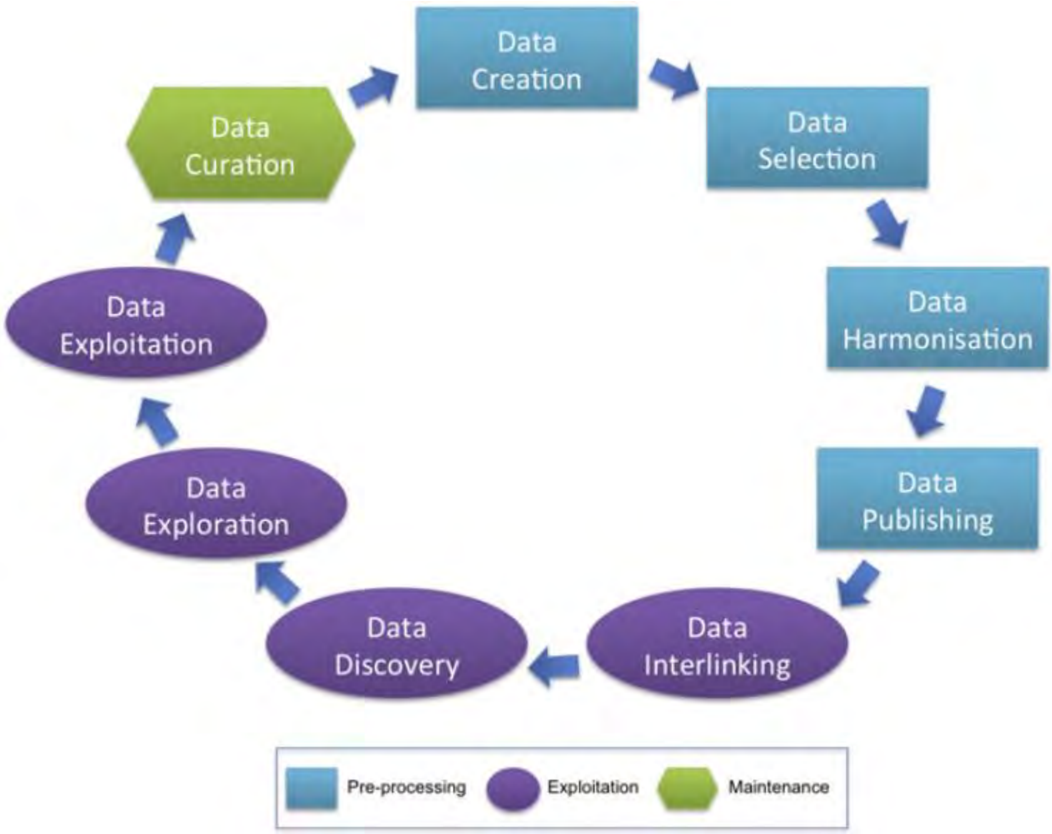


Figure 2-2 Open government data life cycle (Attard et al., 2015)

2.2 Supply-side of OGD

Open data is just data, emphasizing the possibility and desirability of openness (Vassilakopoulou et al., 2018). To make the data valuable, several steps are needed, such as collecting the raw data, making it available to others or further analyzing, combining and presenting the data in ways that make it useful for users to interpret as information (Lindman et al., 2013). Therefore, the OECD has included products and services together with data in the definition of government data (OECD, 2008).

Due to the definition of OGD and the targeted aims of OGD initiatives (Heise & Naumann, 2012) to promote citizen engagement (Kassen, 2013) and government transparency and accountability (Attard et al., 2015), the supply-side of OGD refers to making better quality or more government data available to the public (Ohemeng & Ofosu-Adarkwa, 2015), focusing on the publishing aspect of the

OGD ecosystem (Tim Davies, 2011; Graves-Fuenzalida, 2013). According to the OGD life cycle presented in Section 2.1.3, the supply-side covers the processes of data creation, selection, harmonization, publishing and curation (Attard et al., 2015). In order to prepare the mechanisms of the OGD ecosystem for the publishing and consuming of OGD, usually official portals are launched in OGD programs to make governmental datasets publicly available to citizens (Kassen, 2013; Kostovski et al., 2012; Lourenço, 2013, 2015), which puts OGD portals in an important position on the supply-side. Despite the impressive progress of OGD development on the supply-side, OGD portals suffering from many shortcomings (Agrawal, Kettinger, & Zhang, 2014) including bad design, flawed execution (Peled, 2011) and limited data subjects (Ribeiro, 2017) due to the limitation in IT and information management capabilities of local governments (Agrawal et al., 2014).

2.2.1 OGD evaluation

Due to the important role of OGD portals on the supply-side, many different associations and researchers have presented evaluation frameworks for assessing OGD portals around the world to target the shortcomings in the present development. Until now, no agreed-upon framework has been provided in the literature (Attard et al., 2015). A more detailed comparison of the existing frameworks is presented in Section 3.2.1 as part of our methodology and is the basis for proposing our own framework.

Several limitations of the existing frameworks have also been recognized by reviewers. To start with, although many OGD portal evaluation frameworks have been presented by different associations and researchers (Attard et al., 2015), they have mostly failed to associate these with practical applications (Attard et al., 2015) and several evaluations are not based on a systematic framework (Lourenço, 2013). Only some of these frameworks have been applied to practical evaluations, like the Open Data Barometer by the World Wide Web

Foundation (OpenDataBarometer, 2018a), the Open Data Index by the Open Knowledge Foundation (OpenDataIndex, 2017), the Open Data Monitor by the European Union (OpenDataMonitor, 2017), the OUR Data Index by the OECD (OECD), the Data Openness Indicator (Bogdanović-Dinić, Veljković, & Stoimenov, 2014), the Open Data Portal Index (Thorsby et al., 2017) and the Open Data Poral Requirement (Lourenço, 2015). Reviewers have also pointed out that the evaluations are based on compliance with the law, instead of the usefulness of the data provided on the portal (Attard et al., 2015). Considering these limitations of the present frameworks and the characteristics of local-level OGD portals (see Section 2.2.1) a prioritized evaluation framework for local OGD portals could thus contribute to the present literature. The lack of clarity in the prioritization process in existing frameworks is another gap this thesis seeks to address.

Moreover, while some of these evaluations are based on compliance with the law (Attard et al., 2015), others focus on the data quantity (OpenDataMonitor, 2017) or data quality (Bogdanović-Dinić et al., 2014; Craveiro, Santana, & Pereira, 2013; OpenDataBarometer, 2017b; OpenDataIndex, 2017; Thorsby et al., 2017) on the portal. Although some researchers have considered the interaction of the portal with end-users in their evaluation frameworks, like the help (Thorsby et al., 2017), visualization (DMGLabFudanUniversity, 2018; Thorsby et al., 2017), ranking (DMGLabFudanUniversity, 2018) and voting functions (OECD), none of these evaluation frameworks are specially designed for usability assessment of OGD portals. The Five Star Scheme proposed by Tim Berners-Lee focused on the assessment of the degree of dataset re-usability (Ubaldi, 2013). However, this scheme only considered the usability of the opened data but failed to take the usability of OGD portals or other services into account. Thus, their evaluation of the interaction design is not comprehensive and systematic, which also leaves a gap in the present literature.

2.2.2 Comparative analysis of international OGD

By applying the evaluation frameworks discussed in the last section, different researchers and associations have carried out comparative analysis of the development of OGD in different countries around the world.

The most widely implemented evaluation is Open Data Barometer carried out by World Wide Web Foundation. They included 115 different countries in their latest full version of report of detailed analysis of open data development from three aspects: readiness, implementation, and impacts (OpenDataBarometer, 2017a). Open Knowledge Foundation (OpenDataIndex, 2016) and the Organisation for Economic Co-operation and Development (OECD) (OECD, 2015) also carried out worldwide open data evaluations which included more than 30 different countries by applying the evaluation frameworks proposed by themselves. The European Union published Open Data Monitor, which focuses 32 countries in Europe for their open data development (OpenDataMonitor, 2017). All these studies have shown an imbalanced development of OGD among different countries. However, ranks of the same country in different evaluation frameworks were quite different, which indicated the great difference of these evaluation frameworks and its effect on the evaluation results.

Some scholars also carried out evaluations of OGD development in different countries by applying their own evaluation frameworks. Lourenço (2015) carried out evaluations of OGD portals in 7 different countries from the aspect of transparency and accountability. Some scholars carried out the comparative analysis of OGD portals with a mix of the level and areas. Bogdanović-Dinić et al. (2014) applied their data openness indicator in both countries like the US and Austria as well as areas like Queensland in Australia and Ottawa in Canada in the same study.

Other scholars chose to apply the evaluation frameworks proposed by other scholars as the basis for the comparative analysis. M. Yi (2019) chose the Five Star Scheme proposed by Tim Berners-Lee for a comparative study of the data format and completeness in the UK, the USA and Korea. While Wieczorkowski and Pawełoszek (2018) combined the framework of the Five Star Scheme and the Open Data Barometer for a comparative analysis of OGD portals in Poland, the USA, the UK and Germany.

Scholars' disagreement on the evaluation elements and frameworks lead to various kinds of comparative analysis results and ranks derived from these evaluations. We could notice that usually, the comparison between different countries for its OGD development did not consider the social context including policies and economic development. Open Data Barometer is the only one that looks beyond the data itself. But at the same time, OGD development is an issue which consists of different stakeholders and aspects (Attard et al., 2015) and is influenced by various kinds of social context (OpenDataBarometer, 2016). Thus, it is difficult to include all aspects of OGD initiatives in one evaluation framework for a comparative study.

2.3 Demand-side of OGD

The demand-side of OGD refers to better usage and reusage of the published data (Ohemeng & Ofosu-Adarkwa, 2015). In the OGD ecosystem, the demand-side of OGD refers to the consumption of published OGD, including data interlinking, discovery, exploration and exploitation in the OGD life cycle (Attard et al., 2015). The fast-developing ICTs nowadays have equipped citizens with the ability to discover, share, distribute and innovate with data (Ubaldi, 2013). However, despite the constant growth of OGD on the supply-side, lack of use on the demand-side has been recognized by scholars as a key problem in OGD development (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017; Safarov et al., 2017; H.-J. Wang & Lo, 2016; Zuiderwijk & Janssen, 2013). Although scholars have tried to understand the decisive factors in OGD usage and conditions (H.-J. Wang & Lo, 2016), effective methods for encouraging citizens' consumption of published OGD have not yet been found (Safarov et al., 2017). Also, the interactions between users and OGD portals have been less considered in the literature (A. Meijer et al., 2014), which leaves a gap in understanding the relationships between the supply-side and demand-side of the whole OGD ecosystem.

2.3.1 OGD users

Users are the main actor on the demand-side due to their direct effect on the organization of OGD resources (Zuiderwijk & Janssen, 2013). Researchers have divided OGD users into different types, like citizens (Parycek et al., 2014), business (Magalhaes, Roseira, & Manley, 2014; Susha, Grönlund, & Janssen, 2015), researchers (Gonzalez-Zapata & Heeks, 2015; Whitmore, 2014), developers (Veeckman & van der Graaf, 2015) and journalists (Heise & Naumann, 2012). Among these, citizens are commonly identified as the primary stakeholder and major beneficiary in OGD utilization (Parycek et al., 2014). Being able to access various kinds of OGD, citizens can supervise governments, which then become more transparent and accountable for their actions (T. G. Davies, 2013).

Scholars have asserted that mechanisms of participation in OGD should not be limited to "techies" like developers and scientists (A. J. Meijer, Curtin, & Hillebrandt, 2012). However, despite the important role of citizens on the demand-side for the utilization of OGD and researchers' growing interest in citizen participation in OGD (Harlan Yu & Robinson, 2011), they are actually "less researched as subjects" in the present literature (Safarov et al., 2017, p. 16). Some studies have targeted citizens' participation in the OGD value-extraction process (Safarov et al., 2017), including budgeting initiatives (Baiocchi & Ganuza, 2014) and disaster relief (Jetzek & Avital, 2013). Others focus on citizens' trust in governments' online services (Lim, Tan, Cyr, Pan, & Xiao, 2012; Venkatesh, Thong, Chan, & Hu, 2016). But non-computational skills like understanding users' needs, interface design and marketing have not been embraced substantially for exploring solutions to increase OGD values and impacts (Ribeiro, 2017; Safarov et al., 2017), which leaves a gap in the present literature.

2.3.2 OGD utilization

The utilization of OGD refers to its exploitation by practitioners for a particular purpose (Safarov et al., 2017). According to their separate aims, the usage of OGD has been divided into different types, including innovation (Magalhaes et al., 2014; Veeckman & van der Graaf, 2015), data analytics (Kalampokis, Tambouris, & Tarabanis, 2013; Kuhn, 2011), decision-making (Chakraborty, Wilson, Sarraf, & Jana, 2015; Desouza & Bhagwatwar, 2012), anti-corruption (Leontieva, Khalilova, Gaynullina, & Khalilov, 2015; Rajshree & Srivastava, 2012) and research (Kalampokis et al., 2013; Whitmore, 2014). Similar to the classification of users of OGD, the scope of different types of utilization may also overlap since some, like innovation and decision-making, are very broad, while others, like anti-corruption, are very specific. Scholars have tried to link the OGD utilization types with certain types of users. Some believe that ordinary citizens often cannot "perform the essential operations needed to collect, process, merge, and make sense of the data" (Graves & Hendler, 2013, p. 136). Often used in data journalism for investigation or reporting, OGD has become a critical component for journalists (Appelgren & Nygren, 2014). OGD characteristics like having machine-readable formats and being free of charge also benefit academic research (Arzberger et al., 2004; Whitmore, 2014).

It is expected that "while the supply-side makes data available, the demand-side builds something useful on the data" (Lindman et al., 2013, p. 1241). However, researchers have pointed out that in reality there is a separation of supply and demand in OGD development (Gurstein, 2011). Despite their expectation of participation, citizens are reluctant to use OGD (Gauld, Goldfinch, & Horsburgh, 2010). Lack of use has already become a key problem in the development of OGD (Ruijer, Grimmelikhuijsen, & Meijer, 2017; Safarov et al., 2017). As discussed above, making data publicly available is the first step for end-users to consume OGD services (Lindman et al., 2013), but there is no evidence that an emphasis on data and ICTs supporting data access and interoperability leads to significant increases in users' participation (Evans & Campos, 2013). Data itself cannot automatically create value (Cranefield, Robertson, & Oliver, 2014). Purely releasing government data in its raw form without considering the demands of citizens will not bring meaningful information to the public (Cornford, Wilson, Baines, & Richardson, 2013; Gitelman, 2013). On the other hand, very few studies have been carried out to specifically focus on the demand-side (Gauld et al., 2010). However, "recognizing the needs and acceptance of individuals is the beginning stage of any businesses" and "is crucial for the further development of any new technology" (Taherdoost, 2018, p. 960). Therefore, the acceptance and utilization of OGD from the citizen's perspective has not received enough attention in the present literature (Weerakkody et al., 2017; Zuiderwijk et al., 2015). Also, more work needs to be done to link the supply-side and demand-side to truly realize the value of OGD (Gurstein, 2011).

2.4 OGD portals as the bridge

Researchers have treated OGD programs as the launching of official portals at federal and local levels to make government datasets publicly available (Kassen, 2013; Lourenço, 2013, 2015), thus OGD portals are flagship initiatives of OGD programs (Lourenço, 2015). OGD portals are used by different governments to expose OGD on the internet and make it available to the end-users (Kostovski et al., 2012), which makes them the most common approaches for publishing and consuming OGD (Attard et al., 2015). Considering the concept of OGD as an ecosystem (Graves-Fuenzalida, 2013), we may find that OGD portals actually provide the mechanisms for different stakeholders to provide and consume OGD. Therefore, they act as the "bridge" between the supply-side and the demand-side of OGD. In our study, we treated an OGD portal as a "one-stop shop" for data consumers to obtain access to OGD, saving them the trouble of collecting data from various authorities, offices or websites (Attard et al., 2015). Typical examples of OGD portals include data.gov of the US government, data.gov.uk of the UK government and data.gov.sg of Singapore. According to its definition and functions, researchers have treated OGD as the most important foundation for an OGD portal (Bogdanović-Dinić et al., 2014; Braunschweig, Eberius, Thiele, & Lehner, 2012; Veljković, Bogdanović-Dinić, & Stoimenov, 2014).

2.4.1 Local governments and local OGD portals

Although OGD is not a new concept in e-government worldwide, OGD portal development at local government level remains in an early stage (Thorsby et al., 2017). Compared with federal- or country-level OGD portals, local-level ones have their own characteristics and problems during their development.

Regarding open data policies that support OGD portal development, lower level governments must comply with the policies set at a higher level (Zuiderwijk & Janssen, 2014). Directives from higher levels also put pressure on lower ones to find answers to pressing questions including the challenges, opportunities and barriers in releasing OGD on portals (Conradie & Choenni, 2014). However, in response to the increasing call for OGD, in comparison with national governments many local governments are actually facing unknown tasks and are not fully ready to adopt OGD (Conradie & Choenni, 2014; K. Janssen, 2011). Thus, many issues have emerged during the development process due to the immaturity of local governments regarding how to correctly disclose data (Conradie & Choenni, 2014; Corrêa et al., 2017; Lourenço et al., 2013; Yavuz & Welch, 2014). The limited budgets, IT infrastructure and skills at the local level

(Ebrahim & Irani, 2005) make OGD portal development even harder. A prioritised guide for the development of OGD portals at local levels could help solve these problems.

On the other hand, scholars have noted ministerial-level governments consuming instead of producing data, compared with lower level ones (Zuiderwijk & Janssen, 2014), identifying local governments as the producers of OGD. But present studies lack a focus on local-level OGD portals (Conradie & Choenni, 2014; Petychakis et al., 2014; Zuiderwijk, Janssen, & Parnia, 2013).

2.5 Acceptance and usability of OGD

Because of the expected positive effects of OGD on the whole of society, as discussed in Section 2.1.2, different governments around the world are pushing for more government data to be made publicly available. Although governments have taken proactive steps to improve the availability and ease of use of OGD, scholars have pointed out "barriers oriented around human, organisational and technological factors to accessibility and usability of open data" that have prevented its widespread proliferation (Weerakkody et al., 2017, p. 286). However, the acceptance and usability of OGD and OGD portals have not drawn enough attention in the present literature, especially from citizens' perspective (Safarov et al., 2017; Weerakkody et al., 2017). Moreover, it is recognized by scholars that one of the most significant challenges in the development of OGD is stimulating public interest in effective OGD utilization (M. Janssen et al., 2012; Ubaldi, 2013). Acting as the bridge between the supply-side and demand-side of OGD, OGD portals are burdened with the responsibility of promoting the public's adoption of and access to the data, as well as encouraging their engagement with and utilization of the available data resources (Kassen, 2013). This also calls for more studies focusing on the acceptance and usability of OGD portals (Z. Huang & Benyoucef, 2014), where limited methodical and comprehensive research has been undertaken (Rana et al., 2015). In areas closely related to OGD, which include the open data and e-government services, different theoretical approaches have been used to explain users' adoption of and intention to use the data. Therefore, in this section we first review popular models and theories regarding technology acceptance, since the acceptance of a new technology is the beginning stage for any further development (Taherdoost, 2018). Three popular theories (DOI, TAM and Trustworthiness) in relation to users' adoption of technologies and the concepts were further discussed for their advantages and disadvantage, as well as their fitness for this study. Studies regarding the usability of OGD portals are also discussed.

2.5.1 Technology acceptance models and theories

Acceptance has been defined as "an antagonism to the term refusal and means the positive decision to use an innovation" (Taherdoost, 2018, p. 961). It has attracted the attention of both practitioners and scholars for its positive effect on the design, evaluation and prediction of user's intention and response to new technologies (Mathieson, 1991).

Lots of different models and theories have been developed to explain user's acceptance of new technologies with different factors, such as Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1977), Theory of Planned Behavior (TPB) (Ajzen, 1985), Technology Acceptance Model (TAM) (Davis, 1985, 1989), Diffusion of Innovation theory (DOI) (Rogers, 2010), Motivational Model (MM) (Davis, Bagozzi, & Warshaw, 1992), and Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003). Many studies have been conducted by applying these technology acceptance models and theories, while some others combined previous models or added new constructs to models for their study (Taherdoost, 2018). Comparison of these theories are shown in Table 2.1.

Models and Theories	Core Constructs	Definitions	Reference
Theory of Reasoned Action (TRA)	Attitude Toward Behavior	An individual's positive or negative feelings (evaluative affect) about performing the target behavior	Fishbein & Ajzen, 1977
	Subjective Norm	The person's perception that most people who are important to him think he should or should not perform the behavior in question	Fishbein & Ajzen, 1977
Theory of Planned Behavior (TPB)	Attitude Toward Behavior	Adapted from TRA	
	Subjective Norm	Adapted from TRA	
	Perceived Behavioral Control (PBC)	The perceived ease or difficulty of performing the behavior	Ajzen, 1985
Technology Acceptance Model (TAM)	Perceived Usefulness (PU)	The degree to which a person believes that using a particular system would enhance his or her job performance	Davis, 1989
	Perceived Ease of Use (PEOU)	The degree to which a person believes that using a particular system would be free of effort	Davis, 1989
	Subjective Norm	Adapted from TRA. Included in TAM2 only.	
Diffusion of Innovation (DOI)	Relative Advantage	The degree to which an innovation is perceived as better than the idea it supersedes	Rogers, 2010
	Compatibility	The degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters	Rogers, 2010
	Complexity	The degree to which an innovation is perceived as difficult to understand and use	Rogers, 2010
	Observability	The degree to which the results of an innovation are visible to others	Rogers, 2010
	Trialability	The degree to which an innovation may be experimented with on a limited basis	Rogers, 2010
Motivational Model (MM)	Extrinsic Motivation	The perception that users will want to perform an activity because it is perceived to be instrumental in achieving valued outcomes that are distinct from the activity itself, such as improved job performance, pay, or promotions	Davis et al., 1992
	Intrinsic Motivation	The perception that users will want to perform an activity for no apparent reinforcement other than the process of performing the activity per se	Davis et al., 1992
Unified Theory of Acceptance and Use of Technology (UTAUT)	Performance Expectancy	The degree to which an individual believes that using the system will help him or her to attain gains in job performance	Venkatesh et al., 2003
	Effort Expectancy	The degree of ease associated with the use of the system	Venkatesh et al., 2003
	Social Influence	The degree to which an individual perceives that important others believe he or she should use the new system	Venkatesh et al., 2003
	Facilitating Conditions	The degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system	Venkatesh et al., 2003

Table 2-1 Comparison of technology acceptance models and theories

Since in this thesis, we chose to apply DOI and TAM to our study, further details of these two theories are explained in Section 2.5.2 and 2.5.3. Discussions of other theories are as follows:

TRA is one of the most fundamental and influential theories of human behaviour which was drawn from social psychology (Venkatesh et al., 2003). Recently, it was also used for investigating individual's IT usage behaviour (Kuo, Roldan-Bau, & Lowinger, 2015). The main cognitive components in this model are attitudes toward behaviour and subjective norms (Fishbein & Ajzen, 1977). Main disadvantages of TRA include its lack of addressing the role of habit, the cognitive deliberation, and the moral factors (Taherdoost, 2018).

TPB extended TRA by adding a construct of Perceived Behavioural Control (PBC). It has been used to predict intention and behaviour in a wide variety of settings (Ajzen, 1985) including individual's acceptance of technologies (D. A. Harrison, Mykytyn Jr, & Riemenschneider, 1997; Mathieson, 1991). By adding PBC, a self-efficacy type factor is achieved (Taherdoost, 2018). Taherdoost (2018) claimed that the revised TPB model is more suitable for analyzing individual's voluntariness in choosing to use or not to use of information technology in the workplace.

Drawing from the general motivation theory in psychology for explaining behaviour (Venkatesh et al., 2003), MM was applied to understand new technology adoption and use (Davis et al., 1992). The extrinsic motivation is defined as the perception that users will want to perform an activity because it is perceived to be instrumental in achieving valued outcomes that are distinct from the activity itself, such as improved job performance, pay, or promotions. The intrinsic motivation is defined as the perception that users will want to perform an activity for no apparent reinforcement other than the process of performing the activity per se.

Based on comparison of eight individual acceptance models including TAM, TRA, MM, TPB, Combined TAM and TPB, Model of PC Utilization, DOI, and Social Cognitive Theory, UTAUT identified four core constructs for the acceptance of information systems, which are performance expectancy, effort expectancy, social influence and facilitating conditions.

By comparing the above technology acceptance models and theories, a diverse source of theory bases could be recognized, including psychology and social psychology theories. All these theories have been applied in various studies in predicting and explaining human behaviours in different contexts (Venkatesh et al., 2003). Similarities in core constructs could be recognized from Table 2.1, since some theories like TPB and TAM adopt constructs from other theories like TRA. On the other hand, the focus of these models differs from each other. Compared with DOI, TRA and TPB concentrate on explaining the behaviour of individuals, while DOI focuses more on the acceptance decision influenced by the technology characteristics (Venkatesh et al., 2003). DOI and TAM focus on beliefs about the technology whilst TPB integrates the notion of perceived outcomes when predicting users' behaviour (Taherdoost, 2018).

2.5.2 Diffusion of Innovation

DOI is a popular theory relating to users' adoption of new technologies (Rogers, 2010), allowing for examination of citizens' perceptions and the identifying factors that influence their decisions about their acceptance and utilization of OGD (Weerakkody et al., 2017). The DOI model includes five main constructs (relative advantage, compatibility, complexity, observability, and trialability) which can explain about half of the variance in users' technology acceptance rates (Rogers, 2010).

The acceptance and use of IT has long been of significant importance in information science (Zuiderwijk et al., 2015) because it is the first step in future development of an information system (Taherdoost, 2018). DOI is one of the most used models in issues relating to the acceptance and use of IT (Weerakkody et al., 2017). Rogers has synthesized sixty years of innovation-adoption research to develop DOI, which is regarded as the principal theoretical perspective in technology adoption (Kapoor, Dwivedi, & Williams, 2013). Thus, his theory has been well received and adopted in many research studies of technology acceptance through using the five aspects for data analysis (Pannell et al., 2006; Rijsdijk & Hultink, 2003; Taherdoost, 2018; M. Tan & Teo, 2000; Weerakkody et al., 2017). It has also been adapted to a broader disciplinary background than only frameworks for understanding technology adoption (Carter & Bélanger, 2005). A strength of DOI is the ability to include system characteristics, organizational attributes and environmental aspects in the analysis of the acceptance of a new technology. However, due to its concentration on innovation characteristics, it "has less power in explanatory and less practical for prediction of outcomes" compared with other technology acceptance models like TAM and TPB (Taherdoost, 2018, p. 964). Since the study of this thesis is based on OGD portals, the advantages of DOI in considering different characteristics of OGD portals makes it a better fit for our study. In addition, its ability to analyze users' perceptions of OGD utilization and the factors influencing their acceptance of OGD makes DOI an appropriate choice for our study.

OGD is commonly treated as a kind of "open innovation" (Ribeiro, 2017). While the supply-side tries to diffuse this kind of innovation for its positive effects on society (Lindman et al., 2013; Ubaldi, 2013), the demand-side acts as the adopter and user of OGD including its products and services (Safarov et al., 2017). The operation in the OGD ecosystem together with the nature of OGD fit the scope that the basic DOI model could be applied to, which strengthened our confidence in using DOI as the base for our study.

Although there are other frameworks for understanding technology adoption (Venkatesh & Davis, 2000), some experts in the field have found these frameworks to be stereotypical and lacking sufficient understanding of the nature of innovation-adoption processes (Stokes, Barker, & Pigott, 2014). Researchers want to find out not only "organizational, systemic, and contextual effects" in this process, but also "the push and pull effects of the innovators and the innovation adopters" (Weerakkody et al., 2017, p. 287).

DOI has not yet been applied to the investigation of the relationship between the supply-side and demand-side. Although many frameworks have been used for evaluating the supply-side of OGD (Attard et al., 2015) and examining factors in governments' adoption of OGD (Moon & Norris, 2005), little work has been done regarding citizens' intentions to adopt OGD (Safarov et al., 2017). Zuiderwijk et al. (2015) studied the acceptance and use of open data technologies, but they did not explain citizens' intention to adopt OGD itself. Weerakkody et al. (2017) analyzed predictors affecting users' acceptance and use of open data with DOI from the perspective of citizens, but they failed to link the supply-side with these predictors. Lim et al. (2012) and Venkatesh et al. (2016) investigated citizens' trust of e-government systems. But prior to choosing whether to trust a system or not, citizens first need to know of the existence of the system, which is, according to Rogers (2010), the observability of a system. Because the final purpose of our study is to identify future directions for developing OGD portals to increase citizens' acceptance based on the relationship between the supply and demand sides, the DOI model could thus provide theoretical foundations for the comparison of these two sides, as well as for analyzing citizens' adoption of OGD.

2.5.3 Technology Acceptance Model

The Technology Acceptance Model (TAM) is one of the most widely used (Venkatesh & Davis, 2000) models in the field of technology acceptance (Wu, 2009). TAM states that users' attitude towards the utilization of a system is influenced by its perceived usefulness (PU) and perceived ease of use (PEOU) (Davis, 1989). Davis (1989) (p320) defined PU as "the degree to which a person believes that using a particular system would enhance his or her job performance" and PEOU as "the degree to which a person believes that using a particular system would be free of effort". PEOU can also influence PU, because users benefit more from a system if it is easier to use.

TAM has been tested and validated for various users and systems, including e-government services (Carter & Bélanger, 2005). Many studies found support for PU and PEOU to explain a large portion of the variance for users' intention to use an information system (Davis, 1989). For e-government services, a high level of PU was found to improve users' adoption of a e-government system (Sang, Lee, & Lee, 2009).

Besides its widely used tests and validation, researchers have pointed out that the parsimonious nature of the TAM model may lead to its high frequency of use in the field of e-government services (Rana et al., 2015). However, its ignorance of other external factors, including social influence, limits its applicability beyond the workplace (Taherdoost, 2018). Due to its limitations, researchers have chosen to integrate it with other theories like DOI and trust for better prediction of adoption intent (Safarov et al., 2017).

2.5.4 Trustworthiness

Trustworthiness in this context is defined as “the perception of confidence in the electronic marketer’s reliability and integrity” (Belanger, Hiller, & Smith, 2002, p. 252). Citizens’ perception of trustworthiness can impact on their intention to use e-government services (Carter & Bélanger, 2005; Lim et al., 2012). Having confidence in both the government and the enabled technologies will encourage them to make use of the provided services (Carter & Belanger, 2004; Carter & Bélanger, 2005). McKnight, Choudhury, and Kacmar (2002) established a multidimensional model of trust based on the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1977), focusing on the initial trust of users. In their model, institution-based trust and disposition to trust are the antecedents to trusting beliefs/intentions (McKnight et al., 2002). Institution-based trust is “the sociological dimension of trust”, referring to users’ perception of the institutional environment, which in this case is the government and the portal. The decision to adopt OGD portals requires citizens to have enough confidence in both the government and the enabling technologies (Carter & Bélanger, 2005).

The study of Venkatesh et al. (2016) shows the importance of trust in reducing citizens’ uncertainty in e-government services according to the uncertainty reduction theory. Their study of government websites further confirmed the effect of data quality and website functions on citizens’ feeling of trust, which further affected their intentions to use the websites. The study of Teo, Srivastava, and Jiang (2008) also showed the positive relationship between trust in government and citizens’ trust in the portal. The information quality, system quality and service quality affect users’ trust in the portal as well. Therefore, including trustworthiness in the analysis of the acceptance of OGD will help consider the same issue from the perspectives of citizens' perceptions of risk and insecurity (Carter & Bélanger, 2005; McKnight et al., 2002).

2.5.5 Usability

According to the ISO 9241 standard on Ergonomics of Human System Interaction, the definition of usability in the field of human–computer interaction is "The extent to which a product [service or environment] can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (Petrie & Bevan, 2009, p. 2). Facing the present challenges of OGD development in generating greater user engagement (Safarov et al., 2017), usability is thus treated as a critical element in the success of e-government portals (Youngblood & Mackiewicz, 2012). Because the usability of an OGD portal influences citizens' day-to-day interaction with the portal (Z. Huang & Benyoucef, 2014), usability is also treated as one of the main causes for citizens' low engagement and utilization of OGD (Clemmensen & Katre, 2012).

Due to the importance of usability in the success of OGD portals, many scholars have emphasized for developers to regularly monitor and enhance the usability of portals in order to attract and satisfy users (Scott, 2005). Some studies have tried to evaluate the usability of e-government sites by using the g-quality method (Garcia, Maciel, & Pinto, 2005) and heuristic evaluation (Z. Huang & Benyoucef, 2014). However, usability is closely related to the users, goals and contexts of use under a particular set of circumstances (Petrie & Bevan, 2009). Taking the differences of e-government sites and OGD portals into account, it is thus necessary to carry out specific usability evaluations under the circumstances of the OGD portals, instead of the general e-government websites. However, in the present literature, not enough studies have been carried out specifically focusing on OGD portals. Tang, Gregg, Hirsh, and Hall (2019) have conducted heuristic evaluations of data processing capabilities on U.S. OGD portals, but only basic heuristic principles that fit all websites were included in this study. Although Fajar Marta (2016) has proposed a series of OGD portal design principles, no empirical evaluations based on these principles have been carried out to learn about the usability of OGD portals. All of the above discussions show a gap in the present literature for further studies of the usability of OGD portals.

2.6 OGD development in China

The Regulations of the People's Republic of China on Open Government Information (OGI Regulations) in 2007 ("Regulations of the People's Republic of China on Open Government Information," 2016) marked a change in the Chinese government towards becoming more transparent (Horsley, 2007; Piotrowski et al., 2009). This regulation emphasized the role of government information in serving people's production work and livelihood and their economic and social activities, which shares the same view of similar regulations in western countries such as the "Memorandum on Transparency and Open Government" act of the US[1]. Under the requirement and instruction of this regulation, all levels of governments in China have the responsibility for "promoting, guiding, coordinating and supervising the work of open government information in their respective areas of administration" ("Regulations of the People's Republic of China on Open Government Information," 2016, p. 52), which is similar to the "Open Data Policy" of the US [2] (Timothy Davies, 2014). Since then, lots of regulations and policies at country level have been further proposed by the State Council in China. In 2016, the State Council in China proposed a policy for promotion of the development of big data (StateCouncil,

---

[1] https://www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf

[2] https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

2016), which led to fast development of OGD portals at different levels of government (R. Huang & Wang, 2016). Until the end of 2018, 27 different regulations and policies have gone into effect regarding the development of OGD in China. Their contents cover areas including data life-cycle, data quality, data security and privacy, data infrastructure, as well as open organization of data (Wen, 2019). Due to the introduction of these policies, according to the evaluation of Open Data Barometer, the readiness of government policies of China saw a great improvement in 2017 (OpenDataBarometer, 2017b).

Encouraged by these policies, lots of OGD portals in different local levels have been launched by local governments. Unlike the US data.gov and UK data.gov.uk, at present there is no national level OGD portal in China (R. Huang & Wang, 2016), but at least 29 local-level OGD portals have been built until 2019 (DMGLabFudanUniversity, 2018; R. Li, 2018). This special development process of OGD portals together with the large number of local OGD portals in China makes it possible to carry out comparative study of OGD portals at the same local-government level.

Despite the rapid emergence of local OGD portals in China, although some scholars have carried out studies of certain OGD portals (R. Huang & Wang, 2016), systematic and practical evaluations in China are few in the present literature (Attard et al., 2015; Lourenço, 2013). In the international level, Open Data Census conducted one evaluation of four cities in China in 2014, but no later data has been reported (OpenDataCensus, 2016). Although Open Data Barometer (OpenDataBarometer, 2016) and OUR Data Index (OECD, 2015) have included China in their evaluation scope (see Table 3.1), they focus on datasets from all over the country instead of a particular portal. Nevertheless, the results of Open Data Barometer and OUR Data Index have shown the great gap in OGD development between China and other developed countries including the UK, Canada, Australia, the U.S, etc (OECD, 2015; OpenDataBarometer, 2017b). This indicated the large space for improvement for the OGD development in China and at the same time, it is essential and necessary to figure out future developing directions for OGD in China.

For research in China, DMG Lab in Fudan University proposed China Open Data Index in 2018. Their evaluation framework follows the general structure of Open Data Barometer (DMGLabFudanUniversity, 2018). But their focus of evaluation is the open data themselves instead of the development and services of OGD portals. R. Li (2018) noticed the imbalanced development of different local OGD portals in China, but did not provide a theoretical process to understand the causes for the imbalanced development. Zhou, Xia, and Dai (2019) included both Chinese OGD portals and government websites in their study, their evaluation of the development of open data in China is a combination of different provinces and cities. However, they did not test the effectiveness of the solutions they proposed for the future development of open data in China. Shen, Han, and Hu (2018) looked into the development of OGD from a different aspect. Their evaluation focused on the development speed of province-level OGD portals in China. According to their study, Guangzhou Portal is the one with the fastest developing speed. To conclude, we identified a gap in the literature for an up-to-date and systematic evaluation of local OGD portals in China to reflect present development and also a further study about the direction for future development of OGD portals in China due to the great gap between China and the top countries in OGD development.

2.7 AHP in evaluation

The Analytic Hierarchy Process (AHP) is a general theory of measurement through pairwise comparison and relies on the judgements of experts to derive priority scales (Saaty, 2008). Its process makes it elegant enough "for addressing and analysing discrete alternative problems with multiple conflicting criteria" (Steuer & Na, 2003, p. 502). Its general process contains three main operations, namely, hierarchy construction, priority analysis and consistency verification

(Ho, 2008). AHP starts by subdividing a problem into its component parts and arranging all possible attributes into multiple hierarchical levels constituting goals, criteria, sub-criteria and alternatives (Ho, 2008; Vaidya & Kumar, 2006). Thus, the structure has at least two layers: one layer of criteria and one layer of alternatives. After that, a pairwise comparison of each element in the same level is carried out. A scale ranging from 1/9 for 'least valued than', to 1 for 'equal', and to 9 for 'absolutely more important than' is usually used for this comparison (Saaty, 2008). The final operation is consistency verification to guarantee the judgements of the comparison are consistent. Thus, this is "regarded as one of the most advantages of the AHP" (Ho, 2008, p. 212). Calculations are performed to find the maximum Eigen value, consistency ratio (CR) and normalized value for each criterion/alternative (Saaty, 2008). Once the CR exceeds the limit, then a decision is taken based on the normalized values. Usually, a questionnaire is used for "comparison of each element and geometric mean to arrive at a final solution" (Vaidya & Kumar, 2006, p. 2).

AHP has become one of the most widely used multiple-criteria decision-making tools due to its flexibility in being integrated with different topics and techniques (Vaidya & Kumar, 2006). It is also quite common to use AHP to set priorities for further modelling (Badri, 2001). AHP has the strength as an effortlessly reasonable system by disentangling a complicated issue into several steps. It also requires no authentic information sets. On the other hand, the priorities derived from AHP depends on human judgments, which may be affected by emotions. It is also only fit for direct models (Karthikeyan, Venkatesan, & Chandrasekar, 2016).

Ranking and evaluation are often carried out after determining the importance of each criterion by AHP (Arbel & Orgler, 1990; Babic & Plazibat, 1998). Reviewers have also recognized its popular use in the field of resource planning (Vaidya & Kumar, 2006), especially with a combination of finance (Steuer & Na, 2003), such as allocating space in an academic environment (Benjamin, Ehie, &

Omurtag, 1992) and planning information resources in a healthcare system (C. Lee & Kwak, 1999).

In this thesis, we chose to apply AHP to derive priorities for elements in OGD evaluation framework is due to four reasons. Firstly, the ability of AHP in drawing priorities for modelling and its flexibility in various topics have been demonstrated by related studies (Badri, 2001; Vaidya & Kumar, 2006). Secondly, its strength in solving complicated issues (Karthikeyan et al., 2016) fits our needs for creating a multi-layer framework for OGD evaluation. Thirdly, its applicability in resource planning fits the aim of our proposed evaluation framework for directing the resources arrangements in local governments for future development of OGD. Finally, the process of AHP makes it possible to derive different priorities for different geographical areas which strengthens the generalizability of the evaluation framework proposed in this thesis.

2.8 Virtual agents

Virtual agents are defined as “a pictorial representation of a human in a chat environment” (Bahorsky, Graber, & Mason, 1998). Due to various utilizations, many terms have been used to refer to such representations, including “virtual agents” (Abbattista, Lops, Semeraro, Andersen, & Andersen, 2002), “embodied conversational agents” (Cassell, 2000), “interactive characters” (Isbister & Nass, 2000) and “animated interface agents” (Dehn & Van Mulken, 2000). Similar to a human agent, virtual agents on a screen can present various behaviours including speech, emotions, gestures and eye, head and body movements (Dehn & Van Mulken, 2000).

Virtual agents used on web pages can perform different behaviours and functions, such as gathering information from various web sources and presenting it to users (André, Rist, & Müller, 1998), serving as recommendation agents giving directions to reach a feasible choice set (Chattaraman, Kwon, Gilbert, & In Shim, 2011) and serving as customer service representatives to assist online transaction processes (Chattaraman, Kwon, & Gilbert, 2009). In our study, the virtual agent we created to provide help functions on OGD portals mainly acted as a recommendation agent and customer service representative.

Scholars believe virtual agents have potential benefits in assisting users' interaction with information systems (Lester et al., 1997). Therefore, studies have been carried out analyzing users' preferences in relation to virtual agents (Ossowska, Szewc, Weichbroth, Garnik, & Sikorski, 2016) and testing their performance in assisting users' interaction with websites (Chattaraman et al., 2011). However, no evidence of a clear persona effect, which is the general advantage of an interface with a virtual agent over one without a virtual agent, has been provided in the present literature, although positive effects in some aspects like entertaining have been found (Dehn & Van Mulken, 2000). Additionally, although scholars have tried to use virtual agents on e-commerce websites (Chattaraman et al., 2011) and digital platforms (Hoorn et al., 2004), studies of using virtual agents for OGD users on OGD portals are not found in the present literature, which leaves a gap in the study of virtual agents' potential impacts.

2.8.1 Advantages of virtual agents

Due to the nature of a virtual agent, it can enhance the communication of users and computers, as well as computers' ability to engage and motivate users. Despite its disadvantages including the cumbersome and expensive hardware imposed on its users (Dragone, Holz, & O'Hare, 2007), the high costs of robot programming and modeling, as well as the requirements of video and audio equipment for users compared with traditional ways of human-computer interaction (J. Li, Kizilcec, Bailenson, & Ju, 2016), virtual agents provide various benefits as listed below:

To start with, a virtual agent stands out for its interaction with users. Bryson (1996) emphasized the importance of interaction as one of the three main attributes of virtual reality. Burgoon et al. (2000) found that with an increase in the anthropomorphic features of an interface, people get greater understanding and derive more utility from the website, which consequently shows the interactive and social credentials of virtual agents (McGoldrick, Keeling, & Beatty, 2008).

Secondly, due to its interaction with the user of an information system, a virtual agent can provide better help. Generally, all websites include some forms of help system. However, studies have shown users' resistance to accepting and using the traditional help systems. Some researchers found users felt fear and loathing towards the help menu (Grayling, 1998), while others noted that users chose to use trial and error instead of the help system (Spool, 1997). Grayling (2002) presented three characteristics of a well-designed help system: they should be obvious to invoke, easily available and non-intrusive, which can be fulfilled by virtual agents due to their ability to simulate social interaction (McGoldrick et al., 2008).

Thirdly, different from traditional help pages, virtual agents are able to give nonverbal feedback. A virtual agent is a mixture of several media features including image, animation, message, voice and interactivity (Baylor & Kim, 2009), thus it can use nonverbal communication to influence users like gazes, gestures and facial expressions. Studies have shown that manipulating the motivational contents of a virtual agent can dramatically impact on users' beliefs and attributions (Baylor, Kim, Son, & Lee, 2005; Baylor, Shen, Warren, & Park, 2004).

Finally, rich communication can take place between an information system and its users with the help of virtual agents. Compared with traditional collaborative systems like chat, bulletin boards and discussion lists, virtual agents can offer guidance and direction during or after dialogue sessions (Soller, 2001). Thus, due to the nature of virtual agents, they can provide an adequate platform for rich communication and cooperation to take place (Eschenbrenner, Nah, & Siau, 2009).

2.8.2 Creating rapport with virtual agents

Rapport is originally a concept from psychology, referring to deepening interdependence over time as a result of instant responsiveness (Cassell, 2000). During a reciprocal conversation with a virtual agent, the feeling of mutual understanding and warmth in the interaction is called rapport (WEI-ERN, 2012). Tickle-Degnen and Rosenthal (1990) specified three essential components of rapport: mutual attentiveness, positivity and coordination. Mutual attentiveness refers to the focus of users directed to the conversation and the behaviour towards the other, which may be negative. Positivity covers various positive feelings in rapport with another such as friendliness and caring. Coordination conveys the equilibrium, regularity and predictability between interactants (Tickle-Degnen & Rosenthal, 1990). Although these three components are related, they are not equivalent. Rapport in the context of virtual agents can be divided into two categories (Ranjbartabar, 2016): immediate or instant rapport (Gratch et al., 2006) and long-term relationships (Bickmore & Cassell, 2005). For information systems that only provide short-term relationships with users, attraction from the beginning is important for establishing immediate rapport (Ranjbartabar, 2016), while long-term relationships including intimacy and friendship are built through longer periods of verbal and nonverbal behaviours (Cassell, Gill, & Tepper, 2007).

Users' sense of rapport with virtual agents is correlated with effective communication, greater liking and trust (Gratch et al., 2006), which consequently increases users' engagement in an information system and their intentions to use it (Carter & Bélanger, 2005). Considering the positive effects of virtual agents, as discussed in Section 2.6.1 and results from related studies showing their advantages in usefulness and comfortability (Dehn & Van Mulken, 2000), it is thus reasonable to try to include virtual agents in OGD portals for users' smoother interaction with the computer (Lester et al., 1997).

2.9 Conclusion

In this chapter, we have reviewed a wide range of literature related to OGD and OGD portals, the supply-side and demand-side of OGD, and the acceptance and usability of OGD, as well as virtual agents and rapport. Governments have released large amounts of OGD for its potential economic, social and political benefits. However, there is inadequate evidence for the real impacts of OGD on the whole of society, which may be due to the limited utilization on the demand-side of the OGD ecosystem. The conflicts lying on the supply-side and demand-side of OGD call for a study to find possible solutions to connect these two sides to improve citizens' utilization of OGD, which reflects the core research question of this thesis:

How to improve citizens' utilization of OGD by connecting the supply-side and demand-side through portals?

The review of related literature in this chapter also provided theoretical foundations as well as background information for later design of studies. To start with, the basic concepts of OGD together with its characteristics and impacts and OGD ecosystem provided the general research context of this thesis. The review of OGD development in China shows why it is necessary to conduct this research and possibility for carrying out multiple-case research in China. The discussion of the supply-side and demand-side of OGD as well as the bridge between these two sides supports the study of the evaluation on the supply-side in Chapter 3 and the investigation on the demand-side in Chapter 4 as well as the comparison of the supply-side and demand-side in Chapter 5. DOI and TAM introduced in this chapter provided the theoretical foundation for the research design in Chapter 5 and 7. The overview of the usability of OGD portals shows the necessity for the research of usability evaluation in Chapter 6. Relationships of the concepts discussed in this chapter with the studies carried out in later chapters are shown in the roadmap in Figure 2.3.

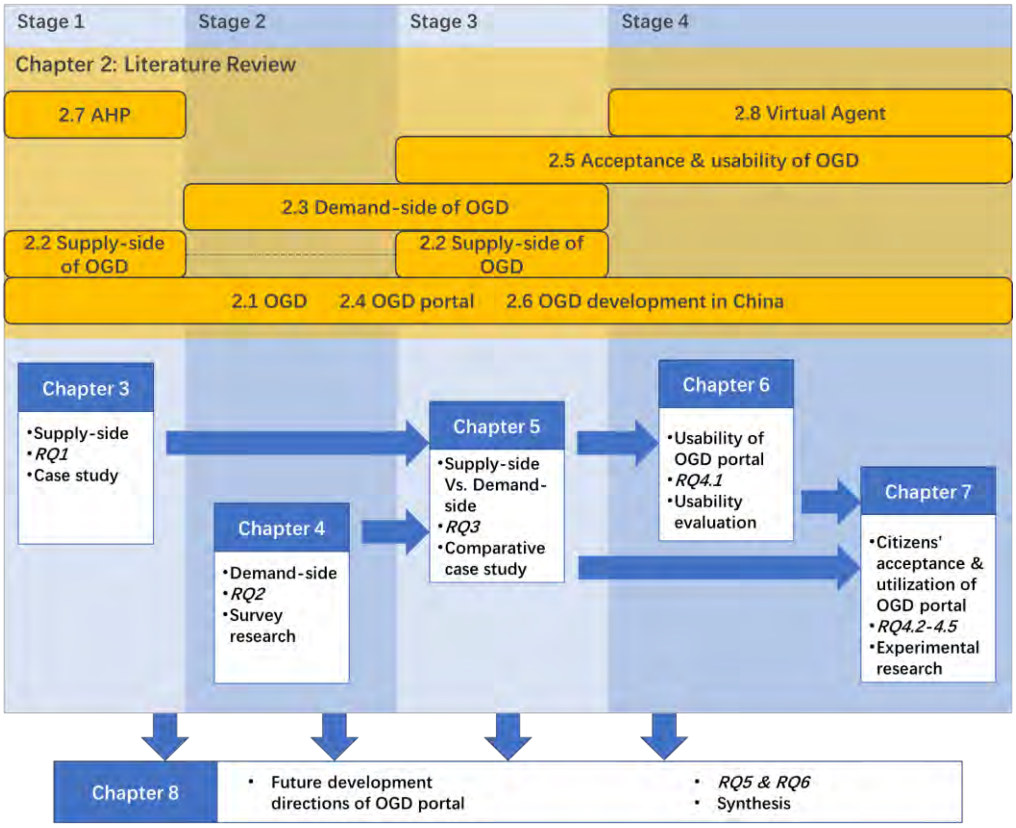
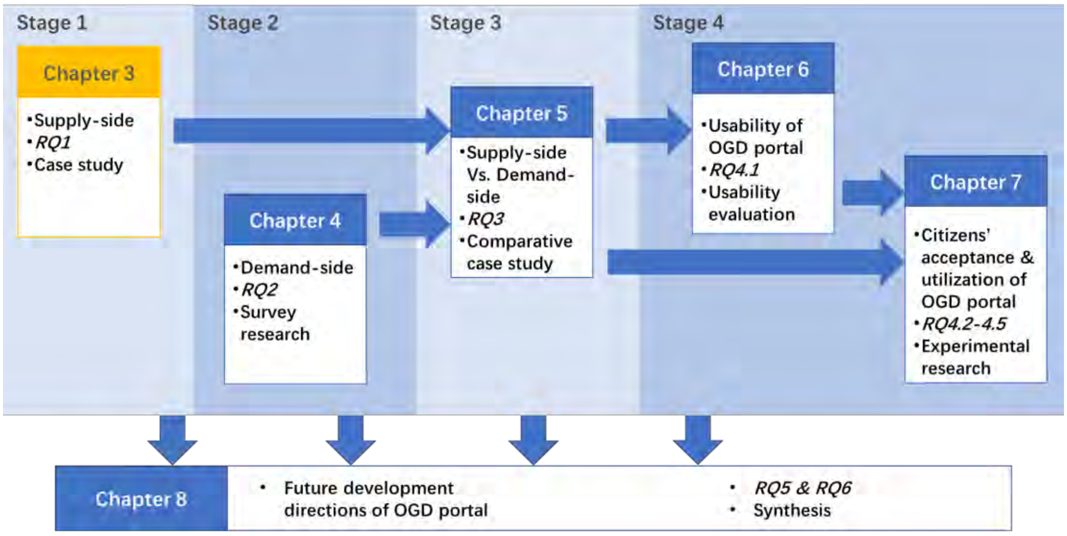


Figure 2-3 Roadmap of concepts in Chapter 2 and later studies

Through synthesizing the existing literature related to our study, several research gaps have been recognized. To start with, local-level OGD portals have not received enough attention for the issues in their development, which leads to RQ1.1, RQ1.3 and RQ1.4 about analyzing and guiding the development of local-level OGD portals. Secondly, few evaluations have been carried out of the development of OGD in China, and few studies focused on the demands and utilization of OGD portals by Chinese citizens. Therefore, knowledge of Chinese OGD development may not reflect the true situation. This not only shows support for carrying out case studies in China but also indicates the necessity for proposing RQ2.1 to RQ2.3 about citizens' OGD demands and utilizations. Thirdly, for the supply-side of OGD, present evaluation frameworks have several limitations like lacking practical application and clarity in the prioritization process, which leads to RQ1.2 about how to obtain priorities for elements of OGD evaluation framework. Also, evaluation of the supply-side has focused more on the data provided on the OGD portal but neglected the usability of the portal and its services, which indicates the necessity of RQ4.1 to RQ4.5 about the usability of OGD portals. Fourthly, for the demand-side, citizens, their demands and their interactions with OGD portals have been less considered in the literature, which justifies further the need of research for RQ2.1 to RQ2.3. A gap in comparing the supply-side and the demand-side of OGD is also recognized in the literature, reflecting RQ3.1 and RQ3.2. Finally, virtual agents can help users to build strong rapport with the information system through smoother interactions, which may consequently encourage users' engagement and intention to use the OGD portal. However, although virtual agents have been used in e-commerce websites, no studies have explored their effect on OGD users, which indicated the need for a study of RQ4.5.

To fulfill the research gaps identified above and to answer the research questions proposed in Chapter 1, we designed a four-stage study presented accordingly from Chapters 3 to 7, as is shown in Figure 2.3. Based on the results from each stage and the discussion presented in Chapter 8, we addressed the core research question of this thesis drawing from this literature review.

3 CHAPTER 3: AN EVALUATION OF THE SUPPLY-SIDE OF OPEN GOVERNMENT DATA IN CHINA[3]



3.1 Introduction

This chapter presents Stage 1 of the thesis to understand the supply-side (i.e. the development and delivery) of OGD in China. As presented in Section 2.2.1, several evaluations regarding country-level OGD portals have been proposed by different organizations and scholars (Attard et al., 2015; Lourenço, 2013; Thorsby et al., 2017) in order to understand the development of OGD. These evaluations provided evidence for the effectiveness of evaluating as a key method in understanding the supply-side of OGD. At the same time, system-oriented evaluations hold the strength in dealing with complex issues, as discussed in Section 1.5. Therefore, a system-oriented evaluation of existing local OGD portals can thus provide an interim step towards understanding the current development of the supply-side of OGD.

[3] This chapter closely follows the paper Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly, 35*(4), 644-656.

Assessments of the potential and capabilities of OGD portals at local levels, such as states, provinces or cities, are limited (Chatfield & Reddick, 2017; Conradie & Choenni, 2014; Kassen, 2013; Petychakis et al., 2014). A gap in the literature in relation to a systematic and empirical evaluation in China has also been recognized (see Chapter 2). Besides, although frameworks can be found in the literature, few seem to have been specifically built for local governments (Lourenço, 2015; Susha, Zuiderwijk, Janssen, & Grönlund, 2015; Thorsby et al., 2017). Although these frameworks all include different aspects as guidance for OGD portal development (Ubaldi, 2013), a key limitation with the present evaluation frameworks is the absence of a prioritization process (Attard et al., 2015; Lourenço, 2013) to guide decision-making and implementation. Considering the heavy investment in open data implementation (Bollettino, 2002), limited budgets and limited IT infrastructure and skills at the local level (Ebrahim & Irani, 2005), local governments are facing particularly tough challenges in providing OGD (Conradie & Choenni, 2014). Thus, an approach that provides prioritized guidance to local governments in the development of OGD portals is required for effective budget and resource allocation (J. Lee & Rao, 2009).

To address the knowledge gaps identified above, we therefore raise the following generic research question for Study 1:

RQ1. How to analyze and guide the development of local OGD portals?

This generic question can be further divided into four more specific sub-questions:

RQ1.1 What framework can be used to assess local-level OGD portals?

RQ1.2 How to obtain priorities for elements of the framework?

RQ1.3 How to use the framework to analyze the present development of local OGD portals?

RQ1.4 How to use the framework to guide the future development of local OGD portals?

To answer the first sub-question, we draw on prior OGD evaluation studies to derive an assessment framework for local-level OGD portal evaluation. Addressing RQ1.2, we then adopt the Analytic Hierarchy Process (AHP) to decide on the priorities of all elements in the framework. A case study in Chinese province-level OGD portals has been carried out to demonstrate the capability of the framework to answer RQ1.3 and RQ1.4.

The structure of the chapter is as follows: Section 3.2 explains the methods we used to address the four research questions. Section 3.3 presents our main results and findings, followed by discussion comparing findings with previous research in Section 3.4. This chapter ends with our conclusions in Section 3.5.

3.2 Methodology

In this section, we first describe the building of an evaluation framework based on related studies, then following the general process of AHP to derive priorities for each element in the framework. To test the capability of the prioritized framework, a case study in China has been carried out. Detailed processes for each step are explained as follows.

3.2.1 Build the evaluation framework

To form the basis of element selection for our evaluation framework, we first searched for related studies, then made selections based on explicit inclusion criteria. After that, we noted evaluation aspects in these studies and made comparisons to select the final elements to build our evaluation framework.

The following search terms together with their combinations were used to form search queries on Google and Google Scholar: government, open data, portal, platform, evaluation, assessment, rank. 261 publications were targeted after the search. A manual selection was performed on these results to guarantee that only relevant studies were used in further comparison. The inclusion criteria for this selection were: firstly, we included studies applying an evaluation framework in practical assessments at least once. Secondly, we included principles that are commonly accepted by governments and non-government organizations as guides for OGD development (Ubaldi, 2013). This resulted in 12 publications that covered 10 evaluation frameworks and 3 principles.

Table 3.1 shows a comparison of the 10 evaluation frameworks. Although the Open Data Census and Open Data Index are both assessments proposed by the Open Knowledge Foundation, we included them separately because they were designed to be used at different levels (local and national). "Prioritize" and "Weight" are descriptions of the framework, indicating whether the framework includes a prioritization process and whether it weights the elements in the framework, respectively. "Object", "Level", "Scope" and "Data Collection" refer to the application and focus of the evaluation using that framework. If the evaluation had been carried out more than once, we recorded the latest details for these four columns relating to the application of the framework.

We can see from Table 3.1 that more than half of these frameworks consider prioritization of the evaluation element. In these prioritized frameworks, only the Open Knowledge Foundation, Open Data Portal Index and China Open Data Index (DMGLabFudanUniversity, 2018) set unequal weights for their elements, while the other two frameworks give equal weight to each element. However, although the frameworks including prioritization explain in their methodology how to set weights according to the Open Definition (OpenDataBarometer, 2016; OpenDataCensus, 2016; OpenDataIndex, 2017), they fail to provide the detailed process of how they reached the prioritization results. In relation to the practical application of these frameworks, there has been a greater focus on national-level portals than local ones. The scope covers cities, states and countries, with most focusing on countries. A range of data collection methods have been used but direct visits to the portals by humans (researchers or crowds) or machines (OpenDataMonitor, 2017) are the most common means of data collection. Evaluation aspects of these frameworks and principles are listed in Table 3.2.

27 assessment elements were identified from these frameworks and principles. The following three exclusion criteria were applied to select final elements for our evaluation framework. Firstly, the element was excluded if it had overlap with other elements, such as “No. of distribution”, “Quality” and “Granularity”. Secondly, the element was excluded if it had little relation to the definition of OGD and OGD portals, such as “No. of unique publishers”, “Voting”, “Ranking”, “Visualization”, “Suggest new datasets” and “Notification”. Thirdly, the element was excluded if it was hard to evaluate through visiting or searching the portal, such as “Primary”, which means the data was collected at the source with the highest granularity (Craveiro et al., 2013; OpenGovernmentWorkingGroup, 2007; SunlightFoundation, 2010). It is hard to prove that data on a portal has never been modified from its original form. “Permanence” is also hard to judge through limited visits to a portal. It is not possible for “Total distribution size” to objectively reflect how much data is disclosed on a portal, because data on portals is stored in different formats, and pictures and PDFs are much larger than CSVs. Elements meeting any one of these three criteria were excluded from the final evaluation framework, which is displayed in Table 3.3. The resulting final framework is presented in Section 3.1.

Table 3-1 Comparison of OGD portal evaluation frameworks

Evaluation Framework	*Producer*	*Prioritize*	*Weight*	*Object*	*Level*	*Scope*	*Data Collection*	*Reference*
China Open Data Index	DMG Lab Fudan University	Yes	Unequal	Portals	Local	19 cities and provinces	Observation	China Open Data Index, (2018)
Data Openness Indicator	Bogdanović-Dinić et al.	Yes	Equal	Portals	National	7 countries	Web Tool	Bogdanović-Dinić et al., (2014)
Open Data Barometer	World Wide Web Foundation	Yes	Equal	Datasets Government	National	115 countries	Expert survey Government survey Secondary data	OpenDataBarometer, (2016)
Open Data Census	Open Knowledge Foundation	Yes	Unequal	Datasets	Local	423 cities of 8 countries	Crowdsource	OpenDataCensus
Open Data Index	Open Knowledge Foundation	Yes	Unequal	Datasets	National	94 countries	Crowdsource	OpenDataIndex, (2017)
Open Data Monitor	European Union	No	-	Portals	National	32 countries	Data Harvesters	OpenDataMonitor
Open Data Portal Index	Thorsby et al.	Yes	Unequal	Portals	Local	36 cities	Visit portals	Thorsby et al., (2017)
Open Data Portal Requirement	Lourenço	No	-	Portals	National	7 countries	Visit portals	Lourenço, (2015)
Open Government Budgetary Data in Brazil	Craveiro et al.	No	-	Portals	Local	26 States	Search Machines	Craveiro et al, (2013)
OUR Data Index	OECD	No	-	Portals	National	35 countries	Expert survey	OECD, (2015)

Table 3-2 Evaluation aspects of frameworks and principles

		Evaluation frameworks									*Principles*			
		China Open Data Index	**Data Openness Indicator**	**Open Data Barometer**	**Open Data Census Open Data Index**	**Open Data Monitor**	**Open Data Portal Index**	**Open Data Portal Requirement**	**Open Government Budgetary Data in Brazil**	**OUR Data Index**	**Eight Open Government Data Principles**	**UK's Public Data Principles**	**10 Principles for Opening Up Government Information**	
	Assess Elements	*China Open Data Index(2018)*	*Bogdanović-Dinić et al. (2014)*	*OpenDataBarometer(2016)*	*OpenDataCensus OpenDataIndex(2017)*	*OpenDataMonitor*	*Thorsby et al.(2017)*	*Lourenço(2015)*	*Craveiro et al(2013)*	*OECD (2015)*	*OpenGovernment WorkingGroup (2007)*	*Ubaldi (2013)*	*SunlightFoundation (2010)*	***SUM***
Quantitative	**No. of categories**					1	1							2
	No. of datasets	1				1	1							3
	No. of distributions					1								1
	Total distribution size	1				1								2
	No. of unique publishers					1								1
	Portal visibility					1								1
Qualitative	**Complete**	1	1	1		1		1	1	1	1	1	1	10
	Data exist			1										1
	Digital format	1	1			1			1		1	1	1	7
	Easy access		1		1				1			1	1	5
	Free of charge			1	1							1	1	4
	Granularity							1						1
	In bulk			1	1							1	1	4
	Machine-readable	1	1	1	1	1	1		1	1	1	1	1	11
	Notification									1				1
	Online access			1	1		1	1	1		1			6
	Open access	1	1		1	1		1			1	1		7
	Open license	1	1	1	1	1			1		1	1	1	9
	Permanence												1	1
	Primary		1						1		1		1	4
	Quality	1						1						2
	Ranking	1								1				2
	Suggest new datasets	1					1	1						3
	Up to date	1	1	1	1			1	1		1	1	1	9
	Update regularly	1	1	1	1			1	1		1	1	1	9
	Visualization	1					1							2
	Voting									1				1

3.2.2 Derive priorities for evaluation elements

We chose to use AHP to derive priorities for each evaluation element. More information about AHP could be found in Section 2.7. According to the general AHP procedure produced by Saaty (2008), we first transferred the evaluation framework into a hierarchical structure with a goal, several layers of criteria and a layer of alternatives. Secondly, an expert questionnaire was created for pairwise comparison of the importance of elements belonging to each of the upper-level elements in the hierarchy. Since we would apply the framework to a case study in China, we used the published literature to identify 10 experts from the field of information and library science working on topics of OGD in China and asked them via email to fill in the questionnaire separately. They ranked the importance of each pair of elements from 1 to 9.

Table 3-3 Exclusion of elements

		Exclusion criterion 1	Exclusion criterion 2	Exclusion criterion 3
Quantitative	No. of category			
	No. of datasets			
	No. of distributions	X		X
	Total distribution size			X
	No. of unique publishers		X	
	Portal visibility			
Qualitative	Complete			
	Data exist			
	Digital format			
	Easy access			
	Free of charge			
	Granularity	X		
	In bulk			
	Machine-readable			
	Notification		X	
	Online access			
	Open access			
	Open license			
	Permanence			X
	Primary			X
	Quality	X		
	Ranking		X	
	Suggest new datasets		X	
	Up to date			
	Update regularly			
	Visualization		X	
	Voting		X	

Exclusion criterion 1: The element had overlaps with other elements
Exclusion criterion 2: The element had little relation to the definition of OGD and OGD portal
Exclusion criterion 3: The element was hard to evaluate through visiting or searching the portal

We examined the consistency of each of the comparison matrices from the expert questionnaires and abandoned those matrices with consistency ratios (CR) over 0.10. Then we calculated the weighted geometric mean of values in each matrix given by the different experts to form the final matrices for calculating the priorities. After that, we derived the local scale of priorities by solving for the principal eigenvector of the matrix and then normalizing the result. We continued this process of weighting until the final priorities of the alternatives in the bottom-most level were obtained. After weighting by the priority of its parent criterion, we finally obtained a global derived scale for each variable in the model. Yaahp 10.5[4] was used to make the calculations.

3.2.3 Case study: evaluation of local OGD portals in China

To test the capability of the evaluation framework, we carried out a case study evaluating local OGD portals in China. A case study is "a research strategy which focuses on understanding the dynamics present within single settings" (Eisenhardt, 1989, p. 534). Thus, it fits to our aims of both testing the framework and describing the present development of local OGD portals in China. We selected China for three reasons. Firstly, no up-to-date, systematic and practical evaluations of Chinese local OGD portals could be found in the literature. Secondly, the development of OGD portals in China commenced at the local level without one united national portal. Thus, evaluation objects did exist in China. Finally, China is the largest developing country, with provinces of different contexts including population, size and wealth. Thus, a case study in China could help test the capability of the framework in areas with different contexts and whether the framework could be widely applied.

We followed the general procedure offered by Lourenço (2013) to carry out the case study. Excluding the steps for building the evaluation framework, there were five further steps.

---

[4] See the introduction to this software at: http://ahpman.metadecsn.com/

Step 1: Select local OGD portals in China as the type of entity to be evaluated

We targeted the official web portals launched by local governments as the type of entity. Thus, firstly, the portal had a structured domain with ".gov". Secondly, the portal had to contain data from more than one department or institution managed by a local government to serve as a "one-stop shop".

Step 2: Identify all province-level OGD portals in China as a subset of entities

We chose province-level OGD portals as evaluation objects for two reasons: Firstly, 34 province-level areas (including 27 provinces and autonomous region, 4 municipalities and 3 special autonomous regions) in China were available for an exhaustive investigation. Secondly, we chose province level instead of city level because larger local governments are more likely to be capable of providing the resources and technology required for the planning, launch and operation of dedicated OGD portals (Chatfield & Reddick, 2017) and are usually "the most innovative in the adoption of new technologies with more need of greater disclosure and lower relative costs for implementing new tools" (Bonsón, Torres, Royo, & Flores, 2012, p. 126). We used Baidu, the biggest and most popular search engine in China, and a literature search to find whether OGD portals had been built in the 34 province-level areas.

Previous studies have shown a relationship between the contextual features of an area and the number of datasets disclosed by local governments. Thorsby et al. (2017) discovered the population size of an area has a positive impact on the number of datasets. Yu (2010) found the size of a local government (equal to the natural logarithm of GDP) and the wealth of a local government (equal to the natural logarithm of general budget revenue) were determinants of the disclosure of local financial data. Thus, we collected population, regional GDP and general budget revenue from each province to support further analysis.

Step 3: Determine the categories expected to be disclosed by each of the entities considered

We referred to the selected frameworks described in Section 3.1 (see Table 3.1) to form an initial list of 36 data categories. We also recorded the categories used by the province-level OGD portals in China, which increased the list to 46 categories. During this process, we noted several problems. Firstly, there were overlaps among these categories, such as statistics with demographics, cultural resources with libraries, locations with maps, environment with air and water quality. Secondly, some categories were not commonly used by other studies or portals, such as contracting, energy, parking, permits and shops. Thirdly, some categories were not appropriate for China, such as 311 data, elections and voting. Therefore, to reduce the overlaps among categories and to ensure applicability in China, we finally selected 15 categories that were commonly used by related studies and portals. Descriptions of the scope for each category are given in Table 3.4.

Table 3-4 Categories for data collection

Category	*Description*
Budget & spend	The budget and spend records of the local government.
Credit records	Credit records of companies and organizations recorded by the local government.
Cultural activity	Data related to cultural and leisure activities and services.
Education	Data related to education system and services.
Environment quality	Data related to environment, including air and water quality, forestation rate, etc.
Government bid	Procurement data of the local government.
Health	Data related to the healthcare system and health services.
International trade	Details of the import and export of specific commodities and/or balance of trade data.
Local statistics	Key statistics in this area, such as demographic and economic indicators (GDP, unemployment, population, etc).
Map	Detailed digital map of this area.
Policies & legislation	Policies and legislations proposed in this area.
Public Safety	Data related to public safety, including crime statistics, food safety, daily products safety.
Registration	Data of companies and organizations registration, including name, unique identifier and other additional information.
Transportation	Details of public transport services, including the construction of public transports.
Weather	Data related to weather and climate records and forecasts.

Step 4: Visit the portals and collect the data required by the evaluation framework

We visited all the province-level OGD portals in China from 13 November 2017 to 17 November 2017 to collect the required data. The data was collected at a moment in time. Thus, the results are based on a snapshot of the available records.

Quantitative elements of the evaluation framework were calculated by visiting each portal. For "Portal visibility" we used as our search engine Yahoo! Japan, which has been used by scholars to get external inlinks (links coming from websites outside the site in question, reflecting the online visibility of a website) (Gao & Vaughan, 2005; Qiu, 2010; K. Yi & Jin, 2008). Search queries "link: URL" and "site: URL" were used to get total inlinks and internal links, whose difference equals the external inlinks.

For qualitative elements of the evaluation framework, we randomly selected 60 datasets for each category shown in Table 3.4 as samples for evaluation. The sample selection for each category was based on keyword searches on a portal. If the search results did not have enough datasets (< 60), we included all datasets with content fitting the description of that category. Each sample dataset was evaluated in the order A4 to A15 by judging whether the statement in the checklist column (Table 3.5) was true or not. If the statement was true, we gave the element a value of 100, otherwise we gave it a value of 0.

Step 5: Compute the assessment result and carry out the analysis

For quantitative elements, to ensure their values were comparable with the qualitative elements, they were normalized prior to aggregation. We treated the largest value in a quantitative variable as 100 and other values of this variable were scaled down accordingly. For the qualitative elements, the calculation was first done by category. We calculated the average score of 60 sample datasets of a category on one assessment element. After scores for all categories of a portal were calculated, we then derived the average score of this assessment element in 15 categories as the final score of this element for this portal. Finally, we added up all quantitative and qualitative elements by their priorities and obtained the assessment result for each portal.

Data analyses were carried out from three different perspectives. Firstly, we made comparison of the result by portal, and secondly by the different categories. Averages and standard deviations were used for these comparisons. Finally, we used correlation analysis to test the relationship between the number of datasets on a portal and the context of an area. The results of the whole process are given in Section 3.3.

3.3 Results and findings

The results and findings of our study are presented in the same order as in Section 3.2. We first present the assessment framework we derived in Section 3.1, then show the results of each AHP step in Section 3.2. Finally, all the detailed results of the case study are given in Section 3.3.

3.3.1 Local OGD portal evaluation framework

Table 3.5 shows the final framework with each element, its description and assessment checklists. To use AHP to derive priorities, a framework must include at least two layers. Thus, we classified the qualitative elements into two groups: one group relates to data accessibility, which is the basic or primary characteristic of open data and thus is the premise of data utilization, since citizens cannot use data without getting access to it first. The other group, data quality, includes other quality attributes including features that make data easier to use and improve data fitness for utilization by data consumers (Attard et al., 2015; Dawes, 2010; R. Y. Wang & Strong, 1996).

3.3.2 Priorities of the evaluation framework

As explained in Section 3.2, we first transferred the framework into a hierarchical structure in AHP with a goal, several layers of criteria and a layer of alternatives. We treated local OGD portal evaluation as the goal, the groups in the framework as the layers of criteria and the elements as the layer of alternatives. The AHP hierarchical structure together with the priorities derived from the AHP process are shown in Figure 3.1.

Table 3-5 Local OGD portal evaluation framework

Group	*Element*	*Description*	*Assessment Checklist*	*Reference*
Data Quantity (C1)	Portal Visibility (A1)	The number of external inlinks of a portal	The no. of total inlinks minus the no. of internal inlinks within the portal itself	Gao & Vaughan, (2005); Qiu, (2010); Yi & Jin, (2008)
	Datasets No. (A2)	The number of datasets on a portal	The number of total datasets on a portal	OpenDataMonitor, (2017); Thorsby et al. (2017)
	Categories No. (A3)	The number of dataset categories on a portal	The number of dataset categories on a portal	OpenDataMonitor, (2017); Thorsby et al. (2017)
Data Accessibility (C2)	Data Exist (A4)	The existence of information/data/dataset of a certain category on a portal.	Data/datasets of a certain category or their information can be searched or browsed on the portal.	OpenDataBarometer, (2016)
	Open Access (A5)	The extent to which data is available or retrievable freely by anyone.	Data/datasets can be accessed through possible ways (Could be ways other than directly from the portal, like off-line application, links to other sites).	the Open Definition; OpenGovernmentWorkingGroup, (2007)
	Online Access (A6)	Offer one-stop-shop online access for end-users to get the data/dataset.	Data/datasets can be accessed online directly through the portal (including after registration).	Attard et al., (2015); Lourenço, (2015)
	Easy Access (A7)	The ease with which information can be obtained through physical or electronic means.	Data/datasets are available with no requirement of registration or any other required procedures.	Sunlight Foundation, (2010); Open Data Index, (2017); OpenGovernmentWorkingGroup, (2007)
	In Bulk (A8)	Possibility to download all of the data stored in a database at once.	Data/datasets on the portal can be downloaded in bulk.	Sunlight Foundation, (2010); Open Data Index, (2017)
	Open License (A9)	Anyone is legally allowed to use, modify and redistribute data for any purpose. Reasonable privacy, security and privilege restrictions may be allowed.	License information that the data/dataset is provided under is stated explicitly on the page for the data/dataset.	Open Data Index, (2017); OpenGovernmentWorkingGroup, (2007); Sunlight Foundation, (2010)
Data Quality (C3)	Digital Format (A10)	The data/dataset is provided in electronic form.	Data/datasets on the portal are in digital format including web pages, PDF files, or scanned copies of paper documents, etc.	OpenDataBarometer, (2016)
	Machine-readable (A11)	Data/dataset is stored in widely-used file formats that easily lend themselves to machine processing, accessing and modifying single elements in the file.	Data/datasets can be directly opened in appropriate data manipulation and analysis software.	OpenGovernmentWorkingGroup, (2007); Sunlight Foundation, (2010); Open Data Index, (2017)
	Complete Metadata (A12)	Data/dataset has metadata that defines and explains what is provided about a particular subject.	Metadata that defines and explains the data/datasets is provided on the portal.	Sunlight Foundation, (2010); Bogdanović-Dinić et al., (2014); OpenDataMonitor, (2017)
	Free of Charge (A13)	Don't have to pay for the data/dataset.	Data/datasets are free of charge for downloading.	Open Data Index, (2017); Public Data Principles, (2012); OpenDataBarometer, (2016)
	Up to Date (A14)	Data/dataset is made available as quickly as necessary to preserve the value of the data.	If the data is updated regularly, it has the latest version. If not updated regularly, the latest version is updated in a week.	OpenDataBarometer, (2016); OpenGovernmentWorkingGroup, (2007); Lourenço, (2015)
	Update Regularly (A15)	To maintain the sustainable utilization of a data/dataset, the data/dataset should be kept regularly updated.	Data/datasets have an update plan.	Sunlight Foundation, (2010); OpenDataBarometer, (2016); Craveiro et al. (2013)

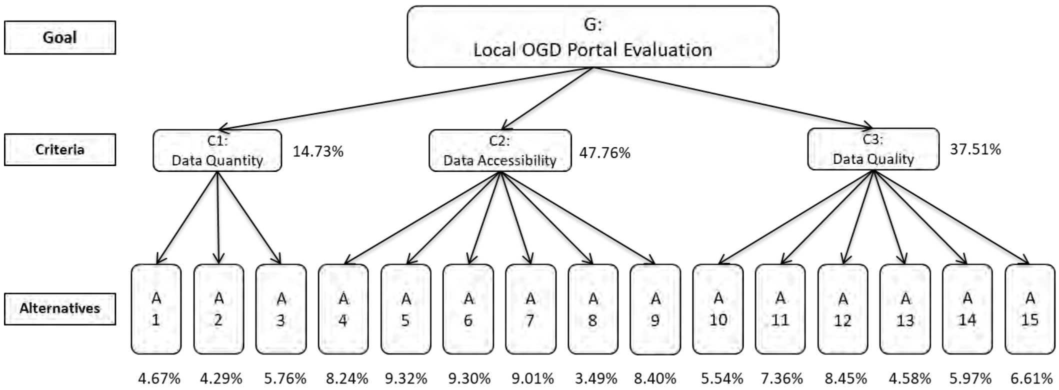


Figure 3-1 AHP hierarchy structure with priorities (see Table 3.5 for A1–A15 descriptions)

3.3.3 Case study in China

Out of 34 province-level jurisdictions in China, we found 9 OGD portals. 2636 datasets from these portals have been selected and evaluated to obtain the results. These portals, together with the context data (until the end of 2016) of these provinces, are listed in Table 3.6.

Table 3-6 Province-level OGD portals in China

Place	Population (Ths)	Regional GDP in RMB (Mio)	Size	General budget revenue in RMB (Mio)	Wealth	OGD Portal URL
Beijing	21,730	25,669.13	28.57	5,081.26	26.95	http://www.bjdata.gov.cn
Guangdong	109,990	80,854.91	29.72	10,390.35	27.67	http://www.gddata.gov.cn
Guizhou	35,550	11,776.73	27.79	1,561.34	25.77	http://www.gzdata.gov.cn
Hong Kong	7,377	5,469.96	27.03	3,503.72	26.58	https://data.gov.hk
Shanghai	24,200	28,178.65	28.67	6,406.13	27.19	http://www.datashanghai.gov.cn
Sichuan	82,620	32,934.54	28.82	3,388.85	26.55	http://www.sc.gov.cn/10462/13797/index.shtml
Taiwan	23,540	33,328.55	28.83	5,780.29	27.08	http://data.gov.tw
Xinjiang	23,980	9,649.70	27.60	1,298.95	25.59	http://data.xinjiang.gov.cn
Zhejiang	55,900	47,251.36	29.18	5,301.98	27.00	http://data.zjzwfw.gov.cn

3.3.3.1 Comparison by portal

The evaluation results are shown in Table 3.7. Taiwan received the top score, followed by Hong Kong. The rankings of portals for data access and data quality were almost the same (Figure 3.2) except for Guizhou, where data quality got a higher score than data access. Differences between data accessibility and quality were large in Zhejiang, Guangdong and Taiwan, but small in Beijing, Guizhou and Hong Kong. By calculating the standard deviation, we noted that differences in “Open licence”, “Easy access” and “In bulk” were significant. The largest difference lay in “Open licence”. “Data existence”, “Open access”, “Online access”, “Digital format”, “Complete metadata” and “Free of charge” were elements showing both high average score and low standard deviation, which indicates that all portals performed well in these aspects.

In relation to data quantity, we saw a great gap between Taiwan, Hong Kong and the others, especially in portal visibility and number of datasets. But we also observed that some portals, like the ones in Guizhou, Sichuan and Zhejiang, might have blocked search engines from collecting their data. Since we relied on the search engine to get portal visibility, this blocking might have led to a relatively low score.

In relation to data accessibility, Taiwan got a much higher score than the others. “Open licence” got the lowest score in all elements. More than half of these portals did not list the licence information for each dataset. Beijing, Shanghai and Guizhou forced users to register before downloading. Sichuan and Xinjiang offered data in PDF or online tables, so users could not download data or datasets in bulk. Sichuan’s portal also offered links to other government websites instead of datasets for downloading in many records, so users needed to search again for the data they wanted on other websites.

In relation to data quality, the OGD portals did well in offering data in digital format with no charge. But there was a wide disparity among portals in offering machine-readable formats. We counted the formats used by each of the portals to obtain the results shown in Figure 3.3. XLS and CSV were the two most commonly used formats. Taiwan and Hong Kong also used XML and JSON. Xinjiang was the only one that did not offer any data in machine-readable format. Datasets in Hong Kong were quite new and up to date. Guizhou updated most of its datasets regularly. However, we noted a gap between “Update regularly” and “Up to date” in Zhejiang, Guangdong and Guizhou, which indicates that although they had a schedule for updating, they did not follow the schedule to provide the latest data. Sichuan and Xinjiang received low scores for offering digital and machine-readable format data, due to offering scanned copies or online tables that could not be downloaded.

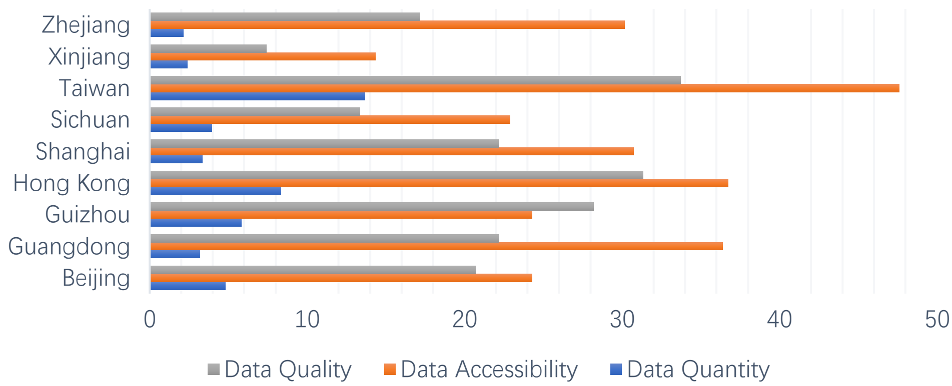


Figure 3-2 Comparison of evaluation scores by portal

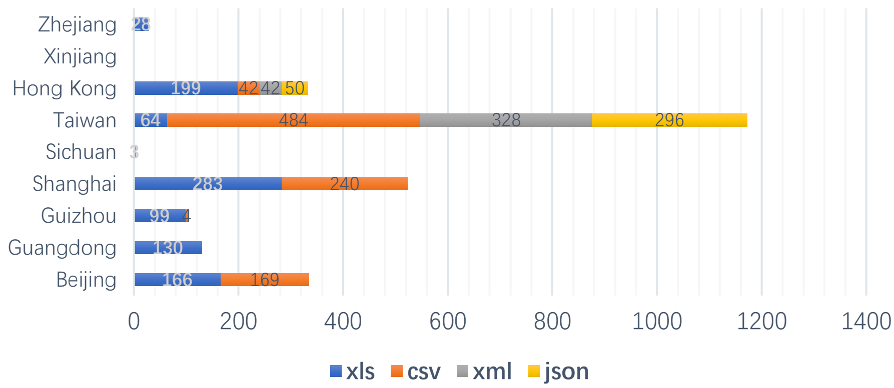


Figure 3-3 Number of different format datasets provided by portals

Table 3-7 Evaluation results

Group	*Element*	*Priority*	Beijing	Guangdong	Guizhou	Hong Kong	Shanghai	Sichuan	Taiwan	Xinjiang	Zhejiang	Avg.	Std.
Data quantity	Portal visibility	4.67%	0.07	0.46	0.01	57.62	0.29	0.00	100.00	3.50	0.00	18.00	36.09
	Datasets no.	4.29%	2.21	0.67	1.76	21.56	4.40	0.35	100.00	3.31	0.86	15.01	32.55
	Categories no.	5.76%	81.82	54.55	100.00	81.82	54.55	68.18	81.82	36.36	36.36	66.16	22.16
	Score by priority		4.81	3.19	5.84	8.33	3.34	3.94	13.67	2.40	2.13	5.30	3.68
Data accessibility	Data exist	8.24%	80.00	80.00	80.00	93.33	86.67	73.33	100.00	40.00	93.33	80.74	17.46
	Open access	9.32%	80.00	80.00	80.00	93.33	77.67	60.92	100.00	40.00	79.67	76.84	17.53
	Online access	9.30%	80.00	80.00	80.00	93.33	86.67	60.92	99.69	40.00	81.00	77.96	17.81
	Easy access	9.01%	0.00	67.72	0.00	93.33	0.00	60.92	99.69	40.00	79.56	49.02	40.69
	In bulk	3.49%	80.00	59.76	80.00	93.33	71.63	0.00	97.38	0.00	9.59	54.63	40.20
	Open licence	8.40%	0.00	80.00	0.00	0.00	68.79	0.00	100.00	0.00	0.00	27.64	42.21
	Score by priority		24.28	36.40	24.28	36.74	30.72	22.87	47.61	14.35	30.15	29.71	9.73
Data quality	Digital format	5.54%	80.00	80.00	80.00	93.33	86.03	60.92	99.69	40.00	83.08	78.12	17.83
	Machine-readable	7.36%	79.22	59.76	80.00	89.06	74.18	1.05	96.29	0.00	9.03	54.29	39.55
	Complete metadata	8.45%	80.00	80.00	80.00	93.33	86.67	54.25	99.69	40.00	80.56	77.17	18.71
	Free of charge	4.58%	80.00	80.00	80.00	93.33	86.67	60.92	100.00	40.00	79.67	77.84	17.83
	Up to date	4.97%	0.44	9.34	43.59	91.14	6.42	24.92	77.03	0.00	8.59	29.05	34.22
	Update regularly	6.61%	0.44	37.38	80.00	43.95	4.76	19.58	64.73	0.00	15.40	29.58	28.91
	Score by priority		20.74	22.19	28.20	31.32	22.15	13.36	33.72	7.43	17.17	21.81	8.47
Total		100.00%	49.83	61.78	58.31	76.39	56.22	40.18	95.01	24.18	49.45	56.81	20.35

3.3.3.2 Comparison by category

Evaluation results for data accessibility, data quality and number of datasets by category are shown in Tables 3.8 and 3.9 in order of the score by weight. Generally, scores for data accessibility were higher than for data quality, except for the category “Weather”. The rankings of categories in their performance concerning data accessibility and data quality were almost the same. The greatest gap between data accessibility and data quality lay in “Map”, while the smallest lay in “Weather”.

In relation to the number of datasets, we noted a large difference among categories (Std=79.27), with “Local statistics”, “Education” and “Health” in the top three, while “Weather”, “Map” and “International trade” were in the bottom three.

In relation to data accessibility, “Credit records” performed the best, while “Health” performed the worst. Referring to evaluation elements, “Data existence” got the highest score, while “Open licence” received the lowest score. By calculating the standard deviation, we found large variations in the scores for “Data existence”, “Online access” and “Open access”, while these variations were small for “Open licence”. This indicates that all categories did not perform well in open licencing.

In relation to data quality, “Local statistics” and “Credit records” performed quite well, while “Environment quality” performed the worst. Referring to evaluation elements, “Digital format”, “Complete metadata” and “Free of charge” received quite high scores, while “Up to date” and “Update regularly” got the lowest scores. By calculating the standard deviation, we found a large spread of scores between categories from 8.16 to 14.2. Less variation in scores was found between categories for “Up to date” and “Update regularly”, while larger standard deviations were found for “Complete metadata” and “Free of charge”. This indicates all categories failed to offer the latest data or update datasets regularly.

Table 3-8 Data accessibility by category

Data Accessibility Data Categories	Data Exist 17.25%	Open Access 19.51%	Online Access 19.47%	Easy Access 18.87%	In Bulk 7.31%	Open Licence 17.59%	Datasets No.	Score by Weight 100.00%
Credit Records	100.00	100.00	100.00	65.08	66.01	33.33	104	79.20
Local Statistics	100.00	100.00	100.00	62.78	62.09	33.33	299	78.48
Budget & Spend	100.00	87.78	91.11	55.93	65.56	33.33	136	73.33
Cultural Activity	88.89	88.89	88.89	55.56	74.71	33.33	249	71.79
Government Bid	88.89	77.78	88.89	53.97	64.10	22.22	134	66.60
Weather	77.78	77.78	77.78	55.56	54.07	33.33	81	64.04
Education	77.78	76.53	77.08	54.24	55.68	32.22	283	63.33
Map	88.89	72.84	72.84	39.51	64.18	33.33	85	61.74
Policies & Legislation	77.78	73.97	73.97	40.63	61.78	30.34	93	59.77
Registration	77.78	77.78	77.78	55.56	45.00	11.11	165	59.47
Public Safety	77.78	69.44	69.44	36.11	52.08	29.86	186	56.36
Transportation	66.67	66.67	66.67	44.44	44.44	22.22	226	53.03
Environment Quality	66.67	63.19	62.68	40.46	32.02	33.33	241	51.87
International Trade	66.67	64.44	66.67	31.11	44.44	22.22	86	50.08
Health	55.56	55.56	55.56	44.44	33.33	11.11	267	44.02
Avg.	**80.74**	**76.84**	**77.96**	**49.02**	**54.63**	**27.64**	**175.67**	**62.21**
Median	**77.78**	**75.25**	**75.53**	**49.21**	**54.88**	**31.28**	**165.00**	**60.76**
Std.	**13.59**	**12.86**	**13.33**	**10.14**	**12.57**	**7.99**	**79.27**	**10.38**

Table 3-9 Data quality by category

Data Quality Data Categories	Digital Format 14.77%	Machine-Readable 19.62%	Complete Metadata 22.53%	Free of Charge 12.21%	Up to date 13.25%	Update Regularly 17.62%	Datasets No.	Score by Weight 100.00%
Local Statistics	98.94	61.72	99.26	100.00	47.34	42.94	299	75.13
Credit Records	100.00	64.70	100.00	100.00	39.59	37.26	104	74.01
Cultural Activity	88.89	76.10	88.89	88.89	24.89	40.21	249	69.32
Budget & Spend	91.11	64.09	91.11	91.11	25.33	29.73	136	66.28
Government Bid	88.89	62.53	88.89	88.89	30.19	23.66	134	64.44
Education	77.08	54.57	77.08	77.08	35.42	27.06	283	58.33
Policies & Legislation	73.97	63.25	73.97	73.97	26.80	30.88	93	58.02
Weather	77.78	54.07	77.78	77.78	27.04	26.48	81	57.36
Map	72.84	63.07	72.84	72.84	17.84	31.91	85	56.42
Registration	77.78	45.00	77.78	77.78	32.22	26.52	165	56.28
Public Safety	69.44	54.86	69.44	69.44	19.07	33.89	186	53.64
International Trade	66.67	43.70	66.67	64.44	33.89	32.35	86	51.50
Transportation	66.67	44.44	55.56	66.67	33.33	33.33	226	49.51
Health	55.56	33.15	55.56	55.56	23.70	23.33	267	41.26
Environment Quality	66.15	29.09	62.68	63.19	19.12	4.18	241	40.58
Avg.	**78.12**	**54.29**	**77.17**	**77.84**	**29.05**	**29.58**	**175.67**	**58.14**
Median	**75.53**	**54.71**	**75.53**	**75.53**	**26.92**	**30.30**	**165.00**	**56.89**
Std.	**12.95**	**12.92**	**14.20**	**13.44**	**8.16**	**9.03**	**79.27**	**10.39**

3.3.3.3 Correlation analysis of contexts and number of datasets

We carried out Spearman's correlation analysis of whether the contexts of an area had a significant relationship with the number of datasets on this area's OGD portal. These results are shown in Table 3.10. Since the p values (sig. (2-tailed)) for the size and wealth of an area with the number of datasets are over 0.05, we found no statistically significant correlations between pairs. However, the p value for the local population and the number of datasets is 0.016 with Spearman's rho = −0.767, showing a significant strong negative correlation between these two. This result indicates that we could not predict the number of datasets on a portal of a local area by its size and wealth, but the local population might indicate the number of datasets on an OGD portal. Further discussion of this correlation is provided in Section 3.4.

Table 3-10 Correlation results for contexts and number of datasets

	Local Population	Size	Wealth
Spearman's rho	-0.767	-0.467	0.017
Sig. (2-tailed)	**0.016***	0.205	0.966

3.4 Discussion

In this section, we present comparisons of the above results with related studies from four perspectives, and then discuss how to plan the future development of OGD portals. To start with, we compare the priorities of our evaluation framework with the two frameworks having unequal priorities. Results derived from AHP show the greater importance of data accessibility and data quality compared with data quantity, which indicates that in the context of big data users should focus on high-quality data instead of a (very) large quantity of data to draw relatively more precise and high-value conclusions (Kaisler et al., 2013). We also note open access gets the highest priority (9.32%) in all elements, followed by online access (9.3%) and easy access (9.01%). These three are also a reflection of the "Open Data Definition" (Leong, Pan, Newell, & Cui, 2016) for data to be freely accessible. However, the Open Data Index sets the highest priorities for open licencing and machine-readability (OpenDataIndex, 2017), followed by other elements with equal priorities. Thorsby et al. (2017) also set a different priority for machine-readability which is higher than for the other elements. Although the priorities we derived for open licencing and machine-readability were also quite high, these two were not emphasized as in the frameworks set up by the Open Data Index and Thorsby et al. We think this is because we asked experts in China during the AHP process. The priority results may, thus, reflect their opinions of Chinese OGD portal development.

Secondly, regarding the evaluation results for province-level OGD portals in China, Taiwan and Hong Kong got much higher scores than the other portals in relation to various aspects. This result is consistent with the Open Data Index. In their evaluation, Taiwan ranked the top in all 94 jurisdictions around the world in 2016 (OpenDataIndex, 2016), getting full scores in 12 out of 15 OGD categories, while Hong Kong ranked 24th. The fact that only 9 out of 34 province-level areas in China have built local OGD portals also indicates that the implementation of OGD portals in China is still new.

Thirdly, regarding the evaluation results by category, there were several areas where present study's results were consistent or inconsistent with the Open Data Barometer's fourth evaluation (OpenDataBarometer, 2017a). In their evaluation, "National statistics" performed the best in data quality in all categories, while "Government spend", "Company registration" and "Map" were the worst. Our results show that although "Local statistics" ranked top, which is the same as the Open Data Barometer, "Environment quality" and "Health" performed the worst. Also, according to the Open Data Barometer "National statistics" was the only category offering machine-readability and free data. It was also the only category that did well in updating data regularly and keeping data up to date. In our evaluation, we found all categories offered data free of charge and offered machine-readable datasets to some extent. For data updating, it is true that all categories did not perform well and "Local statistics" was the best one, but "Cultural activity" received a similar score in updating datasets regularly.

Fourthly, regarding the relationship between local contexts and the number of datasets, we found a conflict in our study with previous ones. Thorsby et al. (2017) indicated that city population size was a critical factor in the number of datasets listed on the portal, which had a positive relationship with the number of datasets. However, the correlation analysis of our case study shows a significant negative relationship between the two, which is opposite to the conclusion of Thorsby et al. (2017). This indicates that, on one hand, further investigation is needed to determine whether population is truly a critical factor in the number of datasets in different parts of the world, since we based our analysis on data from China while Thorsby et al. based their analysis on data from the USA. On the other hand, areas in China with larger populations may have much more potential data or datasets that could be opened to the public. Our different results on the relationship between the size and wealth of local governments and the number of datasets with previous research (He Yu, 2010) could be related to the type of data (Yu's research only included government financial data, while ours included 15 different categories of data).

Finally, for the future development of province-level OGD portals in China, with the prioritized framework and evaluation results it will be quite easy for local governments to discover their present strengths and weaknesses and recognize points of attention for future development. Beijing, for example, has shown a conflict between the priorities viewed by experts and its present development in "Easy access" and "Open licence". Thus, in its future development Beijing should not compel users to register before downloading or viewing data. Also, the portal should add an explicit statement for each dataset about the licence it is provided under. Updating data on the portal regularly and keeping data up to date are other aspects recommended for improving portal quality. For Hong Kong, it is quite easy to see its weakness in open licencing. Thus, it could add an explicit open licence statement for each dataset in its future development. For Xinjiang, since the portal did not perform well in many aspects at present, the local government could first, according to the priority, focus on data accessibility to make its OGD available to the public, and then improve the data quality, especially by putting effort into providing machine-readable data and regularly updating datasets.

3.5 Conclusion

In this chapter, we discussed the building of a prioritized evaluation framework for local OGD portals and presented a case study in China to demonstrate its capability. The aim of our research mainly concerned the analysis and guidance of local OGD portal development. Four sub-questions were proposed to explore this aim, which were:

RQ1.1 What framework can be used to assess local-level OGD portals?

RQ1.2 How to obtain priorities for elements of the framework?

RQ1.3 How to use the framework to analyze the present development of local OGD portals?

RQ1.4 How to use the framework to guide the future development of local OGD portals?

To achieve these aims, we first drew on related evaluations and studies. An evaluation framework was built based on the comparison and synthesis of related studies, which gave the answer to RQ1.1. Analytic Hierarchy Process together with an expert survey was used to obtain priorities for each element in the proposed evaluation framework, which answered RQ1.2. The proposed prioritized evaluation framework was applied to a case study in China for answering RQ1.3. Based on the evaluation results in the case study, we discussed the arrangement of governmental resources by combining the weights of elements in the framework and the evaluation results, which answered RQ1.4. Differences in local and national OGD portal development were found. We identified several research gaps in local OGD portal evaluation frameworks and prioritization of evaluation elements. A dearth of studies of Chinese local OGD portals was also noted. Based on the results of our study, we drew the following conclusions:

- Data accessibility and the data quality of OGD portals draw more attention than data quantity.

- “Open access”, “Online access”, “Easy access”, “Complete metadata” and “Open licence” are the top five aspects emphasized by Chinese experts for the development of local OGD portals, and these reflect the open definition.

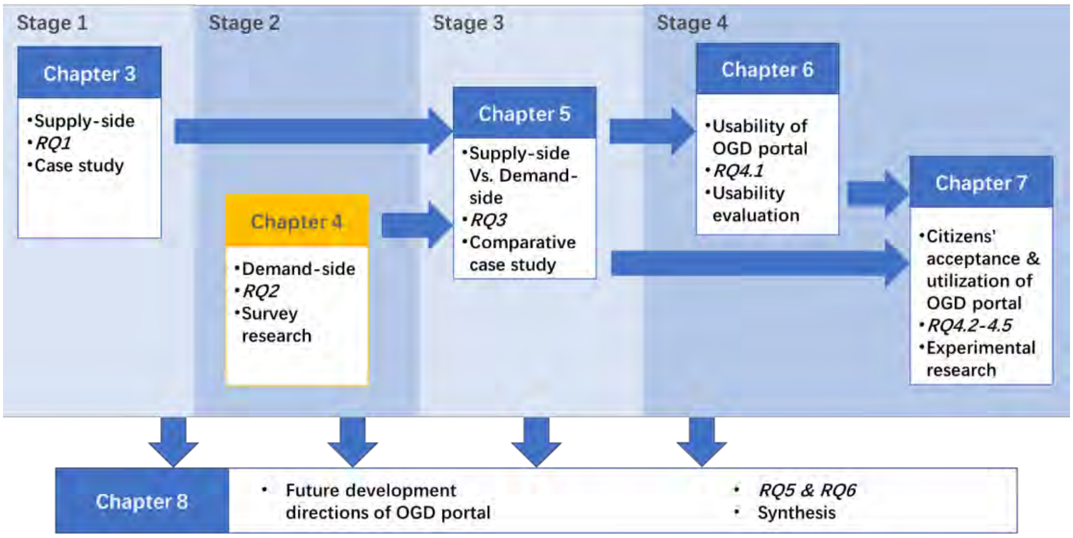
- There are currently imbalances in the development of province-level OGD portals in China, especially in the aspect of open licencing. Great gaps lie between Taiwan, Hong Kong and the other portals.

- “Local statistics”, “Credit records” and “Budget and spend” are data categories well released on province-level OGD portals in China.

- Population is negatively related to the number of datasets on Chinese province-level OGD portals, while the size and wealth of local governments are not currently significantly related to the number of datasets on Chinese province-level OGD portals.

Since the development of OGD portals is rapidly progressing and the case study in this chapter is a snapshot in time, new portals are constantly emerging and existing portals are changing. Thus, findings based on the data used in our case study are likely to change. Our main goal was not to evaluate current portals, but to demonstrate how our framework can be used for measurement so that it can be used to measure change and progress in the field by using the framework at regular timepoints. Similarly, new elements could be added as they emerge, new groupings could be defined or some elements of the framework could be grouped differently (such as moving “Free of charge” from the general quality grouping to the data accessibility group) if a different definition were adopted by evaluators.

4 CHAPTER 4: UNDERSTANDING CITIZENS' DEMANDS FOR OGD AND OGD UTILIZATION[1]



4.1 Introduction

In the previous chapter, we focused on the supply-side of OGD portals. As drawn from the related literature, releasing large amounts of OGD to the public through portals is not a guarantee of the achievement of OGD initiatives in the targeted aims (Attard et al., 2015; Heise & Naumann, 2012) to promote citizen engagement (Kassen, 2013) and provide government transparency and accountability (Attard et al., 2015).

Lack of use has been identified by scholars as a key problem in OGD development, (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017; Safarov et al., 2017; H.-J. Wang & Lo, 2016; Zuiderwijk & Janssen, 2013). Thus, to stimulate the

[1] This chapter closely follows the papers Wang, D., Richards, D., & Chen, C. (2019). *Connecting users, data and utilization: A demand-side analysis of open government data.* Paper presented at the International Conference on Information and Wang, D., Richards, D., & Chen, C. (2019). *Connecting users, data and utilization: A demand-side analysis of open government data.* Paper presented at the International Conference on Information.

utilization of OGD, many studies have been carried out to understand the decisive factors in OGD usage (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017; H.-J. Wang & Lo, 2016). However, reviews have shown that these studies examined this issue from the perspectives of either portals or users, but failed to consider the specific context of OGD utilization or the interaction between portals and users (A. Meijer et al., 2014). Moreover, "users of OGD are relatively less researched as subjects" in the present literature (Safarov et al., 2017, p. 16). Therefore, we recognized a gap in present studies in analyzing OGD users' data demands and acquisition process from the perspective of the demand-side, which puts the users of OGD portals in a central position.

Investigating and understanding users' demands and motivations could be a promising approach to enhance OGD utilization, since they are consumers of utilization effects (Safarov et al., 2017), including transparency (Florini, 2008; Willinsky, 2005), participation in policy (Wijnhoven, Ehrenhard, & Kuhn, 2015) and economic benefits (Willinsky, 2005). In addition, because one of the aims of disclosing government data on the web is for its use, reuse and distribution (Attard et al., 2015), further understanding of OGD users could promote meaningful citizen engagement (Wijnhoven et al., 2015) and improve the efficiency of utilization as well.

As described and justified below (see Section 4.2.1), we focus on demands from the perspective of users who are citizens. To better understand citizens' OGD usage habits and fill the identified knowledge gaps, we proposed the following core research question for study 2:

RQ2. How can citizens' demand for and utilization of OGD be characterized?

We further divided this research question into three more specific sub-questions, which are:

RQ2.1 What are citizens' demands for OGD?

RQ2.2 What are citizens' demands for and utilization of OGD portal services?

RQ2.3 What are the relationships between citizens, their OGD demands and OGD utilization?

To answer these questions, we carried out a survey study in China from the demand-side of OGD, focusing on the OGD demands and utilization of citizens with different characteristics. The objectives, which are also the main contributions of the study in this chapter, are twofold: firstly, to get a clear view of citizens' demands of OGD and OGD portal services; and secondly, to analyze the possible relationship between citizens, their OGD demands and their OGD utilization.

4.2 Research model and hypothesis

To form the theoretical foundation for our study, as well as to support the design of a survey instrument, we have drawn on related research carried out by both scholars and organizations related to OGD portals, OGD users and their demands and utilization.

4.2.1 OGD users

In the study of OGD, users are treated as the main actors in the open data process who can directly affect the coordination of OGD utilization from the demand-side (Zuiderwijk & Janssen, 2013). Thus many studies targeted their participation in the OGD value-extraction process (Safarov et al., 2017). In the scope of human–computer interaction (HCI), user-centred design is one of the key disciplines, focusing on making a system as easy and pleasant to use as possible (Dix, 2009). Therefore, in the second stage of study, we chose to focus on OGD users for understanding their demands and utilization of OGD.

It is commonly accepted by researchers to divide OGD users into different types (King & He, 2006), including citizens (Parycek et al., 2014; Power, Robinson, Rudd, & Reeson, 2015), business (Magalhaes et al., 2014; Susha, Grönlund, et al., 2015), researchers (Gonzalez-Zapata & Heeks, 2015; Whitmore, 2014), developers (Veeckman & van der Graaf, 2015) and journalists (Heise & Naumann, 2012). Among these, citizens are identified as the primary stakeholders who receive major benefits from the utilization of OGD (Parycek et al., 2014). Because the key motivation for releasing government data to the public is reducing the asymmetry of information between citizens and governments bodies (Murillo, 2015; Ubaldi, 2013), therefore, in this study, we chose citizens to represent OGD users, as well as to be our investigation objects for analyzing the hypotheses. Besides their vital importance as participants in OGD utilization as recognized by scholars (Zuiderwijk & Janssen, 2013), choosing citizens as the study objects offered the advantages of wider representation of the population and a range of diverse characteristics.

The demographic characteristics of users are usually treated as an important factor in understanding and predicting ICT adoption in the field of e-government (Dwivedi & Williams, 2008; Venkatesh, Sykes, & Venkatraman, 2014). Demographic characteristics of users, such as gender, age, experience and voluntariness of use were identified to influence user's acceptance of information technology (Venkatesh et al., 2003). Thus, in our research model, we included the demographic characteristics of citizens including their gender, age, education, occupation, and knowledge of OGD (which could indicate their experience with OGD and their voluntariness to use OGD) to analyze their impact on other aspects of OGD demand and utilization.

4.2.2 OGD demand

The source of OGD is the government, thus OGD portals usually provide OGD on certain subjects (Bogdanović-Dinić et al., 2014), including budget, health, public safety, legislation, etc. (Bogdanović-Dinić et al., 2014; OpenDataBarometer, 2017a). According to the Open Data Barometer's fourth global report, providing

OGD according to the demands of citizens could help restore or build citizens' trust of the government (OpenDataBarometer, 2017a). But citizens' need for different OGD categories is not addressed by the report. In order to explore whether citizens' demands for OGD have significant differences in different categories, we choose to include different OGD subjects as sub-elements of citizen's OGD demands in the research model.

In addition, OGD principles including the Eight Open Government Data Principles (OpenGovernmentWorkingGroup, 2007), Ten Principles for Opening Up Government Information (SunlightFoundation, 2010) and UK Public Data Principles (Ubaldi, 2013) have emphasized several attributes such as machine-readability, being up to date, updating regularly, etc. for the development of OGD. But citizens' attitudes towards these OGD quality principles are not clear. Therefore, in order to investigate citizens' attitudes towards the attributes emphasized by scholars and governments, we include OGD attributes as a sub-elements of citizen's OGD demands.

Researchers have noted that different OGD users show different levels of interest in OGD (Gonzalez-Zapata & Heeks, 2015), but the relationship between users and their preferences for data is unclear (Safarov et al., 2017). Considering various types of users in OGD utilization, we proposed the first hypothesis of our study:

H1: Citizens' characteristics have an effect on their OGD demands.

4.2.3 OGD utilization

The utilization of OGD refers to the exploitation of users for a particular purpose (Safarov et al., 2017). Therefore, it is reasonable to include utilization purpose as a sub-element of OGD utilization in the research model. According to users' separate aims, the usage of OGD has been divided into different types, including innovation (Magalhaes et al., 2014; Veeckman & van der Graaf, 2015), data analytics (Kalampokis et al., 2013; Kuhn, 2011), decision-making (Chakraborty et al., 2015; Desouza & Bhagwatwar, 2012), anti-corruption (Leontieva et al., 2015; Rajshree & Srivastava, 2012) and research (Kalampokis et al., 2013; Whitmore, 2014).

Different users may have different reasons and purposes for utilizing OGD, but they all need some access to OGD before they can make use of the data. Officially launched by governments, OGD portals are the main method for making governmental datasets publicly available to end-users (Kassen, 2013; Kostovski et al., 2012; Lourenço, 2013, 2015). Usually, OGD portals have different data-acquisition functions including searching, browsing, ranking, etc. for the end-users to find the data they want. In order to investigate citizens' attitudes and using habits towards these data-acquisition functions, we included them as a sub-element of OGD utilization in the research model.

Besides the data-acquisition functions, all websites including OGD portals are supposed to include some form of help system (Grayling, 2002) to support users' task completion (Z. Huang & Benyoucef, 2014). For OGD portals, user guides, FAQ, online virtual or smart agents and customer telephone lines are usually provided to help their end-users.

In the field of information retrieval, a user-stereotype including a set of demographic and social attributes is commonly used in information filtering to exclude the irrelevant information about a user (Kuflik & Shoval, 2000; Shapira, Shoval, & Hanani, 1997). Because users of different demographic backgrounds may judge the relevance of a certain data element differently, filtering rules based on the information needs of users will increase the accuracy of search results (Kuflik & Shoval, 2000). Due to the effect of users' demographic background, many websites provide adaptive search functions (Sugiyama, Hatano, & Yoshikawa, 2004). Since OGD portals are one-stop shops offering different data-acquisition functions for citizens to access the data they need, regarding comparing the effect of users' demographic background on their utilization of search engines and websites, we thus proposed the second hypothesis of this study:

H2: Citizens' characteristics have an effect on their OGD utilization.

In marketing, consumption refers to a broad set of practices including people's utilization of services and products (Harvey et al., 2001; Warde, 2005), which in OGD is the usage of the data. Economists assume that people's consumption reveals their demand (Harvey et al., 2001). Applying this relation of consumption and demand to the case of OGD, we proposed our third hypothesis:

H3: Citizens' OGD utilization influences their OGD demands.

Combining the concepts of OGD users, OGD demands and OGD utilization together with the hypotheses, we built the research model shown in Figure 4.1.

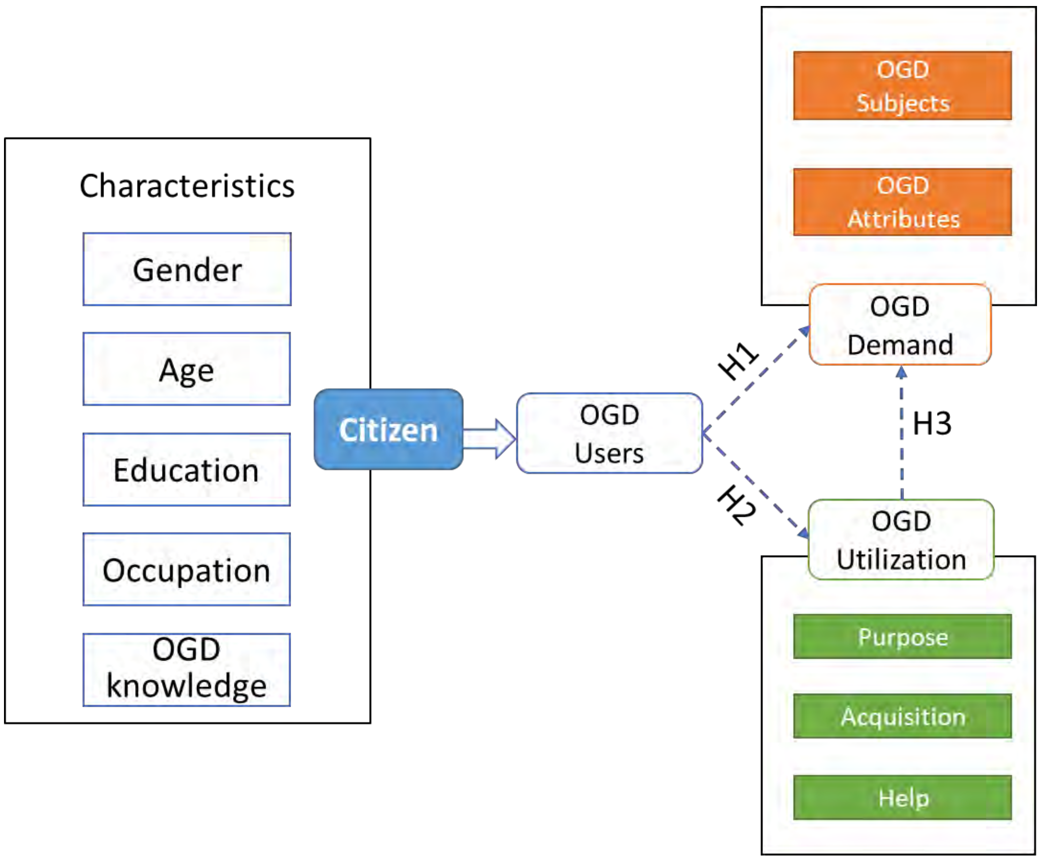


Figure 4-1 Research model

4.3 Methodology

To achieve the primary aim of this chapter, finding out the demands of citizens on the demand-side of OGD, an investigation was carried out on Chinese citizens. A survey was chosen as the appropriate instrument because surveys allow us to capture the preferences, attitudes and situations of the selected participants and generalize the relationships among users, their demands and utilization based on the data analysis (Recker, 2012). Details of the investigation are explained in the following sub-sections.

4.3.1 Survey design

A survey was designed and developed by operationalizing questions based on the research model shown in Section 4.2 that reflect the aims and hypotheses of this study.

The first part included the demographic questions. The first four questions were about the socio-demographic characteristics (gender, age, education and occupation) of the citizens. We also included four questions about citizens' knowledge of OGD, asking whether they had ever heard of OGD and OGD portals and whether they had ever used OGD and OGD portals.

The second part was about the OGD demands of citizens including their demands for different OGD subjects and their demands for OGD attributes. We asked citizens about their demands of OGD in general and their demands in relation to specific OGD subjects. For OGD subjects, we used the same 15 subjects that we selected for Study 1 in Chapter 3. Detailed explanations of each OGD subject can be found in Table 3.4. Before showing the participants the questions about OGD attributes, we first gave them an example of OGD portals with a screenshot of the Shanghai OGD portal, whose total score of performances as discussed in Chapter 3 ranked 5th of the 9 evaluated portals, indicating it to represent the average performance of OGD portals in China. For OGD attributes, we selected 6 out of the 12 elements from the framework in Chapter 3 Study 1. Detailed explanations of each data quality element are shown in Table 4.1. We excluded elements based on the following criteria: data exists, open access, online access and digital format. For "Data exists", according to its description it could be replaced by the questions that we asked about citizens' demands for OGD subjects. For open access, online access and digital format, these three attributes are essential conditions for building an OGD portal, no matter what the perspectives of the citizens. We moved easy access and open license to the next part of the survey. Thus, finally, "Download in bulk", "Machine-readable", "Complete metadata", "Free of charge", "Up to date" and "Update regularly" were selected for OGD attributes. A 7-point Likert scale from "Strongly don't need" to "Strongly need" was used for scoring the demands for OGD subjects and attributes.

Table 4-1 Descriptions of OGD attributes

Attributes	*Description*	*Reference*
Download in bulk	Possibility to download all the needed data stored in a database at once	Sunlight Foundation (2010); Open Data Index (2017)
Machine-readable	Data/dataset is stored in widely-used file formats that easily lend themselves to machine processing, accessing and modifying single elements in the file.	Open Government Working Group (2007); Sunlight Foundation (2010); Open Data Index (2017);
Complete metadata	Data/dataset has metadata that defines and explaines what is provided	Sunlight Foundation (2010); Bogdanović-Dinić et al. (2014); OpenDataMonitor (2017)
Free of charge	Don't need to pay for the data/dataset	Open Data Index (2017); Public Data Principles (2012); OpenDataBarometer (2016)
Up to date	Data/dataset is made available as quickly as necessary to preserve the value of the data.	OpenDataBarometer (2016); OpenGovernmentWorkingGroup (2007); Lourenço (2015)
Update regularly	To maintain the sustainable utilization of a data/dataset, the data/dataset is kept regularly updated.	Sunlight Foundation (2010); OpenDataBarometer (2016); Craveiro et al (2013)
Easy access	The ease with which information can be obtained through physical or electronic means.	Sunlight Foundation (2010); Open Data Index (2017); OpenGovernmentWorkingGroup (2007)
Open licence	Anyone is legally allowed to use, modify and redistribute data for any purpose. Reasonable privacy, security and privilege restrictions may be allowed.	Open Data Index (2017); OpenGovernmentWorkingGroup (2007); Sunlight Foundation (2010)

The third part of the survey was about OGD utilization. Firstly, for utilization purposes, it included one set of multiple-choice questions asking the possible purposes for them to use OGD. We listed six types of OGD utilization purposes derived from the literature for the citizens to choose from, including daily life, news report, business decision, anti-corruption, software development and scientific research. Overlaps are recognized in the scope of different types of OGD utilization, since some, like innovation and decision-making, are very broad, while others, like anti-corruption, are very specific. Thus, in the research model, we only included six specific types and excluded innovation and data analysis, which had too much intersection with other more specific types of utilization. The participants could also add additional utilization purposes beyond these six in their reply.

Secondly, for acquisition methods, we included four different methods of finding data on a portal, namely, keyword search, browse, ranking and recommendation. We selected these methods based on prior visits to more than ten different OGD portals during the data collection process in Chapter 3 when we found that these were the possible methods users could use to discover the data they needed from a portal. “Keyword search” means entering queries to get a matching list of data. “Browse” means looking through records of data gathered according to certain rules like subject, government department, updated time, etc. “Ranking” means the top ranks of data of a certain category, like the most clicked datasets, the most downloaded datasets, etc. “Recommendation” includes two kinds: one is recommending similar or related data/datasets to the dataset that the user is looking at, which we refer to as *related*, while the other is recommending datasets that users may have interest in according to their visiting history, which we refer to as *regular*. We asked the participants whether they would use these methods. For keyword search and browsing, we also asked for the use frequency and preference for the methods. A 7-point Likert scale was used for the frequency and preference, from “Strongly not frequent” to “Strongly frequent” and from “Strongly don’t like” to “Strongly like”. For the browsing method, we also asked the participants what kind of browsing they use, including browsing by subject, browsing by update time and browsing by government department. Besides these four methods, we also included “Easy access” and “Open licence” from the OGD attributes which are related to the data-acquisition process. We asked the participants about their acceptance of these two attributes. A 7-point Likert scale was used from “Strongly don’t accept” to “Strongly accept”.

Thirdly, for help functions, we selected four help functions offered by OGD portals based on our prior visits to those portals in Study 1, which were: user guide webpage, FAQ webpage, online smart/virtual agent to offer instant conversation, and customer phone line. We asked participants for their feelings about the difficulty when using the portal, whether they need the portal to offer help, their need of help functions and their choices of help functions. A 7-point Likert scale was used from “Strongly not difficult” to “Strongly difficult” and from “Strongly don’t need” to “Strongly need”. Because an online smart agent is a relatively new method for offering immediate help on a website, we included five additional questions to further analyze citizens’ attitudes towards this kind of help. We asked the participants about their acceptance of a virtual agent offering help. We also asked the participants for their preferences of the gender, age, look and role of the virtual agent. The whole survey can be found in Appendix B.

4.3.2 Data collection

Due to the recent efforts of the Chinese government in developing open data, including establishing open data policies (StateCouncil, 2016) and building local OGD portals throughout the country (R. Huang & Wang, 2016), we chose to administer the survey to a sample of the Chinese population. Because scholars noticed indistinguishable test results across recruitments of participants via social media and in person (Casler, Bickel, & Hackett, 2013) and online survey has the advantage for a better response rate (Lonsdale, Hodge, & Rose, 2006) as well as saving time and cost (Wright, 2005), we choose to collect data for analysis through social media. Also, since 61.2% of the population in China use the Internet and OGD portals mainly offer online services, it is therefore reasonable to carry out the survey online. By using an online tool called Sojump which is a free website for creating and collecting online survey data, we produced our questionnaire and distributed it from 1 to 10 August 2017 through WeChat and Weibo. We recruited citizens by putting advertisements on the above two social media.

4.4 Data analysis and results

In this section, we first give the results of the recruited sample about their demographic distribution to show a general overview of the participants of this study. Then results for the reliability and validity of all the scales of the survey are shown to support further analysis. Finally, we give the results of the survey about citizens' OGD demands including different demands for OGD subjects and attributes as the answer for RQ2.1, and citizens' OGD utilization preferences including utilization purposes, their preferences for data acquisition methods and help functions as the answer for RQ2.2. Possible correlations of citizens' demographic background and their OGD demands and utilizations are also tested in this section.

4.4.1 Participants and demographics

In total, we received 208 valid responses. In this section, we list the social demographics of the participants including their gender, age, education and occupation, as well as their knowledge of OGD and OGD portals.

4.4.1.1 Socio-demographic background

Table 4.2 shows the socio-demographic characteristics of the respondents. Females represented 63% of all respondents and males represented 36.5%. Our sample covered a wide range of ages and was almost evenly distributed in the age groups 26–30, 31–40 and 41–50 years, which were also the main groups, with a total percentage of 81.8%. Referring to education background, most respondents were in the undergraduate group (52.9%). Our sample covered all kinds of occupations listed in the survey, with student, manager and teacher being the three most popular ones. Due to the possible bias like self-selection that may occur with online surveys (Wright, 2005), we compared the socio-demographic of the recruited sample with the structure of netizens in China. It shows that the distribution of the recruited sample in age is similar to netizens in China, ranging in age from 18 to 50 (CNNIC, 2019). The occupation of the recruited sample also has similar distribution with netizens in China, which include students and manager (CNNIC, 2019). However, the educational qualification of the recruited sample is higher compared to netizens in China (CNNIC, 2019). But since most of the distributions of socio-demographic of the recruited sample are similar to netizens in China, we continued the analysis of the collected survey data.

Table 4-2 Socio-demographic characteristics

Topic	*Dimension*	*Frequency*	*Percentage*	*Topic*	*Dimension*	*Frequency*	*Percentage*
Gender	Male	76	36.5%	**Occupation**	Student	24	11.5%
	Female	131	63.0%		Production worker	9	4.3%
	Other	1	0.5%		Marketing/Salesperson	18	8.7%
Age	Under 18	2	1.0%		Customer service	4	1.9%
	18 - 25	10	4.8%		Logistics	16	7.7%
	26 - 30	54	26.0%		Human resources	7	3.4%
	31 - 40	57	27.4%		Financial/auditor	12	5.8%
	41 - 50	59	28.4%		Civilian post	12	5.8%
	51 - 60	24	11.5%		Technician	15	7.2%
	Over 60	2	1.0%		Manager	33	15.9%
Educational Qualification	Junior high	1	0.5%		Teacher	22	10.6%
	Senior high	28	13.5%		Consultant	4	1.9%
	Undergraduate	110	52.9%		Specialist	9	4.3%
	Postgraduate	57	27.4%		Other	23	11.1%
	Beyond postgraduate	12	5.8%	**Total**		208	100.0%

4.4.1.2 Knowledge of OGD and OGD portals

Regarding whether citizens had ever heard of or used OGD, results show that 59.6% of the respondents had never heard of OGD before, nor did they know about the portals (72.6%). We also found that only 19.71% of the respondents had used the data or the portals before, which is less than the rates of knowing about the data (40.38%) or the portals (27.4%). Thus, we can deduce that some of them chose not to use the data or the portals even if they had heard about them.

We carried out chi-square tests of the choices for the four questions about respondents' socio-demographic characteristics including gender, age, education and occupation, as shown in Table 4.3. Because some cohorts (combinations of characteristics) contain fewer than 5, we chose likelihood ratio instead of Pearson chi-square to test their relationships. Results show that gender was related to whether people knew about OGD. The rate for “Ever heard of OGD” in females (53.6%) was much higher than in males (46.4%). It is also apparent that educational qualification was related to people’s awareness of and usage of OGD. People with higher educational qualifications were more likely to know about and use OGD. Occupation is also related to whether people knew of and used OGD and portals. We noted that students, specialists and human resource workers were the three occupations that were more likely to know about and use OGD and portals. Age did not significantly affect people’s knowledge of OGD and portals.

Table 4-3 Likelihood ratio tests of citizens’ knowledge of OGD

Topic	*Know OGD*		*Use OGD*		*Know Portal*		*Use Portal*	
	L-R	*P*	*L-R*	*P*	*L-R*	*P*	*L-R*	*P*
Gender	6.749	**0.034***	2.889	0.236	1.086	0.581	0.930	0.628
Age	5.538	0.477	6.273	0.393	8.122	0.229	6.500	0.370
Education	19.209	**0.001****	19.505	**0.001****	4.124	0.389	4.505	0.342
Occupation	36.439	**0.001****	17.991	0.158	31.59	**0.003****	32.795	**0.002****

4.4.2 Reliability and adequacy

We examined the reliability of all the scales in the survey with Cronbach’s alpha (Cronbach, 1951). The commonly accepted range for alpha is 0.70 to 0.95 (Tavakol & Dennick, 2011). Our results, seen in Table 4.4, show high reliability for OGD demand and OGD utilization. We also examined the scales with the Kaiser–Meyer–Olkin (Bickmore & Cassell) measure of sampling adequacy (Kaiser, 1970). The results for OGD demand and OGD utilization, seen in Table 4.4, are all above 0.6, which indicates the variables are suitable for factor analysis (Dziuban & Shirkey, 1974). The significance of Bartlett’s test of sphericity is less than 0.05, which also indicates the high validity of the variables. Thus, the measurement of citizens’ OGD demand and utilization in this survey can be seen as reliable and adequate.

Table 4-4 Reliability and adequacy tests of all scales

Reliability	*Variable No.*	*Valid*	*%*	*Cronbach's Alpha*
OGD demand	22	208	100%	0.941
OGD utilization	9	162	77.90%	0.775
Adequacy	*Bartlett's Test of Sphericity*			*Kaiser-Meyer-Olkin Measure of Sampling Adequacy.*
	Approx. Chi-Square	*df*	*Sig.*	
OGD demand	3391.429	231	0.000	0.925
OGD utilization	705.346	36	0.000	0.663

4.4.3 OGD demand

We asked citizens for their demands of different OGD subjects as well as their demands for the OGD attributes emphasized by OGD principles.

4.4.3.1 OGD subjects

For OGD subjects, the survey included questions about citizens' demands of general OGD as well as the 15 specific OGD subjects. Results are shown in Table 4.5. Generally, the average scores for specific categories of data are higher than for general need, which may be due to citizens not knowing what OGD really is. The largest variation in respondents' scores lies in general need. Health and education got the highest average scores for demand. Education also got the smallest standard deviation, which indicates that citizens agreed most that this category is of high demand. International trade got the lowest average score, which is the same as with general need. Although we asked the respondents whether they had any other categories of data need, their replies show that their needs all belonged to our listed categories. Thus, we did not carry out further analysis of these replies.

Table 4-5 Descriptive analysis of citizens' demands for OGD subjects

	N	Min	Max	Mean	SD	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
General Need	208	1	7	5.13	1.738	-0.697	0.169	-0.282	0.336
Budget & Spending	208	1	7	5.56	1.403	-0.915	0.169	0.540	0.336
Credit Records	208	1	7	6.00	1.224	-1.836	0.169	4.322	0.336
Cultural activity	208	1	7	5.58	1.302	-1.099	0.169	1.555	0.336
Education	208	1	7	6.25	1.006	-2.226	0.169	7.570	0.336
Environment quality	208	1	7	6.15	1.064	-1.783	0.169	4.878	0.336
Government bid	208	1	7	5.19	1.633	-0.669	0.169	-0.179	0.336
Health	208	1	7	6.25	1.156	-2.254	0.169	6.289	0.336
International Trade	208	1	7	5.13	1.563	-0.762	0.169	0.128	0.336
Local statistics	208	1	7	6.22	1.058	-2.280	0.169	7.580	0.336
Maps	208	1	7	5.71	1.252	-1.405	0.169	2.691	0.336
Policies & legislation	208	1	7	6.05	1.160	-1.755	0.169	4.040	0.336
Public Safety	208	1	7	6.22	1.099	-2.193	0.169	6.561	0.336
Registration	208	1	7	5.46	1.503	-1.063	0.169	0.836	0.336
Transportation	208	1	7	6.17	1.120	-2.033	0.169	5.636	0.336
Weather	208	1	7	5.85	1.174	-1.218	0.169	1.930	0.336

We carried out two independent sample T-tests for the association between gender and citizens' demands for different OGD subjects. Results show no significant differences in the average and standard deviation of male and female scores for different OGD subjects.

We carried out one-way ANOVA to analyze the effects of age, education and occupation on people's OGD demand, as shown in Table 4.7. We found that age had a significant effect on respondents' demands for registration and international trade data. Compared with citizens older than 25, citizens of age 18–25 had a significantly lower demand for registration data. Citizens of age 31–40 had a significantly higher demand for international trade data than citizens of age 41–50. People of different educational background had significantly different demands for registration, budget and spending, and government bid data. Citizens of master's degree level showed a significantly higher demand for registration data than citizens of bachelor's degree level. Citizens of doctoral degree level showed a significantly higher demand for budget and spending data than citizens of bachelor's degree level. Citizens of doctoral degree level also showed a significantly higher demand for government bid data than citizens of bachelor's and master's degree levels. Occupation affected the demand for cultural activity and transportation data significantly. Citizens with civilian posts had a significantly lower demand for cultural activity and transportation data than citizens of other occupations.

We carried out two independent sample T-tests for the association between citizens' OGD knowledge and their demands for different OGD subjects, as shown in Table 4.6. There were significant differences in citizens' demands of OGD in general between those who knew of and used OGD and OGD portals before and those who did not. Citizens who knew of OGD also showed higher demands for policies and legislation data compared with those who did not know of OGD. Citizens who had used OGD before showed higher demands for map data than those who had not.

Table 4-6 T-tests of citizens' knowledge of OGD and their demands of OGD subjects

	Know OGD						Use OGD						Know OGD Portal						Use OGD Portal					
	Avg.		F	P	t	p	Avg.		F	P	t	p	Avg.		F	P	t	p	Avg.		F	P	t	p
General Need	Yes	5.67	1.437	0.232	3.782	0.000**	Yes	5.78	3.096	0.080	2.874	0.004**	Yes	5.68	3.069	0.081	2.875	0.004**	Yes	5.73	2.487	0.116	2.506	0.013*
	No	4.77					No	4.95					No	4.92					No	4.98				
Budget & Spending	Yes	5.69	0.401	0.527	1.124	0.262	Yes	5.71	0.685	0.409	0.828	0.409	Yes	5.65	0.001	0.976	0.577	0.565	Yes	5.59	0.376	0.540	0.141	0.888
	No	5.47					No	5.52					No	5.52					No	5.55				
Credit Records	Yes	6.07	0.561	0.455	0.692	0.490	Yes	6.00	0.676	0.412	0.000	1.000	Yes	6.16	0.101	0.751	1.144	0.254	Yes	5.9	5.202	0.024	-0.461	0.647
	No	5.95					No	6.00					No	5.94					No	6.02				
Cultural activity	Yes	5.54	1.762	0.186	-0.375	0.708	Yes	5.47	0.022	0.882	-0.641	0.522	Yes	5.54	0.017	0.895	-0.225	0.823	Yes	5.34	0.502	0.480	-1.295	0.197
	No	5.60					No	5.61					No	5.59					No	5.63				
Education	Yes	6.32	0.069	0.794	0.785	0.433	Yes	6.27	0.100	0.752	0.089	0.929	Yes	6.39	0.423	0.516	1.156	0.249	Yes	6.24	0.219	0.640	-0.077	0.938
	No	6.21					No	6.25					No	6.21					No	6.26				
Environment quality	Yes	6.29	1.986	0.160	1.530	0.128	Yes	6.27	1.657	0.200	0.837	0.404	Yes	6.33	2.294	0.131	1.540	0.125	Yes	6.29	1.296	0.256	0.965	0.336
	No	6.06					No	6.12					No	6.08					No	6.11				
Government bid	Yes	5.14	6.221	0.013	-0.344	0.731	Yes	5.29	1.616	0.205	0.447	0.655	Yes	5.00	3.028	0.083	-1.043	0.298	Yes	4.93	3.252	0.073	-1.162	0.246
	No	5.23					No	5.17					No	5.26					No	5.26				
Health	Yes	6.20	0.115	0.735	-0.439	0.661	Yes	6.00	0.351	0.554	-1.614	0.108	Yes	6.26	0.017	0.897	0.137	0.891	Yes	6.07	0.378	0.539	-1.064	0.289
	No	6.27					No	6.31					No	6.24					No	6.29				
International Trade	Yes	5.21	1.517	0.219	0.641	0.522	Yes	5.42	0.168	0.682	1.421	0.157	Yes	5.18	0.894	0.346	0.258	0.797	Yes	5.15	4.747	0.030	0.066	0.948
	No	5.07					No	5.05					No	5.11					No	5.13				
Local statistics	Yes	6.17	0.177	0.674	-0.610	0.542	Yes	6.04	0.009	0.923	-1.267	0.207	Yes	6.23	0.350	0.555	0.058	0.954	Yes	6.07	0.274	0.601	-0.999	0.319
	No	6.26					No	6.27					No	6.22					No	6.26				
Maps	Yes	5.86	2.828	0.094	1.383	0.168	Yes	6.02	5.044	0.026	2.092	0.039*	Yes	5.93	3.293	0.071	1.550	0.123	Yes	5.95	2.514	0.114	1.370	0.172
	No	5.61					No	5.63					No	5.63					No	5.65				
Policies & legislation	Yes	6.25	4.223	0.041	2.151	0.033*	Yes	6.13	0.030	0.863	0.525	0.600	Yes	6.28	0.638	0.425	1.749	0.082	Yes	6.12	0.259	0.611	0.425	0.671
	No	5.92					No	6.03					No	5.97					No	6.04				
Public Safety	Yes	6.32	1.285	0.258	1.084	0.280	Yes	6.20	0.298	0.586	-0.146	0.884	Yes	6.39	0.942	0.333	1.332	0.184	Yes	6.22	0.144	0.705	-0.011	0.992
	No	6.15					No	6.23					No	6.16					No	6.22				
Registration	Yes	5.40	2.429	0.121	-0.448	0.655	Yes	5.38	2.664	0.104	-0.421	0.674	Yes	5.44	2.293	0.132	-0.135	0.893	Yes	5.32	8.062	0.005	-0.579	0.565
	No	5.50					No	5.48					No	5.47					No	5.5				
Transportation	Yes	6.27	1.309	0.254	1.068	0.287	Yes	6.20	0.134	0.715	0.182	0.856	Yes	6.40	0.570	0.451	1.833	0.068	Yes	6.27	0.088	0.767	0.606	0.545
	No	6.10					No	6.17					No	6.09					No	6.15				
Weather	Yes	5.95	3.333	0.069	1.075	0.284	Yes	5.89	0.551	0.459	0.275	0.783	Yes	6.00	0.416	0.520	1.163	0.246	Yes	5.95	0.000	0.999	0.639	0.524
	No	5.77					No	5.83					No	5.79					No	5.82				

Table 4-7 One-way ANOVA of age, education, occupation and OGD subjects

	Age		Education		Occupation	
	F	P	F	P	F	P
General Need	0.840	0.540	0.459	0.766	0.943	0.510
Budget & Spending	1.885	0.085	2.472	**0.046***	1.419	0.153
Credit Records	1.971	0.071	1.107	0.354	0.977	0.475
Cultural activity	1.098	0.365	0.416	0.797	1.809	**0.044***
Education	0.452	0.843	1.268	0.284	1.474	0.130
Environment quality	0.694	0.655	1.345	0.255	0.820	0.639
Government bid	1.971	0.071	2.712	**0.031***	1.550	0.103
Health	1.340	0.241	0.312	0.870	1.058	0.398
International Trade	2.283	**0.037***	1.460	0.216	1.195	0.285
Local statistics	0.817	0.558	0.254	0.907	1.130	0.336
Maps	0.514	0.797	0.831	0.507	1.147	0.322
Policies & legislation	1.240	0.287	1.674	0.157	0.905	0.549
Public Safety	0.475	0.826	1.359	0.249	0.737	0.725
Registration	2.762	**0.013***	2.552	**0.040***	1.520	0.113
Transportation	0.573	0.751	0.022	0.999	1.852	**0.038***
Weather	0.958	0.455	0.211	0.932	0.836	0.621

4.4.3.2 OGD attributes

We first calculated respondents' demands for different OGD attributes on OGD portals. Results in Table 4.8 show that "Up to date" got the highest mean value and the lowest standard deviation, showing citizens' agreements on its great need, while "Download in bulk" got the lowest score but the highest standard deviation.

Table 4-8 Descriptive analysis of OGD attributes

	N	Min	Max	Mean	Median	SD
Download in bulk	208	1	7	5.48	6.00	1.404
Free of charge	208	1	7	6.17	7.00	1.190
Up to date	208	1	7	6.21	6.00	1.014
Update regularly	208	1	7	6.20	6.00	1.056
Complete metadata	208	1	7	5.66	6.00	1.309
Machine-readable	208	1	7	5.79	6.00	1.229

We carried out two independent sample T-tests for the association between gender and citizens' demands for different OGD attributes. Results show no significant differences in the average and standard deviation of male and female scores for different OGD attributes.

We carried out one-way ANOVA to analyze the effects of age, education and occupation on respondents' demands for different data qualities. Results are shown in Table 4.9. We found that age had a significant effect on "Download in bulk", "Free of charge", "Up to date" and "Complete metadata". Citizens of age 41–60 showed significantly lower demand for "Download in bulk" than citizens of age 18–40. Citizens of age 41–50 showed significantly lower demand for "Free of charge" and "Up to date" than citizens of age 26–40. Citizens of age 41–50 also showed significantly lower demand for "Complete metadata" compared with citizens of age 26–40 and 51–60. People of different educational background had significantly different demands for all OGD attributes except "Machine-readable". Citizens of master's and doctoral degree levels showed higher demands for "Download in bulk" than others. Citizens of master's degree level showed higher demands for "Free of charge", "Up to date", "Update regularly" and "Complete metadata" than citizens of bachelor's degree level. Occupation affected the demands for "Download in bulk", "Up to date", "Complete metadata" and "Machine-readable". Students showed significant higher demands for "Download in bulk", "Complete metadata" and "Machine-readable" compared with citizens of other occupations. Citizens of civilian posts showed significantly lower demand for "Up to date" compared with citizens of other occupations.

Table 4-9 One-way ANOVA of age, education, occupation and OGD attributes

	Age		Education		Occupation	
	F	P	F	P	F	P
Download in bulk	6.276	**0.000****	5.238	**0.000****	1.891	**0.033***
Free of charge	3.186	**0.005****	3.493	**0.009****	0.939	0.513
Up to date	2.539	**0.022***	3.007	**0.019***	1.792	**0.047***
Update regularly	1.801	0.100	2.637	**0.035***	1.595	0.089
Complete metadata	2.817	**0.012***	3.121	**0.016***	3.645	**0.000****
Machine-readable	1.688	0.126	1.760	0.138	2.417	**0.005****

We carried out two independent sample T-tests for the association between citizens' OGD knowledge and their demands for different OGD attributes, as shown in Table 4.10. Citizens who knew of OGD showed significantly higher demands for OGD attributes than those who did not know of OGD. Citizens who had used OGD before showed higher demand for "Free of charge" compared with those who had not used OGD before. Citizens who knew of OGD portals showed significantly higher demands for "Download in bulk", "Free of charge" and "Machine-readable" compared with those who did not know of OGD portals. Citizens who had used OGD portals before showed higher demand for "Download in bulk" than those who had not used OGD portals.

Table 4-10 T-tests of citizens' knowledge of OGD and OGD attributes

	Know OGD						Use OGD					
	Avg.		F	P	t	p	Avg.		F	P	t	p
Download in bulk	Yes	5.76	5.099	0.025	2.526	**0.012***	Yes	5.76	4.749	0.030	1.676	0.098
	No	5.28					No	5.40				
Free of charge	Yes	6.42	4.697	0.031	2.745	**0.007****	Yes	6.49	2.682	0.103	2.058	**0.041***
	No	6.00					No	6.08				
Up to date	Yes	6.39	3.807	0.052	2.142	**0.033***	Yes	6.38	3.306	0.070	1.244	0.215
	No	6.09					No	6.17				
Update regularly	Yes	6.38	4.610	0.033	2.255	**0.025***	Yes	6.36	2.626	0.107	1.138	0.257
	No	6.07					No	6.15				
Complete metadata	Yes	5.81	0.597	0.441	1.327	0.186	Yes	5.76	1.237	0.267	0.532	0.595
	No	5.56					No	5.64				
Machine-readable	Yes	3.01	5.539	0.020	2.250	**0.026***	Yes	5.98	4.465	0.036	1.281	0.204
	No	5.64					No	5.74				
	Know OGD Portal						Use OGD Portal					
	Avg.		F	P	t	p	Avg.		F	P	t	p
Download in bulk	Yes	5.77	8.735	0.003	2.132	**0.035***	Yes	5.83	5.347	0.022	2.093	**0.040***
	No	5.36					No	5.39				
Free of charge	Yes	6.46	3.574	0.060	2.163	**0.032***	Yes	6.44	1.892	0.170	1.632	0.104
	No	6.06					No	6.1				
Up to date	Yes	6.42	1.724	0.191	1.842	0.067	Yes	6.41	2.124	0.147	1.435	0.153
	No	6.13					No	6.16				
Update regularly	Yes	6.39	2.360	0.126	1.590	0.113	Yes	6.37	1.784	0.183	1.142	0.255
	No	6.13					No	6.16				
Complete metadata	Yes	5.82	0.461	0.498	1.091	0.276	Yes	5.71	1.351	0.246	0.239	0.811
	No	5.60					No	5.65				
Machine-readable	Yes	6.07	4.687	0.032	2.255	**0.026***	Yes	6.05	0.654	0.419	1.518	0.130
	No	5.68					No	5.72				

4.4.4 OGD utilization

In this section, we give the results for citizens' purposes of using OGD, their data-acquisition methods and their preferences of help functions offered by OGD portals.

4.4.4.1 Utilization purpose

We gave respondents seven specific choices of utilization purposes for OGD, which were daily life, software development, scientific research, news report, business decision, anti-corruption and other. These are the commonly accepted utilizations of OGD found in the literature (Safarov et al., 2017) and have little intersection with one another. Respondents could choose as many types as they wanted. Results show that citizens mainly used OGD for daily life (25.7%), followed by anti-corruption (20%). However, these two types of data involve direct utilization of OGD that does not require deeper exploration of the data or innovative uses. The least utilized data concerned software development (3.6%).

We examined the correlation between users' socio-demographic characteristics and their knowing of OGD with their choices of utilization, with chi-square tests. We chose likelihood ratio instead of Pearson chi-square to test the results. From Table 4.11, there is a high possibility that citizens' education affected their OGD utilization. Users' ever knowing of and using OGD also affected their choices of utilization. For users' education background, results show that highly educated users were less likely to use OGD for daily life, but more likely to use it for scientific research. The percentage of use of business decision and anti-corruption data were almost the same in senior high, undergraduate and postgraduate users. Results also show that citizens who had known of and used OGD were more likely to choose scientific research and business decision as their purpose of utilization, while citizens who had not known of and used OGD preferred to use OGD for daily life and news reports. The differences based on users' demographics in usage of software development and anti-corruption data were not significant.

Table 4-11 Chi-square tests of OGD utilization purpose

Socio-demographic	Likelihood Ratio			Knowledge of OGD	Likelihood Ratio		
	Value	df	sig.		Value	df	sig.
Gender	4.631	6	0.592	Know OGD	22.83	6	**0.001****
Age	41.78	36	0.234	Use OGD	15.62	6	**0.016***
Education	39.66	18	**0.002****	Know portal	6.761	6	0.343
Occupation	59.93	78	0.936	Use portal	9.874	6	0.130

4.4.4.2 Acquisition methods

Figure 4.2 shows results regarding acceptance of the four methods of data acquisition, namely, keyword search, browse, ranking and recommendation. We found over 90% of the respondents used keyword search for the data they wanted on a portal, followed by the related datasets automatically recommended by the portal. About 25.5% of the respondents chose not to refer to rankings when looking for datasets on a portal, making it the method with the lowest acceptance rate.

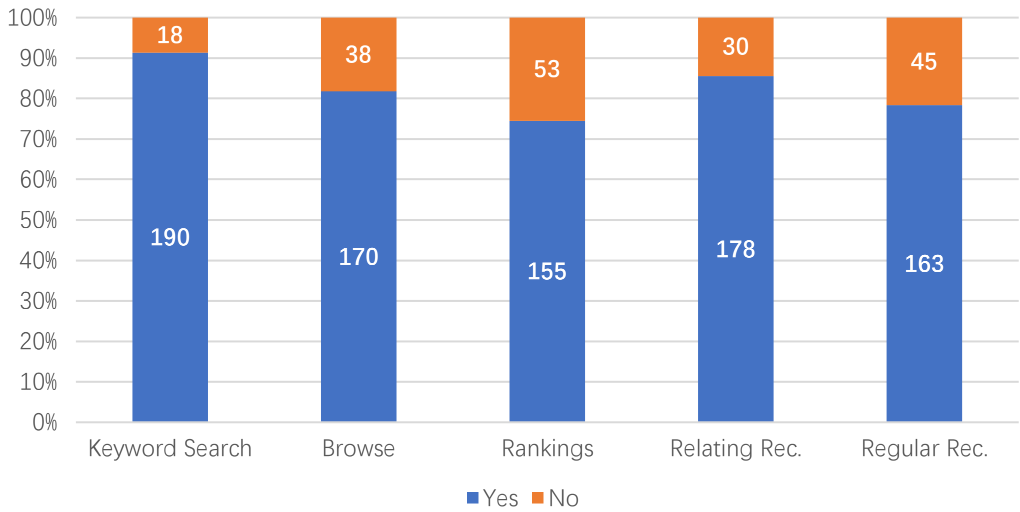
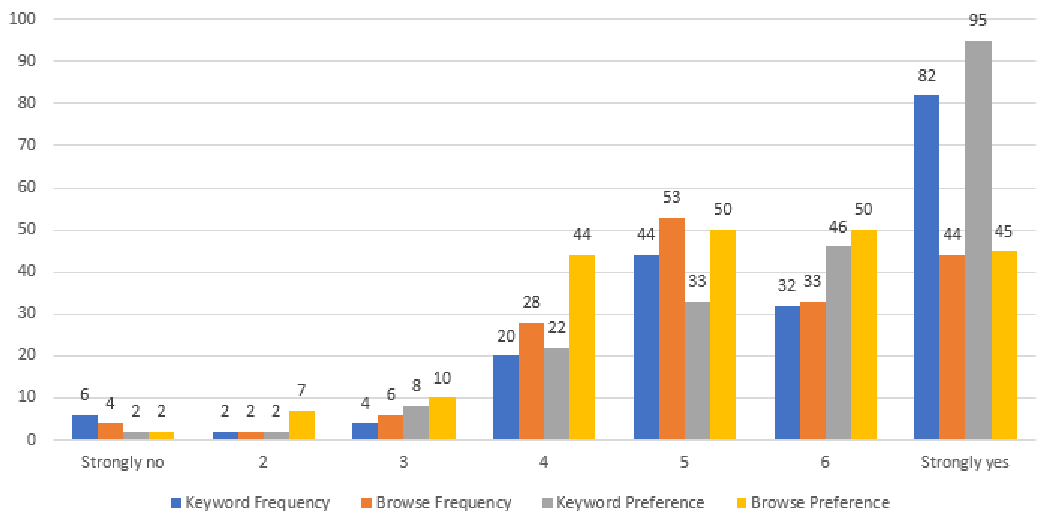


Figure 4-2 Citizens' acceptance of data-acquisition methods

We compared respondents' reported usage frequency and preferences for keyword search and browsing, as shown in Figure 4.3. According to the results, we found that more respondents chose to use keyword search to find the data they needed, since 190 out of 208 respondents had ever used keyword search, while only 170 had used browsing. However, the mean score for the usage frequency of browsing was much higher than for keyword search, with a relatedly small standard deviation. This indicates that using frequency for browsing resulted in a higher value than with keyword search. On the other hand, keyword search got a higher preference score than browsing.



	Keyword Frequency	Browse Frequency	Keyword Preference	Browse Preference
Valid N.	190	170	208	208
Mean	5.73	5.88	5.35	5.23
Std. Deviation	1.472	1.332	1.381	1.387

Figure 4-3 Reported usage frequency and preferences of keyword search and browse

We carried out chi-square tests of whether gender influenced people's acceptance of regular recommendation of datasets based on their visit history. The likelihood ratio (=10.501) shows a strong relationship between the two (sig. (2-sided) p=0.005). Compared with males, females were more likely to accept regular recommendations, as shown in Figure 4.4.

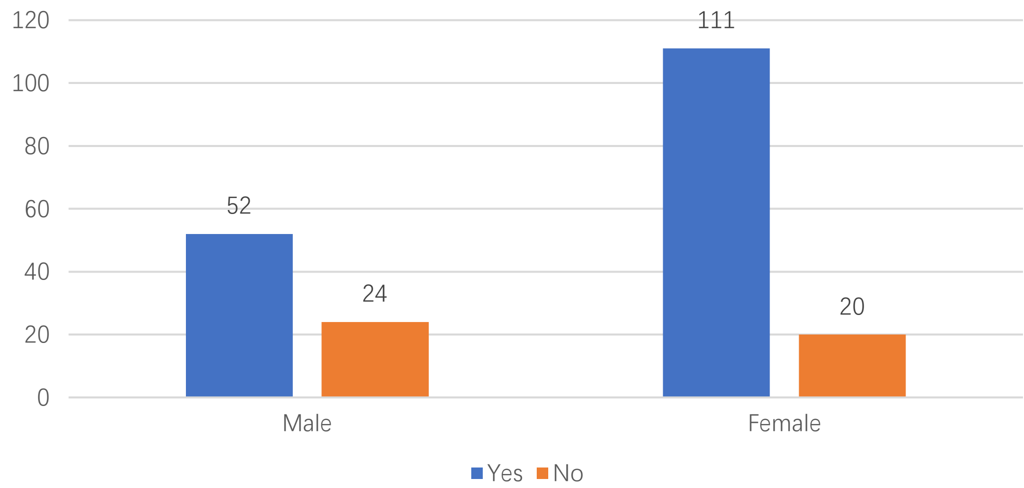


Figure 4-4 Differences in acceptance of regular recommendations between genders

4.4.4.3 Help functions

Generally, users of OGD portals felt it was quite difficult to get the data they needed from the portal, with an average score of 4.94 and standard deviation of 1.654. Since the sample size for users' socio-demographic characteristics was the same as for their feeling of difficulty, it was statistically robust to carry out two independent sample T-tests to determine the relationship between their feeling of difficulty and their gender, as well as a one-way ANOVA test of users' feeling of difficulty with their age, education background and occupation. Results show that only educational level had a significant effect on people's feelings about data access difficulty ($F=2.95$, $P=0.021$). Compared with citizens with senior high school degrees and bachelor's degrees, citizens with doctoral degrees showed a higher score of difficulty in finding the data they needed from the portal.

Regarding the help functions, we first calculated the percentage of respondents' acceptance of help functions of OGD portals. 90.9% of the respondents needed help from the portal. The chi-square tests of citizens' acceptance of help functions of a portal show no relationship with their gender, age, occupation or educational background. But the chi-square tests show a significant relationship between citizens' acceptance of help functions of a portal with whether they had ever used OGD portals before, with likelihood ratio=8.851, $p=0.003$. Compared with participants who had used OGD portals before, more who had never used them selected that they needed help from the OGD portal. We examined the correlations between users' need of help with their feeling of difficulty of data acquisition from OGD portals. Results show that there was a strong possibility (sig. (2-tailed) $p = 0.001$) that these two have a positive relationship, the Pearson correlation being 0.236.

Referring to citizens' extent of need of help functions, results show users' strong demand for OGD portals to offer help, with a mean of 5.34 and standard deviation of 1.331. Their choices gathered on 5 (26.4%) and 7 (25%) of the 7-point scale.

The two independent sample T-tests and one-way ANOVA tests of citizens' need of help functions of a portal show no relationship to their gender, age, occupation or educational background. The two independent sample T-tests also show no significant differences in the need of help functions between citizens who knew of or used OGD and OGD portals before and those who did not.

We also calculated the frequencies and percentages of users' choices of different help functions, as shown in Figure 4.5. The differences were not significant among different kinds of help functions. FAQ was preferred by most respondents, while customer service phone line received fewest selections. The chi-square tests of people's choices of help functions show no relation to their gender, age, occupation or education background.

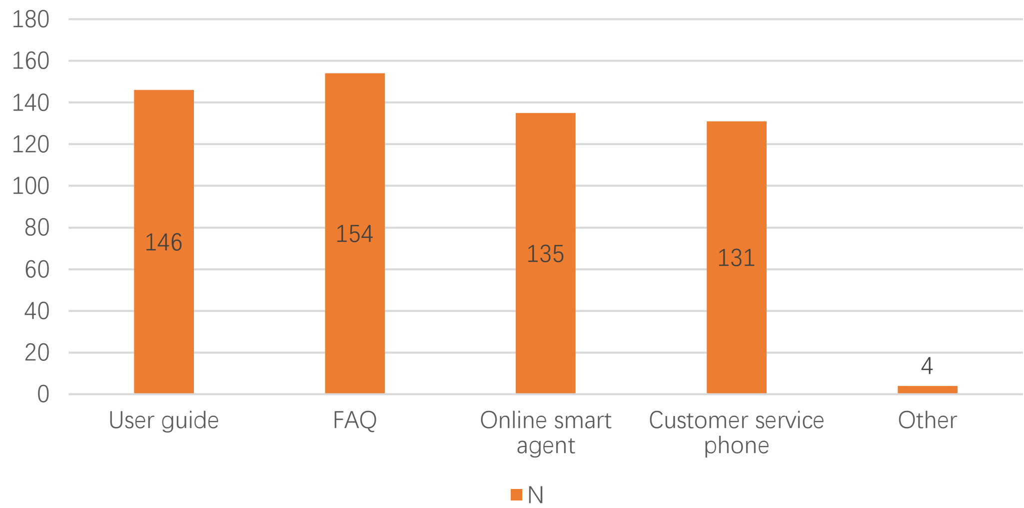


Figure 4-5 Frequency and percentage of help functions

4.4.4.4 Virtual agents

We further analyzed questions regarding the use of a virtual agent, which is a relatively new approach used by websites to offer help to end-users. The investigation of virtual agents focused on two aspects: citizens' acceptance of the virtual agent, and their preferences of the appearance of the virtual agent.

For the first aspect, among the 7-point scales of their acceptance of a virtual agent, respondents gathered on 5 (25.5%), followed by 6 (23.1%). The average score of their acceptance (5.07) also shows their high acceptance of this kind of help. The two independent sample T-tests and one-way ANOVA show that people's acceptance of virtual agents had no relation with their gender, age, occupation or education background. But the two independent sample T-tests show there was a significant difference in participants' acceptance of a virtual agent between those who knew of OGD portals and those who did not, with $F=0.576$, $P=0.449$, $t=2.344$, $p=0.02$. The average acceptance score for participants who knew of OGD portals is 5.44, while the average acceptance score for participants who did not know of OGD portals is 4.93.

For the second aspect of people's preferences of virtual agents' appearance, we can see from Figure 4.6 that most people did not care about the gender of the virtual agent, but female virtual agents were preferred by more respondents than male virtual agents. Most people preferred the virtual agent to be younger than themselves or of the same age. They did not mind whether the virtual agent looked like themselves or not, but there were more people who preferred the virtual agent to not look like themselves. Friend was the most preferred role of a virtual agent, much more than the other options.

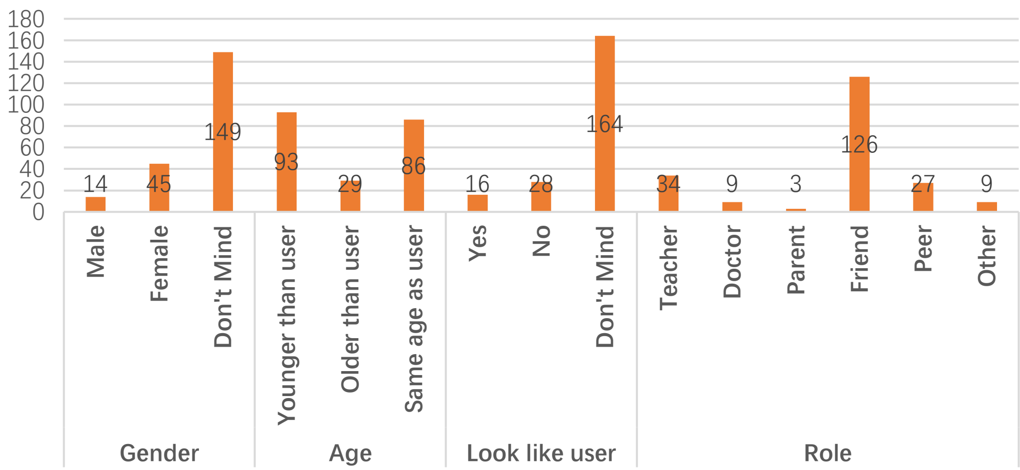


Figure 4-6 Frequency of people's preference of virtual agents' appearance

We carried out chi-square tests of people's preferences of virtual agents' appearance with their demographic characteristics including gender, age, education and occupation. We chose likelihood ratio instead of Pearson chi-square to test these relationships because some groups had fewer than 5. The results in Table 4.12 show that gender was related to people's preferences of virtual agents' role. Males preferred the virtual agent to be parent and teacher, while females preferred peer, friend and doctor much more than males, as shown in Figure 4.7. Age had a strong relationship with people's choice of age and whether the virtual agent looked like the user themselves, as shown in Figure 4.8. Generally, with an increase in age, people preferred virtual agents to be younger than themselves, instead of being the same age. People aged between 18 to 25 showed a strong preference for the virtual agent being not like themselves. People's education background affected their preference of the gender of the virtual agent, as shown in Figure 4.9. People with higher education background showed less preference for male virtual agents. People's knowing of OGD affected their choice of gender for the virtual agent. Compared with those who knew of OGD, there were more participants who did not know of OGD who chose male virtual agents. Whether the participants had ever used OGD affected their choice of role for the virtual agent. Compared with those who had used OGD before, those who had never used OGD were more likely to accept a virtual agent acting like a doctor.

Table 4-12 Chi-square tests of citizens' preferences of virtual agents' appearance

Topic	*Gender (A)*		*Age (A)*		*Look like user*		*Role (A)*	
	L-R	*P*	*L-R*	*P*	*L-R*	*P*	*L-R*	*P*
Gender	7.403	0.116	6.208	0.184	2.488	0.647	19.283	**0.037***
Age	16.099	0.187	31.816	**0.001****	32.763	**0.001****	37.986	0.150
Education	16.181	**0.040***	5.269	0.728	8.082	0.426	22.392	0.320
Occupation	26.263	0.449	25.661	0.482	34.148	0.131	71.161	0.280
Know OGD	17.548	**0.000****	0.275	0.871	0.184	0.912	7.384	0.194
Use OGD	3.558	0.169	1.347	0.510	1.083	0.582	11.138	**0.049***
Know Portal	5.211	0.074	0.855	0.652	0.162	0.922	4.154	0.528
Use Portal	4.165	0.125	1.743	0.418	1.966	0.374	1.378	0.927

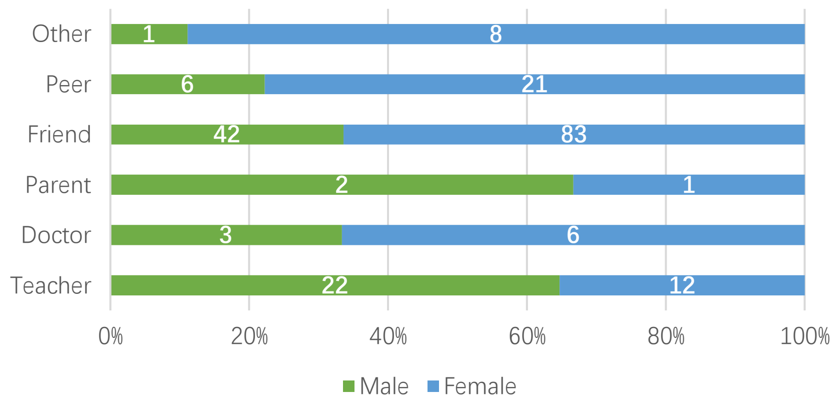


Figure 4-7 Distribution of different gender participants' preferences of virtual agent's role

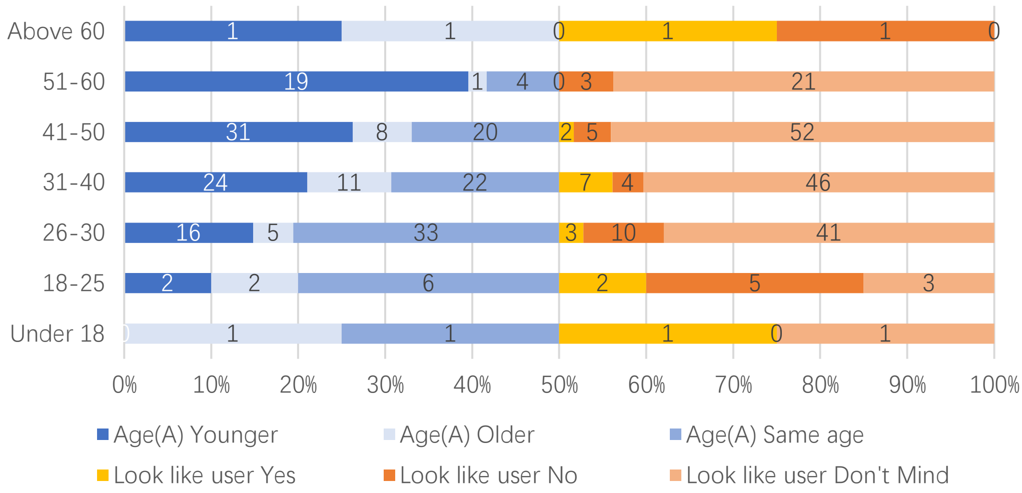


Figure 4-8 Distribution of different aged participants' preferences of virtual agent's appearance

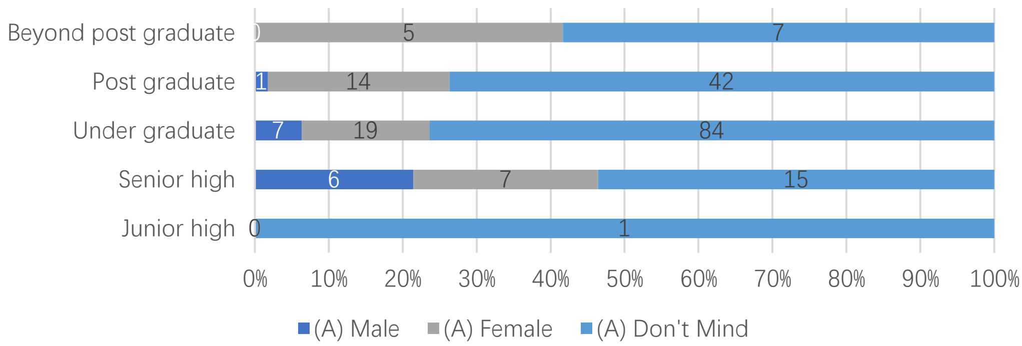


Figure 4-9 Distribution of different education background participants' preferences of virtual agent's gender

4.5 Discussion

In this section, we first give a further discussion about Chinese citizens' present knowledge of OGD and OGD portals due to the interesting results shown in the last section. Then analysis and discussion of the collected data according to our hypotheses proposed in Section 4.2 are displayed, as the answer to RQ2.3.

4.5.1 Knowledge of OGD

Our analysis of citizens' knowledge of OGD shows that currently, relatively few of them knew about or had ever used OGD and the OGD portals established by the government. This finding is in line with previous studies showing that a serious problem with OGD development is its usage, compared with its disclosure (Heise & Naumann, 2012; Kuhn, 2011; Whitmore, 2014; Zuiderwijk & Janssen, 2013). Since OGD is intended to reduce the asymmetry of information between citizens and government bodies (Fuentes-Enriquez & Rojas-Romero, 2013; Murillo, 2015), it becomes a precondition for OGD utilization that citizens are aware of the data being released by the government. Although governments have built portals to make their open data accessible to the public, it seems that their goals of promoting transparency and facilitating accountability (Attard et al., 2015) have not been achieved due to citizens' lack of awareness of their efforts. Therefore, it is important for governments to make efforts to improve citizens' awareness of the data that they have disclosed. Referring to Chinese OGD portals, they could stop blocking search engines (D. Wang, Chen, & Richards, 2018) for collecting portal information and also put links for OGD portals on other e-government websites to increase online visibility (K. Yi & Jin, 2008).

Our results also show some citizens chose not to use OGD or portals even after they were informed about them, which goes against the conclusion of Wijnhoven et al. (2015) that citizens who already knew about an OGD project were more willing to participate than those who had never heard of OGD before. Thus, researchers could study citizens' motivations for OGD utilization in future studies. If we understood why citizens chose not to use OGD, it would help governments to improve OGD quality and services accordingly to promote utilization.

4.5.2 Citizens and their OGD demands

Our analysis in Section 4.4.3 has shown citizens of different demographic backgrounds have different preferences in both the OGD subjects and the OGD attributes. The results support our first hypothesis (H1) that citizens' characteristics have an effect on their OGD demands. The statistics also support their greater demands for health and education data, which has the potential to make key public services more effective and inclusive (OpenDataBarometer, 2017b). However, according to the Open Data Barometer's fourth report in 2017, although the rate of open datasets published by governments in health (7%) and education (8%) is not the lowest (1%) among all OGD subjects, their availability shows a significant decline. In addition, health data also shows low quality in updating and sustainability. This would affect public services for citizens as a whole. Therefore, governments should improve their data quality and open data that citizens show great need of in order to reach their aim of OGD programs. For Chinese OGD portals, it would be helpful in encouraging citizens to use the data if they could release more data about health and education, and update the data regularly to ensure availability of the latest version. Our results about citizens' preference for map data for scientific research support the OpenDataBarometer (2017b) view of using map data for innovation and to produce significant economic value.

4.5.3 Citizens and OGD utilization

The analysis in Section 4.4.4 has shown citizens' socio-demographic characteristics and their knowledge of OGD and OGD portals affect their OGD utilization purposes. This finding supports the view that different end-users have different issues to solve (Susha, Grönlund, et al., 2015) when visiting an OGD portal. Females' preference for the recommendation function shows the effect of gender on citizens' selection of OGD data-acquisition methods. Citizens' socio-demographic characteristics do not affect their selection of different help functions. But the characteristics of citizens have an effect on their choices of the appearance of the virtual agent offering help on OGD portals. These results show support for our second hypothesis (H2) that citizens' characteristics have an effect on their OGD utilization.

4.5.4 OGD demands and utilization

To examine our third hypothesis about citizens' utilization affecting their OGD demand, we first used two independent sample T-tests to test the 15 subjects of OGD with the six different purposes of utilization. The results, shown in Table 4.13, indicate that citizens who chose to use OGD in daily life scored a much higher demand for health data than those who did not choose daily life. Citizens were also in greater need of map data if they used OGD for scientific research. For business decisions, users showed higher demands for maps, credit records, international trade, and budget and spending data from government. For users interested in anti-corruption data, citizens reported higher needs for health, cultural activity, policies and legislation, weather and registration data. For other subjects that are not listed in Table 4.13, results did not show a significant difference in the scores for demand within each purpose of OGD utilization. The relationships between citizens' OGD utilization purpose and their demand of certain data subjects show support for the view that different societal issues involve different needs of different users (Susha, Grönlund, et al., 2015). Thus, it is possible to improve the usage of open data in the future development of OGD portals by recognizing users' requirements for specific purposes or contexts during their visits to OGD portals or platforms (Power et al., 2015; Ruijer, Grimmelikhuijsen, Hogan, et al., 2017).

Table 4-13 T-tests of OGD utilization purposes and demands for OGD subjects

Utilization	Subject	Select		Not Select		F	P	t	df	p
		Mean	Std. D	Mean	Std. D					
Daily Life	Health	6.41	1.034	5.88	1.327	4.823	**0.029***	-2.860	98.396	**0.005****
Scientific Research	Maps	5.91	1.040	5.56	1.380	7.332	**0.007****	-2.125	205.813	**0.035***
Business Decision	Maps	5.99	1.083	5.46	1.344	6.066	**0.015***	-3.151	203.190	**0.002****
	Credit Records	5.75	1.466	5.20	1.496	14.317	**0.000****	-4.225	173.917	**0.000****
	International Trade	5.63	1.314	4.68	1.638	4.568	**0.034***	-4.620	202.965	**0.000****
	Budget & Spending	5.95	1.173	5.20	1.502	7.224	**0.008****	-4.021	201.566	**0.000****
Anti-corruption	Health	6.49	0.880	5.96	1.360	7.500	**0.007****	-3.291	157.989	**0.001****
	Cultural activity	5.88	1.092	5.23	1.440	10.903	**0.001****	-3.597	175.248	**0.000****
	Policies & legislation	6.32	0.932	5.74	1.316	8.653	**0.004****	-3.623	167.730	**0.000****
	Weather	6.10	0.986	5.55	1.305	11.003	**0.001****	-3.360	174.899	**0.001****
	Registration	5.73	1.294	5.15	1.667	7.412	**0.007****	-2.798	177.903	**0.006****

Secondly, we analyzed the relationship between users' data-acquisition method and their need for different OGD attributes. We carried out two independent sample T-tests to see whether users' need for different OGD attributes had any relationship with their data-acquisition method. The results, in Table 4.14, show that users who used keyword search ranked "Download in bulk" and "Machine-readable" higher than those who did not use keyword search. Users who browsed sought data that was "Free of charge", "Up to date", "Complete metadata" and "Machine-readable" more than users who did not browse. Users who referred to the ranks provided were in greater need of all six attributes than users who did not. Users who accepted recommendations showed greater need for all attributes except "Complete metadata".

Table 4-14 T-tests of data-acquisition methods and OGD attributes

		Download in bulk						Free of Charge						Up to date					
		Mean	Std.D	F	P	t	P	Mean	Std.D	F	P	t	P	Mean	Std.D	F	P	t	P
Keyword Search	Yes	5.56	1.363	0.115	0.735	2.779	**0.006****	6.22	1.119	9.326	0.003	1.475	0.157	6.26	0.928	10.589	0.001	1.549	0.139
	No	4.61	1.577					5.61	1.720					5.67	1.609				
Browse	Yes	5.53	1.333	5.595	0.019	1.003	0.321	6.28	1.056	20.257	0.000	2.321	**0.025***	6.31	0.899	9.655	0.002	2.399	**0.021***
	No	5.24	1.684					5.66	1.582					5.76	1.344				
Rankings	Yes	5.70	1.301	1.291	0.257	4.019	**0.000****	6.34	1.047	8.669	0.004	3.195	**0.002****	6.34	0.921	1.355	0.246	3.077	**0.002****
	No	4.83	1.503					5.66	1.427					5.85	1.183				
Relating Rec.	Yes	5.57	1.339	2.980	0.086	2.459	**0.015***	6.27	1.123	2.318	0.129	3.053	**0.003****	6.30	0.919	3.700	0.056	3.256	**0.001****
	No	4.90	1.647					5.57	1.406					5.67	1.348				
Regular Rec.	Yes	5.57	1.352	1.366	0.244	1.861	0.064	6.28	1.125	2.579	0.110	2.667	**0.008****	6.34	0.924	1.384	0.241	3.500	**0.001****
	No	5.13	1.546					5.76	1.334					5.76	1.190				

		Update Regularly						Complete metadata						Machine-readable					
		Mean	Std.D	F	P	t	P	Mean	Std.D	F	P	t	P	Mean	Std.D	F	P	t	P
Keyword Search	Yes	6.25	0.935	22.336	0.000	1.296	0.211	5.72	1.282	0.551	0.459	1.885	0.061	5.84	1.180	3.795	0.053	2.061	**0.041****
	No	5.67	1.879					5.11	1.491					5.22	1.592				
Browse	Yes	6.26	0.981	2.076	0.151	1.792	0.075	5.76	1.265	1.755	0.187	2.387	**0.018***	5.88	1.161	5.580	0.019	2.054	**0.045***
	No	5.92	1.323					5.21	1.417					5.37	1.441				
Rankings	Yes	6.34	0.921	9.351	0.003	3.213	**0.002****	5.80	1.250	1.344	0.248	2.609	**0.010***	5.93	1.179	2.166	0.143	2.870	**0.005****
	No	5.85	1.183					5.26	1.403					5.38	1.289				
Relating Rec.	Yes	6.33	0.894	17.030	0.000	3.258	**0.003****	5.72	1.271	0.357	0.551	1.498	0.136	5.88	1.177	2.433	0.120	2.547	**0.012***
	No	5.40	1.522					5.33	1.493					5.27	1.413				
Regular Rec.	Yes	6.34	0.905	7.917	0.005	3.141	**0.003****	5.74	1.285	0.227	0.634	1.531	0.127	5.88	1.180	2.405	0.122	1.998	**0.047***
	No	5.67	1.365					5.40	1.372					5.47	1.358				

On the other hand, we examined the correlation between users' usage frequency and preference of keyword search and browsing with their demands for different OGD attributes. According to the Pearson correlation scores, in Table 4.15, we found that generally, there were significant positive correlations between citizens' need of OGD attributes and their preferences and usage frequency of keyword search and browsing. Moreover, the correlations were stronger with their preferences of data-acquisition methods than with their usage frequencies.

Table 4-15 Pearson correlation analysis of data acquisition and OGD attributes

		Download in bulk		Free of charge		Up to date	
		r	P	r	P	r	P
Keyword Search	Frequency	0.309	**0.000****	0.371	**0.000****	0.370	**0.000****
	Preference	0.430	**0.000****	0.530	**0.000****	0.505	**0.000****
Browse	Frequency	0.202	**0.008****	0.249	**0.001****	0.308	**0.000****
	Preference	0.240	**0.000****	0.349	**0.000****	0.351	**0.000****
		Update regularly		**Complete Metadata**		**Machine-readable**	
		r	P	r	P	r	P
Keyword Search	Frequency	0.368	**0.000****	0.306	**0.000****	0.292	**0.000****
	Preference	0.511	**0.000****	0.410	**0.000****	0.425	**0.000****
Browse	Frequency	0.326	**0.000****	0.257	**0.001****	0.321	**0.000****
	Preference	0.352	**0.000****	0.284	**0.000****	0.311	**0.000****

*r stands for Pearson Correlation, P stands for 2-tailed significant

Finally, we analyzed the relationship between help functions and users' demand of data quality. The Pearson correlation scores, in Table 4.16, show that users' needs for help had a significant positive relation with their needs for all the data attributes considered.

Table 4-16 Pearson correlation analysis of need of help and OGD attributes

Download in bulk	Free of charge	Up to date	Update regularly	Complete Metadata	Machine-readable
r=0.399*	r=0.452*	r=0.433*	r=0.443*	r=0.405*	r=0.405*

r = Pearson Correlation, 2-tailed, *significance p<0.05

All the above results support our third hypothesis (H3) that citizens' OGD utilization is influenced by their OGD demands.

In summary, based on the above data analysis of the survey, our three hypotheses are supported. We further refined the research model to show the specific relationships among different variables, as shown in Figure 4.10.

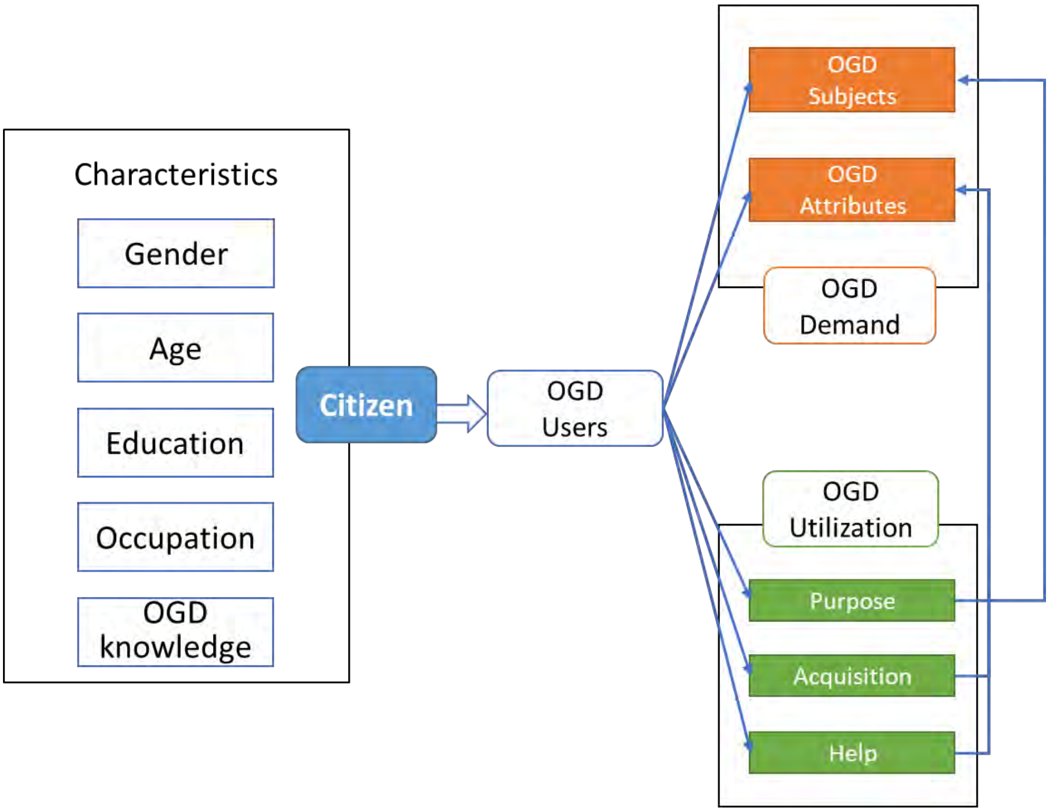


Figure 4-10 The refined research model

4.6 Conclusion

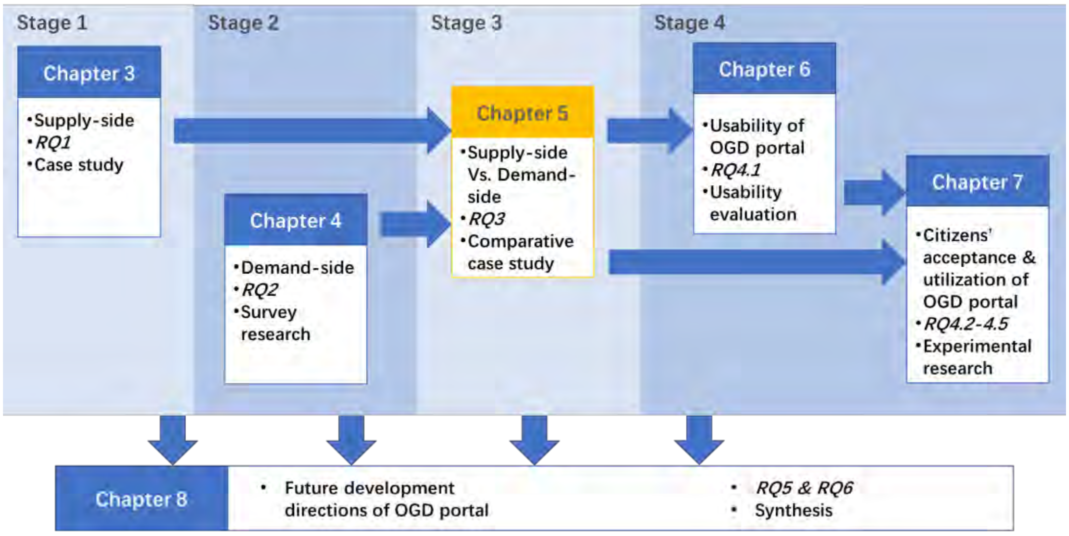
OGD portals are the most commonly implemented approach for the publishing and consuming of OGD by citizens (Attard et al., 2015), and users of OGD portals are consumers of OGD (Safarov et al., 2017) who have a direct effect on OGD utilization (Zuiderwijk & Janssen, 2013). Thus, looking into the interaction between users and OGD portals could help improve OGD utilization as well as the benefits derived from its usage.

In order to analyze the interaction between OGD users, their OGD demands and OGD utilization, we carried out this study from the perspective of citizens representing OGD users, covering three aspects including the characteristics of citizens, their OGD demands and their OGD utilization, as well as the relationships among these three. We designed and carried out an online survey among Chinese citizens to collect the data for analyzing and drew the following key conclusions:

- Few users have ever known of or used OGD or portals set up by governments. Users' occupation and education have an effect on their knowledge of OGD.
- Users show different degrees of demand for different OGD subjects. Their age, occupation and education affect their preferences for certain data subjects, while gender shows no effect on their preferences.
- Citizens choose to use OGD mainly to access data concerning daily life and anti-corruption, which are in line with the main aims of OGD programs. Their purposes of utilization have associations with their education and knowledge of OGD. Gender, age and occupation show no significant effects on the purposes of OGD utilization.
- Different purposes of utilization could lead to differences in the demand for OGD subjects.
- Users of OGD portals prefer keyword search most of four data-acquisition methods but browsing has the highest and most stable usage frequency.
- Females are more likely to accept regular recommendations of data/datasets than males.
- Users' age, education background and occupation affect their demands for different OGD attributes.
- Users' demands for OGD attributes are positively related to their preferences and usage frequency of data-acquisition methods.
- Users' needs for help are positively related with their feelings of difficulty in getting data from OGD portals.
- Users' needs for help are positively related to their demands for different OGD attributes.

Based on the above conclusions drawn from the data analysis and comparison with related studies, we recommend governments make efforts to improve public awareness of OGD programs and portals. Only if citizens know about OGD can they think about using it. Governments should also improve the quality of datasets and disclose more data in subject areas that citizens have shown greatest need of and interest in, such as health and education. Additionally, providing OGD according to users' specific scenarios in visiting OGD portals to meet their specific data demands is a possible way to improve OGD utilization. Since our study shows some citizens choose not to use OGD even if they are aware of its existence, researchers could carry out further study of citizens' motivations and needs concerning OGD and OGD portals. Future development of OGD portals could offer better help functions to reduce data-acquisition difficulties. Improving data quality, especially keeping data on portals up to date, could also help reduce the difficulty of data access from the portal.

5 CHAPTER 5: A COMPARISON OF THE SUPPLY-SIDE AND DEMAND-SIDE OF OGD PORTALS



5.1 Introduction

One of the main aims of disclosing government data is for its use, reuse and distribution (Attard et al., 2015). In Chapter 3, we found a fast development of OGD and OGD portals in China. However, in contrast to the continual development on OGD's supply-side, several studies have pointed out lack of use as a key problem for OGD (Ruijer, Grimmelikhuijsen, & Meijer, 2017; Safarov et al., 2017; Zuiderwijk & Janssen, 2013). But the reasons for OGD not being used are still unknown (Safarov et al., 2017).

Researchers believe that understanding the decisive factors in OGD utilization is important for stimulating OGD usage (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017; H.-J. Wang & Lo, 2016). A review of the literature revealed that although related studies have been carried out, they analyze OGD utilization either from the supply-side (data and portals) or the demand-side (users) (A. Meijer et al., 2014). Furthermore, too few studies have considered the demand-side of OGD, placing an emphasis on data and its access (Evans & Campos, 2013; Ohemeng & Ofosu-Adarkwa, 2015) and failing to connect the two sides of OGD. No studies have been carried out to compare the two sides of OGD or to analyze the relationship between these two sides. Therefore, what is the present relationship between supply-side and demand-side of OGD becomes the key concern of this chapter. Moreover, considering the problem of low utilization of the opened data (Safarov et al., 2017), a possible mismatch may exist between these two sides which affected the acceptance and utilization of OGD.

Against this backdrop and recognized gaps, our objective of this chapter is thus to figure out the possible relationships between the supply-side and demand-side of OGD, and to apply our findings to the development of OGD portals by eliminate the possible mismatch from the supply-side. Thus, we raised the following research question for this chapter:

RQ3. What is the relationship between the supply-side and demand-side of OGD?

We further divided RQ3 into two sequential sub-questions:

RQ3.1 What research model could be used for the comparison of the supply-side and demand-side of OGD?

RQ3.2 How to analyze the present relationship between the supply-side and demand-side of OGD by using the research model?

To answer the first sub-question RQ3.1, an extended research model based on the Diffusion of Innovation theory (DOI) was derived by drawing on prior studies of OGD's supply-side and demand-side. To test the capability of this research model in identifying the possible relationship between the two sides as well as to address the second sub-question RQ3.2, a comparative case study of the OGD development in China was carried out by applying the extended research model. Data collected in the case study were analyzed to determine the concord and conflict of the two sides of OGD.

The main contributions of this chapter are as follows: Firstly, by synthesizing the present studies and DOI, a multidimensional model is constructed to guide the comparison of the supply-side and demand-side of OGD. Secondly, by applying the developed model to a comparative case study in China, we not only demonstrate the capability of the model but also derive a series of developmental prescriptions for OGD portals. Thirdly, the case study helps to show the present relationship between two sides of OGD in China, which may shed light on the relationship of OGD's two sides in other geographical areas. Lastly, based on DOI, this study deepens the understanding of citizens' acceptance of OGD.

5.2 Research model

In this section, we develop our research model, shown in Figure 5.1, with further descriptions and support references shown in Table 5.1. The research model is derived from the DOI theory (column 1 of Table 5.1) and related literature (column 4 of Table 5.1) and serves to connect the supply and demand sides of OGD by drawing the factors from Chapters 3 and 4 together (columns 2 and 3 of Table 5.1, respectively). DOI includes five characteristics of innovation: observability; relative advantage; compatibility, complexity and trialability. We include the first four characteristics in our research model. In DOI, trialability is defined as "the degree to which an innovation may be experimented with on a limited basis" (Rogers, 2010, p. 16), and is suited to analysis of a product or service that is available for users to try over a limited time period. Since our study did not include an experimental or trial period of OGD usage by users, and OGD is not offered to the public temporarily according to its definition, trialability has been omitted from this study. The other four aspects of DOI in our research model are explained below in detail.

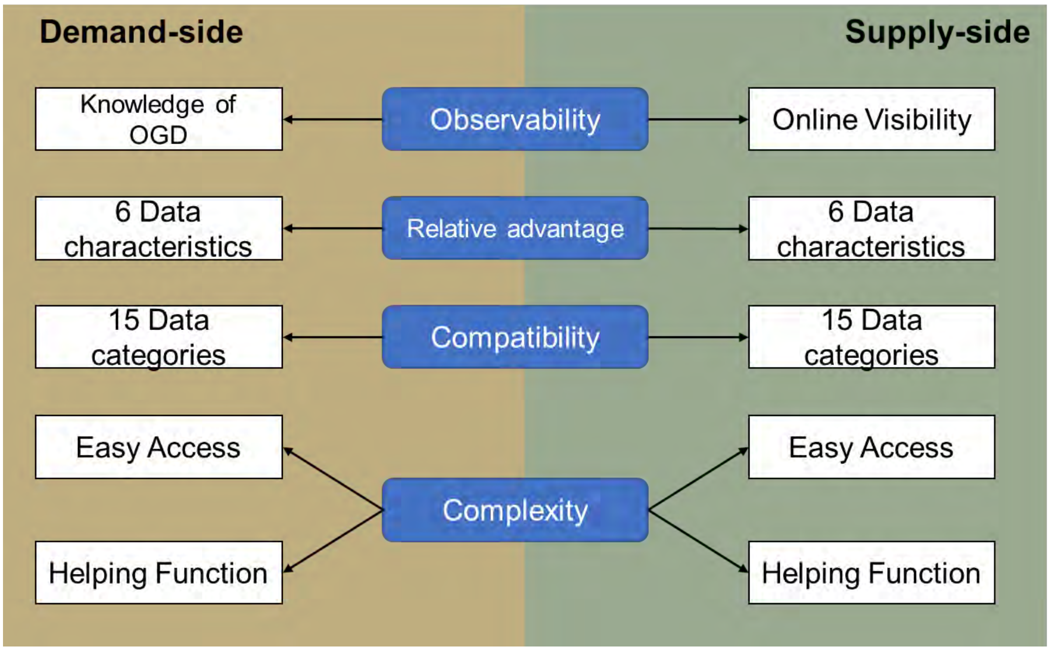


Figure 5-1 Modified DOI research model for connecting OGD demand and supply

Table 5-1 Comparison model of OGD's supply-side and demand-side

Aspects	Description	Supply-side Factor	Demand-side Factor	Reference
Observability	Citizens' awareness of OGD portal	Online visibility	Ever Known OGD portal	Gao & Vaughan (2005); Thelwall (2002); Qiu (2010); Weerakkody et al (2017); Yi & Jin (2008)
Relative advantage	OGD portal's benefits in comparison to physical offices and other platforms	Download in bulk	Download in bulk	Sunlight Foundation (2010); Open Data Index (2017)
		Free of charge	Free of charge	Open Data Index (2017); Public Data Principles (2012); OpenDataBarometer (2016)
		Up to date	Up to date	OpenDataBarometer (2016); OpenGovernmentWorkingGroup (2007); Lourenço (2015)
		Update regularly	Update regularly	Sunlight Foundation (2010); OpenDataBarometer (2016); Craveiro et al (2013)
		Complete metadata	Complete metadata	Sunlight Foundation (2010); Bogdanović-Dinić et al. (2014); OpenDataMonitor (2017)
		Machine-readable data	Machine-readable data	OpenGovernmentWorkingGroup (2007); Sunlight Foundation (2010); Open Data Index (2017)
Compatibility	Types of information citizens are expecting to access through OGD portals	15 types of information offered by the portal	Citizens' attitudes towards 15 types of information	Bogdanović-Dinić et al. (2014); OpenDataBarometer (2016); Open Data Index (2017); Weerakkody et al (2017)
Complexity	The complex of use of OGD portal	OGD easy access	Demand of OGD easy access	Bogdanović-Dinić et al. (2014); Craveiro et al (2013); Open Data Index (2017)
		Help functions offered	Demand of help functions	Grayling (2002); Weerakkody et al (2017)

5.2.1 Observability

Rogers (2010) (p16) defined observability as "the degree to which the results of an innovation are visible to others". Applying this characteristic to OGD, we use it to identify citizens' awareness of OGD portals, because OGD portals are the main channel for citizens to access OGD (Lourenço, 2015). For the supply-side of OGD, scholars commonly apply a webometric approach to measure the impact, visibility and connectivity of websites (Smith & Thelwall, 2002; Thelwall, 2002; K. Yi & Jin, 2008) by counting the hyperlinks, since they are essential for accessing digital resources on the web (Bar-Ilan, 2001). The online visibility of a website means its possibility of discovery by other websites (Qiu, 2010). Because in webometrics the hyperlinks pointing to and being pointed to are treated as analogous to referring to and being referred to in bibliometrics and citation analysis (Ingwersen, 1998), thus the number of inlinks (a link pointing into a site (Björneborn & Ingwersen, 2004)) is used as an indicator of the visibility of the web contents (Chu, He, & Thelwall, 2002; Gao & Vaughan, 2005; Vreeland, 2000; Wormell, 2001). In our comparison model, we accept the concept of online visibility of a website as the indicator for the observability of the supply-side of OGD. For the demand-side of OGD, citizens' awareness of the existence of OGD portals is examined to indicate the observability of OGD on the demand-side.

5.2.2 Relative advantage

Rogers (2010) (p15) defined relative advantage as "the degree to which an innovation is perceived as better than the idea it supersedes". According to its definition, the relative advantage of OGD could be the delivery of greater benefits for citizens when comparing the benefits delivered via other physical offices and portals that citizens have access to (Weerakkody et al., 2017). Some characteristics of open data are emphasized by researchers when carrying out evaluations of OGD portals (Bogdanović-Dinić et al., 2014; Craveiro et al., 2013;

Lourenço, 2015; OpenDataBarometer, 2016; OpenDataIndex, 2017; OpenDataMonitor, 2017; OpenGovernmentWorkingGroup, 2007; PublicSectorTransparencyBoard, 2012; SunlightFoundation, 2010), characteristics which are treated as a guarantee of OGD's benefits (Dawes, 2010). Thus, to make a comparison of the demand-side and supply-side, we selected the six most commonly accepted characteristics of OGD for both sides: "Download in bulk", "Free of charge", "Up to date", "Update regularly", "Complete metadata" and "Machine-readable". The selection of OGD characteristics was based on an analysis of existing guidelines and principles as listed in the reference column of Table 5.1. Previous evaluations of OGD portals in Chapter 3 Study 1 listed in Table 5.1 set clear assessment criteria for each of the selected characteristics. Definitions and assessment criteria for each characteristic that we adopted from previous assessments in Chapter 3 Study 1 are listed in Table 5.2. For the demand-side, these six characteristics are easy to observe by users during the utilization of OGD, thus it is possible to evaluate the extent to which citizens need each characteristic.

Table 5-2 Open government data characteristics

Element	Description	Assessment Criteria for supply-side	Reference
Download in bulk	Possibility to download all the needed data stored in a database at once.	Can the data/dataset on the portal be downloaded in bulk?	Sunlight Foundation. (2010); Open Data Index. (2017)
Free of charge	Don't have to pay money for the data/dataset.	Is the data free of charge for downloading?	Open Data Index. (2017); Public Data Principles. (2012); Open Data Barometer. (2016);
Up to date	Data/dataset is made available as quickly as necessary to preserve the value of the data.	If the data is updated regularly, does the data have the latest version? If not updated regularly, has the data been updated in a week?	Open Data Barometer. (2016); Open Government Working Group. (2007); Lourenço, (2015)
Update regularly	To maintain the sustainable utilization of a data/dataset, the data/dataset should be kept regularly updated.	Does the data have an update plan?	Sunlight Foundation. (2010); Open Data Barometer. (2016); Craveiro et al. (2013)
Complete Metadata	Data/dataset has metadata that defines and explains what is provided.	Does the data/dataset provide metadata that defines and explains the provided contents?	Sunlight Foundation. (2010); Open Government Working Group. (2007);
Machine-readable	Data/dataset is stored in widely-used file formats that easily lend themselves to machine processing, accessing and modifying single elements in the file.	Can the data/dataset be opened in appropriate data manipulation and analysis software?	Open Government Working Group. (2007); Sunlight Foundation. (2010); Open Data Index. (2017)

5.2.3 Compatibility

Rogers (2010) (p15) defined compatibility as "the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters". It can be used to predict users' intention to use OGD (Putzer & Park, 2010). OGD is believed to bring citizens various kinds of information that they have interest in but previously had no access to (Weerakkody et al., 2017), therefore it is supposed to meet the expectations of citizens including potential adopters and the existing values of transparency and citizens' right to know. But scholars have questioned whether such data can really cater to the needs of individuals (Behkamal, Kahani, Bagheri, & Jeremic, 2014). Thus, we want to know if OGD portals are offering citizens the types of information they are interested in. That is, does the OGD portal contain the content that citizens desire? To make this comparison between the supply-side and demand-side, we selected the same content categories of OGD for both sides. The selection was based on an initial list of 46 categories of data obtained from related studies and OGD portals. After reducing the overlaps among categories, 15 categories were left, listed in Table 5.3. On the supply-side, the number of datasets of each category offered on OGD portals was calculated. On the demand-side, citizens' demands for each category was evaluated.

Table 5-3 OGD categories for the comparison of compatibility

Category	Description
Budget & spend	The budget and spend records of the local government.
Credit records	Credit records of companies and organizations recorded by the local government.
Cultural activity	Data related to cultural and leisure activities and services.
Education	Data related to education system and services.
Environment quality	Data related to environment, including air and water quality, forestation rate, etc.
Government bid	Procurement data of the local government.
Health	Data related to the healthcare system and health services.
International trade	Details of the import and export of specific commodities and/or balance of trade data.
Local statistics	Key statistics in this area, such as demographic and economic indicators (GDP, unemployment, population, etc.).
Map	Detailed digital map of this area.
Policies & legislation	Policies and legislations proposed in this area.
Public safety	Data related to public safety, including crime statistics, food safety, daily products safety.
Registration	Data of companies and organizations registration, including name, unique identifier and other additional information.
Transportation	Details of public transport services, including the construction of public transports.
Weather	Data related to weather and climate records and forecasts.

5.2.4 Complexity

Complexity was defined by Rogers (2010) (p16) as "the degree to which an innovation is perceived as difficult to understand and use". The less complex an innovation is to use/operate, the more easily it is accepted (Weerakkody et al., 2017). Researchers have noted existing difficulties in accessing open data (Kassen, 2013) and the unfriendly design of the interfaces (C. Martin, 2014), which lead to fewer users. Since user-centred design has become one of the key methods in human–computer interaction, it is commonly accepted that user interfaces should be made as easy and pleasant to use as possible (Dix, 2009). For OGD, scholars have emphasized the importance of data being obtainable with no requirement for registration or any other procedures to download or read the data, to ensure easy access (OpenDataIndex, 2017; SunlightFoundation, 2010). Thus, we include OGD easy access as a comparative factor for both sides of OGD.

Generally, all websites include certain forms of help system to reduce users' difficulties (Grayling, 2002). But users' attitudes towards different help functions vary (Grayling, 1998), which also leads to different effects of the help functions on users. Typical help functions used by OGD portals include a user guide webpage, FAQ webpage, online smart/virtual agent to offer instant conversation and customer phone support. Help functions provided by the supply-side and users' demands towards these help functions have been compared.

5.3 Methodology

To address the core research question of this study and to justify the capability of the research model, an in-depth case study was carried out to make a comparison between the demand-side and supply-side of OGD. The case study method enables researchers to closely explore and investigate contemporary real-life phenomena through detailed analysis (Zainal, 2007). Given our objective to make comparison between both sides of OGD, we chose a multiple-case study as our research design, because multiple-case studies can provide stronger evidence of the robustness and generalizability of the results compared with single-case studies, due to their retaining of elements that are replicated across most of the cases (Eisenhardt & Graebner, 2007; Ellonen, Wikström, & Jantunen, 2009).

5.3.1 Case study in China

We selected the region of China to carry out the case study for four reasons. Firstly, encouraged by the central government (StateCouncil, 2016), OGD has developed at a fast rate in China in recent years and many local OGD portals have been built (R. Huang & Wang, 2016). Thus, it is possible to carry out OGD case studies in China. Secondly, China is the largest developing country, with sub-areas of different economic and political contexts for the development and utilization of OGD. Its diversity in OGD development benefited the selection of cases for both the supply-side and demand-side, and at the same time strengthened the robustness and generalization of the results. Thirdly, the OGD development in many regions in China is less investigated in the literature, such as in Shanghai, Guangzhou and Shandong. Thus, our study focused on these regions that are novel in the literature. Finally, as we identified from the present literature, until now no case studies of comparing the supply-side and demand-side of OGD have been carried out in China. Thus, our study filled this research gap.

5.3.2 Construct operationalization

For the supply-side of the case study, we adhered to the general procedure offered by Lourenço (2013) to carry out the study on OGD portals in China. Firstly, we adopted the term "portal" to refer to the official web portal launched by local governments (Attard et al., 2015). A portal had to have a structured domain with ".gov" and contain data from different departments of the local government or institutions managed by the local government. Secondly, province-level OGD portals in China were selected as multiple cases because right now there is no one united country-level portal in China for all open data from different levels of government, while the number of existing province-level portals in China permits a reasonably sized and exhaustive study. Also, portals at the same political level have similar contexts and characteristics. Province-level areas are better to study than lower-level ones like OGD portals for cities because they are more capable of providing the resources and technology needed for the development of OGD portals (Chatfield & Reddick, 2017), thus they are information-rich cases (Ellonen et al., 2009). Twelve province-level OGD portals were identified from the related literature (DMGLabFudanUniversity, 2018; D. Wang, Chen, et al., 2018) and online searching, as shown in Table 5.4. Thirdly, the data categories to be used in the data collection procedure were selected from related research (see Table 5.3). Finally, personal visits to each of the OGD portals were carried out for collecting sample datasets. Evaluations of each dataset were then implemented according to the assessment criteria in Table 5.2.

Table 5-4 Description of cases for supply-side

Place	Population (Ths)	Regional GDP in RMB (Mio)	General budget revenue in RMB (Mio)	OGD Portal URL
Beijing	21,730	25,669.13	5,081.26	http://www.bjdata.gov.cn
Guangdong	109,990	80,854.91	10,390.35	http://www.gddata.gov.cn
Guizhou	35,550	11,776.73	1,561.34	http://www.gzdata.gov.cn
Hong Kong	7,377	5,469.96	3,503.72	https://data.gov.hk
Jiangxi	45,920	18,499.00	2,151.47	http://data.jiangxi.gov.cn
Ningxia	6,750	3,168.59	387.66	http://ningxiadata.gov.cn/odweb
Shandong	99,470	68,024.49	63,002.33	http://data.sd.gov.cn
Shanghai	24,200	28,178.65	6,406.13	http://www.datashanghai.gov.cn
Sichuan	82,620	32,934.54	3,388.85	http://www.sc.gov.cn/10462/13797/index.shtml
Taiwan	23,540	33,328.55	5,780.29	http://data.gov.tw
Xinjiang	23,980	9,649.70	1,298.95	http://data.xinjiang.gov.cn
Zhejiang	55,900	47,251.36	5,301.98	http://data.zjzwfw.gov.cn

For the demand-side of the case study, a five-part survey was designed based on our research model for data collection. The first part contained questions reflecting citizens' socio-demographic characteristics that may affect their usage of OGD portals (Venkatesh et al., 2014), including their gender, age, educational background and occupation. The second part was citizens' awareness of OGD portals with questions asking whether the participant had ever known or heard of OGD portals before. The third part included questions asking citizens' attitudes towards OGD characteristics. The fourth part asked about participants' demands for 15 kinds of OGD content categories. A 7-point Likert scale (from "Strongly don't need" to "Strongly need") was used for both these parts. The last part included two sets of questions. One set asked the participants' attitudes towards ease of access of OGD by selecting from a 7-point Likert scale (from "Strongly don't need" to "Strongly need"). The other asked the participants to select as many help functions as they preferred.

An instrument validation was carried out following the procedures described by (DeVellis, 2016). After drawing the instrument elements from related literature (see Table 5.1), the survey was then examined by two experts for minor modifications. Then a pilot test of the questions was administered to 10 graduate and PhD students studying in the field of information systems to test the validity and reliability of the instrument. Slight modifications were made accordingly.

5.3.3 Data collection

For the supply-side, the data was collected by using the same process described in Chapter 3. This was conducted in October 2018 to include the most current dataset by using three different methods. Data was collected objectively, aiming at reducing possible bias of the researchers that might influence the final results. Firstly, for observability, we used the search engine Baidu with the search query "domain: URL" and Google with the search query "link: URL" to get the number of inlinks collected by each search engine. Secondly, for relative advantage, compatibility and OGD easy access in complexity, data was collected through random selection of dataset samples from the 12 OGD portal cases. On each OGD portal, keyword searches for each category listed in Table 5.3 were carried out. Then 60 datasets were randomly selected from the results. If the search results for a category did not have enough datasets (< 60), we then included all the searched datasets with contents that fit the description of that category. Each selected dataset was evaluated according to the listed criteria in Table 5.2 and marked as to whether it could be downloaded without any procedure like registration or application (OGD easy access). If the dataset met each criterion, we marked a value of 1 for each element, otherwise we marked a value of 0. Thirdly, for help functions offered by OGD portals, we manually visited each of the portal cases to mark whether the four kinds of help functions (user guide webpage, FAQ webpage, online smart/virtual agent and customer phone support) were provided or not. We marked a value of 1 for each help function that was offered and a value of 0 if it was not offered.

For the demand-side, the data was collected during the data collection process described in Chapter 4. An online tool called Sojump was used to create an online questionnaire and to distribute it among Chinese citizens from 1 to 10 August 2017. Participants were randomly recruited through the top two social media sites in China, called WeChat and Weibo. We chose citizens to represent OGD users and to be the objects for the supply-side of case studies for two reasons. On one hand, the key motivation for releasing government data to the public is to reduce the asymmetry of information between citizens and government bodies (Murillo, 2015; Ubaldi, 2013), thus citizens are major beneficiaries of OGD utilization and are identified as primary stakeholders (Parycek et al., 2014). On the other hand, due to the diversity of citizen stakeholders, choosing them as objects offered the advantages of wider representation of the population and a range of diverse characteristics.

5.4 Data analysis and results

In this section, we explain in detail the process of data analysis including data pre-processing, reliability and adequacy tests, general description analysis and parametric tests, and show the results for each step.

5.4.1 Data pre-processing

Because data in the demand-side for relative advantage, compatibility and OGD easy access used a 7-point Likert scale, to ensure the values on the supply-side for these three aspects are comparable, the data collected on the supply-side was normalized prior to further analysis. We first calculated the number of collected sample datasets from each OGD portal case (D_i, i = [1,12]). For each element in relative advantage, we added the value for each of the OGD portal cases (R_{ij}, j = [1, 6]). Then by using the formula $R_{ij}' = 7 * R_{ij} / D_i$ (i= [1,12], j= [1,6]) we calculated the value for each element in relative advantage for each OGD portal case after the normalization step. We repeated a similar step to calculate the value for OGD easy access in complexity. For each data category in compatibility, we added the number of datasets for each of the OGD portal cases (C_{ik}, k= [1,15]). Then by using the formula $C_{ik}' = 7 * C_{ik} / 60$ (i= [1,12], k= [1,15]) we derived the value for each data category in compatibility for each of the OGD portal cases after the normalization.

5.4.2 Reliability and adequacy tests

We examined the reliability of all the scales in our study with Cronbach's alpha (Cronbach, 1951). Results are shown in Table 5.5. Because observability and OGD easy access in complexity have only one variable, it is not appropriate to calculate Cronbach's alpha for these two factors. The commonly accepted range for Cronbach's alpha is 0.70 to 0.95 (Tavakol & Dennick, 2011). The results for relative advantage and compatibility thus show high reliability with scores of 0.88 and 0.95, respectively. The alpha for complexity is below 0.7, which may be due to the low number of variables (Tavakol & Dennick, 2011). But the overall alpha for the whole model including all variables is 0.93, thus we continued our analysis. We also examined all the scales with the Kaiser–Meyer–Olkin (Bickmore & Cassell) measure of sampling adequacy (Dziuban & Shirkey, 1974). All the results are above 0.6, indicating the variables were suitable for factor analysis (Kaiser, 1970). The significance of Bartlett's test of sphericity is less than 0.05, which also indicates the high validity of the model.

Table 5-5 Reliability and adequacy tests of the model

	Aspects	Variable No.	Valid	%	Cronbach's Alpha
Reliability	Relative Advantage	6	220	100	0.88
	Compatibility	15	220	100	0.95
	Complexity	4	220	100	0.44
	Overall	27	220	100	0.93
Adequacy	Aspects	Bartlett's test of sphericity			Kaiser-Meyer-Olkin measure of sampling adequacy
		Approx. Chi-square df	df	Sig.	
	Relative Advantage	818.33	15	0	0.80
	Compatibility	3003.48	105	0	0.94
	Complexity	36.02	6	0	0.61
	Overall	4315.83	325	0	0.92

5.4.3 Observability

Results for online visibility of the supply-side are shown in Table 5.6. The number of inlinks of different OGD portals varies. A gap was also noted between the two search engines. Some portals like those of Guangdong and Sichuan might have blocked search engines from collecting their data. Since measuring portal visibility relies on a search engine for accessing the data, blocking of search engines may lead to a relatively low rate of visibility for that portal and cause differences between different search engines. For the demand-side, results show that 27.4% of the participants knew about OGD portals.

Table 5-6 Online visibility of OGD portals

Portal	Beijing	Guangdong	Guizhou	Hong Kong	Jiangxi	Ningxia
Inlinks Collected by Baidu	2293	10904	531	1064	65	210
Inlinks Collected by Google	2430	1	36	9200	330	219
Portal	Shandong	Shanghai	Sichuan	Taiwan	Xinjiang	Zhejiang
Inlinks Collected by Baidu	8159	6128	0	20	37	988
Inlinks Collected by Google	1000	9560	1	139000	907	1200

5.4.4 Relative advantage

Results for relative advantage are shown in Table 5.7. On the supply-side, "Free of charge" received the highest score and had the smallest standard deviation, indicating that it offered the best performance aspect of OGD portals, while "Up to date" received the lowest score. On the demand-side, all variables had an average score above 5 out of the 7-point scale, which shows citizens' positive needs for these data characteristics. "Up to date" received the highest score and the smallest standard deviation, indicating it to be the top focus of citizens, while "Download in bulk" received the lowest score concerning needs but had the highest standard deviation, which means citizens had different opinions about this kind of data quality. Two independent sample T-tests were carried out to analyze the relationship between the supply-side and demand-side of OGD in relative advantage. Results, in Table 5.7, show that significant differences between the two sides lie in "Free of charge", "Up to date", "Update regularly" and "Complete metadata", with $p < 0.05$. For "Free of charge" and "Complete metadata", the scores for the supply-side are higher than the demand scores of the citizens. For "Up to date" and "Update regularly", the scores of the supply-side are much lower than for the demand-side.

Table 5-7 Comparative results for supply-side and demand-side in relative advantage

Relative Advantage	Side	Avg	Std	F	Sig.	t	Sig.
Download in bulk	Supply	4.94	2.96	33.06	0.00	-0.62	0.55
	Demand	5.48	1.40				
Free of charge	Supply	6.83	0.37	7.14	0.01	4.94	0.00**
	Demand	6.17	1.19				
Up to date	Supply	2.65	2.32	32.89	0.00	-5.29	0.00**
	Demand	6.21	1.01				
Update regularly	Supply	3.46	2.53	36.41	0.00	-3.73	0.00**
	Demand	6.20	1.06				
Complete metadata	Supply	6.81	0.43	13.29	0.00	7.47	0.00**
	Demand	5.66	1.31				
Machine-readable data	Supply	5.21	2.93	36.92	0.00	-0.68	0.51
	Demand	5.79	1.23				

5.4.5 Compatibility

Results for compatibility are shown in Table 5.8. On the supply-side, in total 4196 dataset samples were selected from the 12 OGD portals. Referring to the number of datasets for different categories, we noted obvious differences among categories, with “Education” and “Local statistics” being the most popular categories that OGD portals generally released to the public, and “Policies & legislation” the least popular one. On the demand-side, generally, average scores for citizens’ needs of OGD were all above 5 on the 7-point scale, indicating their high demands for OGD. The largest variation in respondents’ scores of needs lay in general need. “Health” and “Education” received the highest average scores. “Education” also got the lowest standard deviation, indicating citizens agreed most on their high demand of this category. “International trade” got the lowest average score. Two independent sample T-tests were carried out to analyze the relationship between the supply-side and demand-side of OGD in compatibility. Results show that differences in the average scores for “Education”, “Cultural activity” and “Local statistics” were not statistically significant between the supply-side and demand-side of OGD, while for all other data categories there were statistically significant differences between the two sides. The scores for the supply-side were much lower than for the demand-side. The greatest difference lay in "Policies & legislation".

Table 5-8 Comparative results for supply-side and demand-side in compatibility

Data type	Side	Avg	Std	F	Sig.	t	Sig.
Health	Supply	3.80	2.98	57.10	0.00	-2.83	**0.016***
	Demand	6.25	1.16				
Education	Supply	4.48	2.98	104.15	0.00	-2.06	0.064
	Demand	6.25	1.01				
Cultural Activity	Supply	4.09	2.74	27.54	0.00	-1.87	0.088
	Demand	5.58	1.30				
Local Statistics	Supply	4.46	3.19	101.86	0.00	-1.90	0.083
	Demand	6.22	1.06				
Transportation	Supply	3.53	2.84	40.48	0.00	-3.21	**0.008****
	Demand	6.17	1.12				
Map	Supply	1.54	2.24	5.38	0.02	-6.41	**0.000****
	Demand	5.71	1.25				
Public Safety	Supply	3.18	3.02	62.69	0.00	-3.47	**0.005****
	Demand	6.22	1.10				
Policies& Legislation	Supply	1.06	2.14	7.74	0.01	-8.01	**0.000****
	Demand	6.05	1.16				
Weather	Supply	1.97	2.57	13.96	0.00	-6.01	**0.000****
	Demand	5.85	1.17				
Environment Quality	Supply	3.33	3.07	95.72	0.00	-3.16	**0.009****
	Demand	6.15	1.06				
Registration	Supply	3.29	2.68	12.66	0.00	-2.79	**0.017***
	Demand	5.46	1.50				
Credit Records	Supply	2.17	2.48	14.52	0.00	-5.31	**0.000****
	Demand	6.00	1.22				
International Trade	Supply	1.23	1.92	0.22	0.64	-8.31	**0.000****
	Demand	5.13	1.56				
Budget & Spend	Supply	1.49	2.61	7.02	0.01	-9.22	**0.000****
	Demand	5.56	1.40				
Government Bid	Supply	1.77	2.59	4.68	0.03	-6.80	**0.000****
	Demand	5.19	1.63				

5.4.6 Complexity

Results for complexity are shown in Table 5.9. For OGD easy access, there were no statistically significant differences in the average scores for the supply-side and demand-side. The average score for the supply-side was higher than for the demand-side. For help functions, on the supply-side "Customer phone support" received the highest average score, while "Online smart agent" received the lowest average score and standard deviation. Three out of the 9 portals offered three different kinds of help. Those of Ningxia, Sichuan and Xinjiang did not offer help for customers at all. On the demand-side, differences among the average scores and standard deviations for the four help functions were very small. “FAQ” received the highest score, while “Customer phone support” received the lowest score. Two independent sample T-tests were carried out to analyze the relationship between the supply-side and demand-side of OGD in complexity. There were statistically significant differences in the average scores for “User guide”, “FAQ” and “Online smart agent”, where the scores for the demand-side were much higher than for the supply-side. The greatest difference lay in “Online smart agent”.

Table 5-9 Comparative results for supply-side and demand-side in complexity

Complexity		Side	Avg	Std	F	Sig.	t	Sig.
OGD easy access		Supply	4.49	3.33	14.58	0.00	0.44	0.665
		Demand	4.06	2.06				
Helping Functions	User guide	Supply	0.42	0.51	1.58	0.21	-2.08	**0.039***
		Demand	0.70	0.46				
	FAQ	Supply	0.42	0.51	2.74	0.10	-2.46	**0.015***
		Demand	0.74	0.44				
	Online smart agent	Supply	0.08	0.29	46.79	0.00	-6.31	**0.000****
		Demand	0.65	0.48				
	Customer phone support	Supply	0.67	0.49	0.33	0.57	0.26	0.80
		Demand	0.63	0.48				

5.5 Discussion

Theoretical and pragmatic implications can be drawn from the comparative case study of the supply-side and demand-side of OGD in China. In this section, we relate our empirical findings to extant literature in OGD research, and then summarize the theoretical contributions and pragmatic implications of this study.

5.5.1 Discussion of analytical findings

In relation to observability, the findings from our survey of the demand-side are in line with the present literature identifying lack of use of OGD by citizens as a serious problem compared with the disclosure of OGD (Heise & Naumann, 2012; Kuhn, 2011; Ruijer, Grimmelikhuijsen, & Meijer, 2017; Zuiderwijk & Janssen, 2013). Citizens showed little awareness of OGD. Based on this result, we further analyzed whether citizens' awareness of OGD portals had an effect on their demands for OGD characteristics and categories. According to the two independent sample T-test results in Table 5.10, citizens who had known about OGD portals had higher requirements for the data characteristics of "Download in bulk", "Free of charge" and "Machine-readable". Their demands for OGD in general were also higher than for those who had never heard of OGD before. On the supply-side, a correlation analysis of the number of datasets on each of the OGD portal cases and the portal's online visibility was carried out. Results have shown there to be a strong possibility (sig. (2-tailed) p <0.001) of a positive strong correlation between the two, with the Pearson correlation being 0.939. Thus, increasing the number of datasets on the portal could help increase the online visibility of an OGD portal. For help functions, few differences were found in the scores for different help functions. Hui and Png (2015) concluded that online and telephone services both had their own supporters.

Table 5-10 Correlations between observability, relative advantage and compatibility on demand-side

Aspect	Awareness of OGD Portal	Avg	Std	F	Sig.	t	Sig.
Download in bulk	Yes	5.77	1.12	8.74	0.00	2.13	**0.035***
	No	5.36	1.49				
Free of charge	Yes	6.46	0.80	3.57	0.06	2.16	**0.032***
	No	6.06	1.29				
Machine-readable data	Yes	6.07	1.03	4.69	0.03	2.26	**0.026***
	No	5.68	1.28				
OGD demands in general	Yes	5.68	1.39	3.07	0.08	2.88	**0.004****
	No	4.92	1.81				

We also analyzed the correlation of relative advantage and compatibility on both sides with Pearson's correlation analysis. Results are shown in Table 5.11. Significant correlations between relative advantage and compatibility were found on both the supply-side and demand-side of OGD. For the supply-side, significant strong correlations lay in the scores for "Download in bulk" and "Health", "Machine-readable" and "Education", and "Download in bulk" and "Environment quality”. For the demand-side, all pairs showed weak or moderately significant correlations. Although more pairs of relative advantage and compatibility on the demand-side showed statistically significant correlations, the correlations on the supply-side were statistically stronger than on the demand-side.

Due to the above discoveries, we derived the modified research model shown in Figure 5.2 to explain the relationships between the supply-side and demand-side of OGD based on DOI. Dots beside each box for both sides (i.e. demand and supply) show the results of comparison between the supply and demand sides for each aspect, with orange dots for significant differences and green ones for non-significant differences. “M” stands for a mixed result, “H” for higher than the other side and “L” for lower than the other side. For example, the orange “H” dot for compatibility on the demand-side indicates that the scores for citizens’ demands of the 15 data categories were significantly higher than the scores for the amount of data for the 15 categories on the supply-side. Ideally, a perfect model for the relationship between the supply-side and demand-side should have balanced scores or insignificant differences for all five boxes on each side. However, at present, significant conflicts between the two sides are evident for compatibility and complexity, revealing that the supply-side could not meet the expectations of citizens on the demand-side. For relative advantage, “Up to date” and “Update regularly” are the two characteristics where the supply-side could not meet the requirements of the demand-side, while “Free of charge” and “Complete metadata” are the two characteristics where the supply-side surpassed the expectations of the demand-side, which led to a mix result in this aspect. Arrows on the outside of boxes between the supply and demand sides show the correlations between different aspects of DOI. For the demand-side, significant relationships lie in observability and relative advantage, as well as observability and compatibility. For both supply and demand sides, significant relationships lie between relative advantage and compatibility.

Table 5-11 Correlation analysis of relative advantage and compatibility

Relative Advantage		*Correlation*	*Compatibility*														
			Health	Education	Cultural Activity	Local Statistics	Transpor-tation	Map	Public Safety	Policies & Legislation	Weather	Environment Quality	Registration	Credit Records	International Trade	Budget & Spend	Government Bid
Supply side	Download in bulk	Pearson's r	**0.621**	0.550	0.491	-0.176	0.481	0.275	0.475	0.004	0.319	**0.603**	0.451	0.444	0.346	-0.233	-0.234
		p-	**0.031***	0.064	0.105	0.585	0.114	0.387	0.119	0.991	0.313	**0.038***	0.142	0.148	0.271	0.467	0.465
	Free of charge	Pearson's r	0.436	0.531	0.215	0.068	0.27	0.146	0.340	-0.015	0.211	0.386	0.231	0.265	0.239	0.254	0.254
		p-	0.157	0.075	0.502	0.833	0.397	0.652	0.279	0.962	0.509	0.215	0.471	0.405	0.455	0.426	0.426
	Up to date	Pearson's r	0.401	0.422	0.427	0.405	0.4	0.515	0.470	0.348	0.506	0.221	0.523	0.344	0.427	-0.008	0.1
		p-	0.196	0.172	0.166	0.192	0.198	0.087	0.123	0.267	0.093	0.491	0.081	0.274	0.167	0.980	0.758
	Update regularly	Pearson's r	0.358	0.382	0.190	0.106	0.414	0.321	0.570	0.225	0.281	0.228	0.356	0.288	0.095	-0.161	-0.064
		p-	0.253	0.220	0.553	0.742	0.18	0.309	0.053	0.481	0.376	0.475	0.256	0.364	0.769	0.618	0.844
	Complete metadata	Pearson's r	0.441	0.545	0.227	0.102	0.275	0.143	0.353	-0.011	0.194	0.383	0.255	0.261	0.220	0.233	0.243
		p-	0.151	0.067	0.478	0.752	0.387	0.657	0.260	0.974	0.546	0.219	0.424	0.413	0.492	0.466	0.446
	Machine-readable data	Pearson's r	0.609	**0.616**	0.489	-0.131	0.473	0.256	0.487	0.083	0.279	0.534	0.434	0.397	0.277	-0.252	-0.228
		p-	0.035	**0.033***	0.107	0.685	0.120	0.423	0.108	0.798	0.381	0.074	0.159	0.201	0.384	0.429	0.477
Demand side	Download in bulk	Pearson's r	**0.336**	**0.297**	**0.259**	**0.391**	**0.353**	**0.334**	**0.257**	**0.266**	**0.262**	**0.331**	**0.303**	**0.371**	**0.342**	**0.336**	**0.293**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****
	Free of charge	Pearson's r	**0.427**	**0.388**	**0.374**	**0.527**	**0.475**	**0.406**	**0.430**	**0.431**	**0.378**	**0.449**	**0.291**	**0.468**	**0.344**	**0.366**	**0.279**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****
	Up to date	Pearson's r	**0.487**	**0.387**	**0.390**	**0.542**	**0.533**	**0.501**	**0.474**	**0.521**	**0.486**	**0.508**	**0.389**	**0.510**	**0.388**	**0.453**	**0.384**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****
	Update regularly	Pearson's r	**0.483**	**0.366**	**0.426**	**0.588**	**0.588**	**0.507**	**0.474**	**0.504**	**0.461**	**0.515**	**0.341**	**0.478**	**0.368**	**0.385**	**0.328**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****
	Complete metadata	Pearson's r	**0.384**	**0.264**	**0.316**	**0.427**	**0.383**	**0.318**	**0.294**	**0.356**	**0.296**	**0.369**	**0.369**	**0.362**	**0.305**	**0.398**	**0.320**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****
	Machine-readable data	Pearson's r	**0.448**	**0.384**	**0.397**	**0.474**	**0.476**	**0.409**	**0.382**	**0.425**	**0.446**	**0.449**	**0.370**	**0.395**	**0.412**	**0.436**	**0.357**
		p-	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****	**0.000****

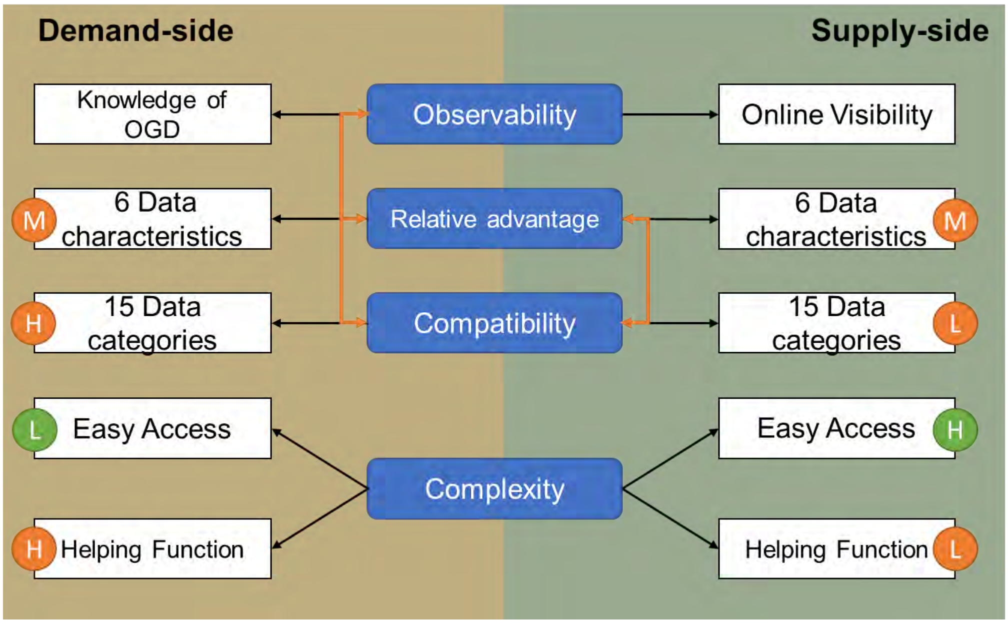


Figure 5-2 Modified research model with conclusions

5.5.2 Theoretical contributions

The theoretical contributions of this study mainly concern the research question of this chapter, which asks what is the relationship between the supply-side and demand-side of OGD? Firstly, since the main aim of our study is to find the possible relationship between the two sides of OGD, a model has been developed from the theory of DOI by including elements from related literature to guide the comparison of the supply-side and demand-side of OGD. Using that literature, we first described each aspect in the area of OGD, then evaluating factors for both sides were selected from previous studies. Online visibility on the supply-side and citizens' knowledge of OGD portals on the demand-side were used to compare the observability. The six most commonly accepted OGD characteristics were used to compare the relative advantage on both sides. The number of 15 OGD categories on the supply-side and citizens' demand towards the same kinds of OGD were used to compare the compatibility. OGD easy access and help functions were used to compare the complexity of both sides. By applying this research model, it is possible to link the two sides of OGD which have previously been studied separately and make comparisons of these two sides using the same constructs.

Secondly, a systematic and practical case study in China has been carried out to test the capability of our model. This case study not only collected data about the development of OGD portals in China, but also focused on Chinese citizens' perspectives on OGD, including their needs for certain data content, requirements for data quality and demands for help functions. Based on this case study, strong conflicts between the supply-side and demand-side of OGD have been recognized in the four aspects of the research model. For observability, most citizens were not aware of OGD portals, exacerbated by the situation where some OGD portals were blocking search engines from collecting their data, which resulted in the data having low online visibility. For relative advantage, the supply-side could not meet the demands of citizens for data to be up to date and updated regularly, with "Up to date" receiving the lowest average score on the supply-side while it was the highest average score on the demand-side. "Free of charge" and "Complete metadata" are the only two OGD characteristics that have higher average scores on the supply-side than on the demand-side. For compatibility, the number of datasets on OGD portals for all OGD categories except "Education", "Cultural activity" and "Local statistics" could not meet the demands of citizens, especially for "Policies and legislation". For complexity, the supply-side of OGD could not meet citizens' demands for help functions via "User guide", "FAQ" and "Online smart agent". The greatest gap lies in use of an online smart agent. "Customer phone support" is the only help function that received a higher score on the supply side. These conflicts between the supply and demand side of OGD help to explain why citizens are not adopting OGD, which leads to the low data utilization.

Thirdly, as presented in Section 5.5.1, the relationship between the four aspects in our research model has been discussed, although it is seldom analyzed in studies applying DOI theory. Understanding these relationships strengthens and has improved our research model. Significant links were found between observability and compatibility in both the supply-side and demand-side of OGD. Our analyses also revealed strong relationships between observability and relative advantage.

5.5.3 Pragmatic implications

The pragmatic implications of this study are mainly related to the future directions for the design of OGD portals based on the relationship between supply and demand sides. The developing directions could be derived from comparative analysis based on our research model. Firstly, OGD portals need to improve their online visibility in order to raise citizens' awareness of the data and the portal. Since observability has a positive relationship with compatibility on the supply-side, OGD portals could improve online visibility by uploading more datasets. Increasing more inlinks and allowing search engines to collect information from the portal could also help to improve online visibility. Secondly, OGD portals need to make sure the data on the portal is up to date and updated regularly, which are among the key concerns of citizens. Thirdly, although scholars have noted the fast development of OGD, it seems that the number of datasets on OGD portals still cannot meet citizens' needs. Thus, governments should continue to open up new OGD for the public as a possible way to encourage OGD utilization. Finally, the help functions on OGD portals cannot meet the needs of citizens. In future development, we recommend that OGD portals could improve their online help services in general, and in particular try adding an online smart agent to their services. Since we carried out the case study in China, which is the largest developing country, with areas of different OGD development contexts and OGD development is still in its infancy in China (R. Huang & Wang, 2016), our results could provide suggestions for other developing countries in forming strategies for developing OGD.

5.6 Conclusion

Our review of the literature confirmed that lack of utilization has become the key problem in the development of OGD (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017), rather than the lack of disclosure of large amounts of data to the public. Although studies evaluating the supply-side of OGD and understanding the perspectives of OGD users have been carried out, no previous studies have systematically compared the two sides and sought to reconcile them. The aim of this chapter was to find the relationships between the supply-side and demand-side of OGD and determine directions for the future development of OGD portals. To reach this aim, this study: (1) built a modified DOI research model as the foundation for comparing the supply-side and demand-side of OGD; and (2) carried out a multiple case study in China to test the capability of the model. Data for the supply-side and demand-side of OGD in China were collected separately. After normalizing the collected data, a direct comparison of these two sides from four aspects were possible, which include observability, relative advantage, compatibility and complexity. The proposed research model can be used not only for analyzing the relationship between supply-side and demand-side of OGD, but also provided a mechanism of communication between these two sides in the OGD ecosystem. Based on the results of the comparison from four aspects, important conclusions have been drawn as follows:

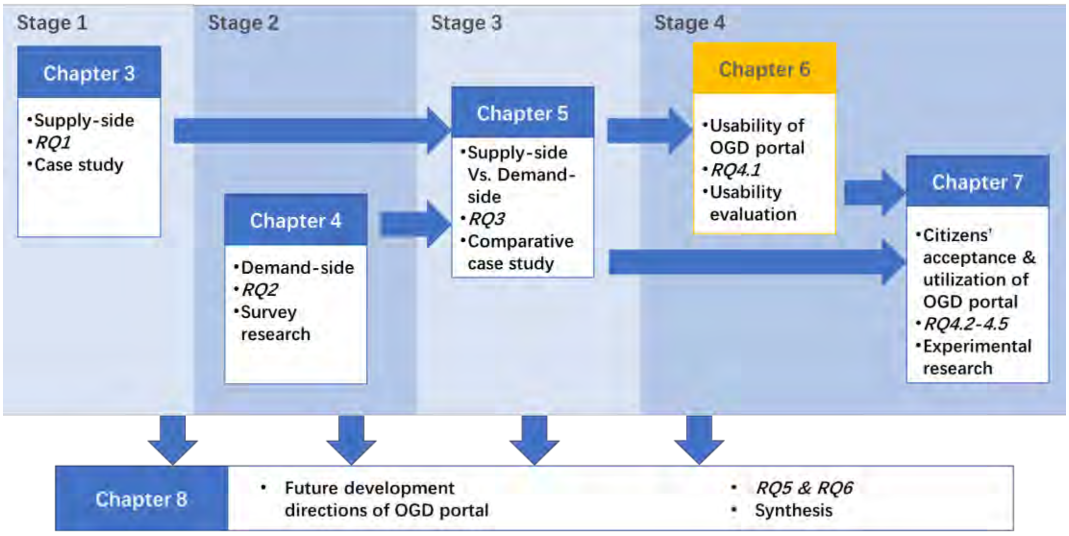
(1) At present, citizens are not fully aware of OGD portals.

(2) Significant conflicts lie in the relative advantage, compatibility and complexity of the two sides of OGD, among which the supply-side cannot meet the needs from the demand-side.

(3) Observability has a strong relationship with relative advantage and compatibility on the demand-side, and has a positive relationship with compatibility on the supply-side.

Based on these comparison results of OGD's two sides, we recommend that future OGD portals should: (1) improve their online visibility; (2) keep their data up to date and perform regular updates; (3) continually release more datasets to the public; and (4) improve help functions and provide online smart agents to assist usage.

6 CHAPTER 6: THE USABILITY OF OGD PORTALS



6.1 Introduction

In the previous chapter, our comparison of the supply-side and demand-side of OGD portals showed several conflicts in the present development of OGD portals. According to supply and demand theory in business, if the supply-side of OGD portals corresponds with the demands of users on the demand-side, users are more likely to use the data on the portals. Thus, how to match the supply-side with the demand-side of OGD portals has become the key issue explored in this thesis for improving OGD utilization.

Usability is prerequisite for technology acceptance and utilization (Nielsen, 1994a). This deals with the question of how well users can use the functionality of an information system (Grudin, 1992). Together with utility, it belongs to usefulness, which is a sub-aspect of system acceptability, basically dealing with the question of whether a system is good enough to satisfy the needs and requirements of its users and other potential stakeholders (Nielsen, 1994b). Usability is a relatively narrow concept compared with the larger issue of the acceptability of the whole information system. But it "applies to all aspects of a system with which a human might interact" (Nielsen, 1994b, p. 25).

Due to the critical effects of usability on the success of a website, in this chapter we focus on the usability aspect of OGD portals and have proposed the following research question for the chapter:

RQ4.1 How to evaluate the usability of current OGD portals?

To answer the above research question, we carried out an exploratory study. Firstly, extended usability principles for OGD portals were selected from the literature. Evaluation criteria for each principle were further developed. Then, we carried out a heuristic evaluation on a set of OGD portals in China to gain understanding of the usability of these portals.

Being the first part of Stage 4 of this thesis, the objectives, which are also the main contributions of this chapter, are twofold: firstly, to find a systematic and objective method to evaluate the usability of a current OGD portal; and secondly, to evaluate the existing province-level OGD portals in China and identify the one with the greatest usability in preparation for the study reported in the next chapter. We also anticipate that the developed evaluation framework and our evaluation findings can be leveraged by designers of OGD portals to improve the usability of these portals in future development.

6.2 Usability and heuristic evaluation

In this section, we discuss the importance of including usability in the study of OGD portals and the approach we chose for the usability inspection, which is heuristic evaluation.

6.2.1 Usability of OGD portals

In the field of human–computer interaction (HCI), usability is a well-known and well-defined concept referring to how well and how pleasantly the user and the computer can interact through the interface (Chou & Hsiao, 2007). For software products, scholars have defined usability as the ability to attract and be understood, learned and operated by its users (Fernandez, Insfran, & Abrahão, 2011). For websites, usability is often treated as a qualitative appraisal of the user-friendliness and ease of use (Y. Lee & Kozar, 2012). Since OGD portals launched by governments are a type of website on the internet, we accepted this concept of qualitative appraisal for usability in our study, focusing on an interaction design to guarantee pleasant and effective interaction between users and OGD portals (Nielsen, 1994b).

Previous studies have shown great benefits from a high level of usability for OGD portals. On one hand, because OGD portals work as a bridge between users and government (Lourenço, 2015), they can thus provide an impression of the accountability of a government and its online services (Z. Huang & Benyoucef, 2014). Therefore, democratic values (Baker, 2009) and the purposes of OGD programs (Attard et al., 2015) require an OGD portal to be user-friendly. Failure of an OGD portal in relation to usability could not only hinder users' access to the portal and its services, but also stymie citizens' positive evaluation of the government (Youngblood & Mackiewicz, 2012). On the other hand, positive effects on users' performances and satisfaction with the portal can be acquired from good usability (Z. Huang & Benyoucef, 2014). Related studies have shown a positive relationship between users' adoption and use of e-government services and the level of usability of the website (Verdegem & Verleye, 2009). Conversely, failure of access to and execution of a website's services because of low levels of usability will lead to user dissatisfaction and prevent them from returning to the website or recommending it to others (Anthopoulos, Siozos, & Tsoukalas, 2007). In order to improve the participation of citizens in making use of the data provided by OGD portals, it is thus essential to consider usability for a portal to make a good impression on its users. However, in the present literature, not enough studies have been carried out focusing on the usability of OGD portals.

Heuristic Evaluations of data processing capabilities have been carried out on U.S. OGD portals, but they only considered the basic heuristic principles which fit all kinds of websites (Tang et al., 2019). The special characteristics of OGD were not included in the previous usability study of OGD portals. On the other hand, similar studies were not found in areas other than the U.S.

6.2.2 Heuristic evaluation

According to the concept of usability that we have accepted, evaluation of the website is regarded as the major approach to study portals' usability, and is often known as usability inspection or expert review (Fernandez et al., 2011). Different methods including heuristic evaluation (Nielsen, 1994c), cognitive walkthrough (Khajouei, Hasman, & Jaspers, 2011) and user testing (Sonderegger & Sauer, 2010) can be used for usability inspection. Among these different approaches, heuristic evaluation is a broadly used method for usability inspection (Nielsen, 1994c) in different studies because of its efficiency, ease of operation and effectiveness (Youngblood & Mackiewicz, 2012). Compared with other methods such as user testing, studies have shown heuristic evaluation's better capacity in trapping a high proportion of problems. In the study of W.-s. Tan, Liu, and Bishu (2009), which used heuristic evaluation and user testing for a web usability evaluation, heuristic evaluation detected 150 out of 183 problems, while user testing only identified 69 problems. Moreover, because of its effectiveness for in-depth inspections, heuristic evaluation can thus recognize serious and specific problems, which can also benefit the future design of an information system (Garcia et al., 2005). Due to the above advantages of heuristic evaluation, we chose this method in our study of OGD portals.

Heuristic evaluation is defined as "a systematic inspection of a user interface design for usability" (Nielsen, 1994b, p. 155). Its main aim is to find usability problems in a user interface design for future improvements (Nielsen & Molich, 1990). Usually, a heuristic evaluation involves using a set of evaluators to examine the interface and judge its compliance compared with recognized usability principles, which are treated as the “heuristics” (Nielsen, 1994b; Nielsen & Molich, 1990). Applying this concept to OGD portals, a heuristic evaluation can identify where the problems lie in the present portal interface and offer suggestions for generating a better design.

6.3 Methodology

In this section, we explain in detail the approach of this exploratory study of the usability of OGD portals in China, including the selection of portals, the building of the usability principles and evaluation criteria for the OGD portals, and the heuristic evaluation procedure.

6.3.1 Selection of OGD portals

The design of this chapter is based on the results of previous chapters, and all case studies and surveys in previous chapters are carried out in China, thus it made sense to also select China for this empirical study of the usability of OGD portals in order to make comparison of the research results with those of other chapters.

To select the target OGD portals for the empirical study, we used two search engines (Google and Baidu) and related studies in the literature to locate province-level OGD portals in China during June 2018. Portals include should be official web portals launched by provincial governments. Thus, the targeted portals were those ones that included a structured domain with “.gov” and contained data from different government departments or institutions of that area. Nine such portals were found. But at the time of usability evaluation, the portal for Xinjiang Province could not be visited. Thus, we deleted this one from the sample and confirmed the other eight portals as the final samples for the usability evaluation, as shown in Table 6.1.

Table 6-1 List of selected OGD portals

Place	Population (Ths)	Regional GDP in RMB (Mio)	General budget revenue in RMB (Mio)	OGD Portal URL
Beijing	21,730	25,669.13	5,081.26	http://www.bjdata.gov.cn
Guangdong	109,990	80,854.91	10,390.35	http://www.gddata.gov.cn
Guizhou	35,550	11,776.73	1,561.34	http://www.gzdata.gov.cn
Hong Kong	7,377	5,469.96	3,503.72	https://data.gov.hk
Shanghai	24,200	28,178.65	6,406.13	http://www.datashanghai.gov.cn
Sichuan	82,620	32,934.54	3,388.85	http://www.sc.gov.cn/10462/13797/index.shtml
Taiwan	23,540	33,328.55	5,780.29	http://data.gov.tw
Zhejiang	55,900	47,251.36	5,301.98	http://data.zjzwfw.gov.cn

6.3.2 Extension of usability principles

Thousands of usability principles have been proposed to guide the developers of information systems (Nielsen, 1994b). Molich and Nielsen published 10 usability principles explaining most of the problems that can be observed in interface designs, which should be followed by all user interface designers (Molich & Nielsen, 1990). Based on this original version of the principles and a comparison with several other published usability heuristics, Nielsen then developed an enhanced version of the usability heuristics (Nielsen, 1994a) which is the most popular, accepted set of principles for evaluating website usability (Budd, 2007; Z. Huang & Benyoucef, 2014). It has been applied to various kinds of websites including e-commerce websites (Hasan, Morris, & Probets, 2013), e-government websites (Delopoulos, 2015; Saeed, Malik, & Wahab, 2013), and OGD portals in the US (Tang et al., 2019), which not only proved its capability in recognizing the usability problems of websites but also supports its applicability in the field of OGD. Therefore, it was reasonable to include Nielsen's 10 usability principles in our study.

However, these principles were developed more than 20 years ago and were originally designed for usability evaluation of software in general, with later extension to evaluate websites. In addition, although OGD portals are a subtype of websites, they have their own characteristics compared to other kinds of websites. Firstly, due to their nature, they aim to offer data or datasets to the end-users

(Kassen, 2013; Kostovski et al., 2012; Lourenço, 2013, 2015). Thus, they need to include data services on the interface. Also, the data provided on the portal needs to follow the definition of OGD. All of these issues together called for further development of Nielsen's existing 10 usability principles to address the specific needs of OGD portals.

Considering the characteristics of OGD portals, Fajar Marta proposed a set of design principles for OGD portals to reflect the three core values of open data: transparency, privacy and information quality (Fajar Marta, 2016). The proposed OGD portal design principles were validated based on observation of five OGD portals. Since the objects of the heuristic evaluation of this study are portals specially designed for OGD, we chose to also include these principles in our study.

However, repetition could be recognized in some of the included principles. For example, the fourth OGD portal design principle by Fajar Marta requires the portal to provide help functions, which is the same as the tenth principle of Nielsen about providing help and documentation. Some of the principles are not suitable for heuristic evaluation carried out by non-operators, such as the thirteenth OGD portal design principle by Fajar Marta requiring the dataset on the portal to remove privacy information before being uploaded.

Therefore, a further selection of usability principles was then carried out to exclude duplicated ones and those that are unable to implement objective evaluations. After the selection, we finally confirmed 21 usability principles for OGD portals. The results of the selection of usability principles are shown in Table 6.2.

6.3.3 Development of evaluation criteria

Usability principles are called "heuristics" because they are quite broad and can be applied to any type of user interface, but are not specific usability principles

(Nielsen, 1994b). Therefore, the extended usability principles for OGD portals in Table 6.2 are too general for carrying out heuristic evaluation and in-depth assessment. In traditional heuristic evaluations, the discovery of usability problems depends on the observation and experiences of the evaluator to find violations of the usability principles (Tang et al., 2019). However, lack of detail in usability principles will result in failure of heuristic evaluation in identifying the usability problems (Z. Huang & Benyoucef, 2014), especially those problems related to links on the websites (Hasan et al., 2013), because these problems need evaluators to test them in order to be found. Thus, to both avoid the disadvantage of heuristic evaluation in recognizing problems that need user testing and to ensure the heuristic evaluation in this study was objective and practical, it was essential to devise specific criteria for judgement of each selected principle. In our approach, the focus of evaluators was on identification of compliance with criteria rather than on violation of principles. By adding these criteria to the general usability principles, it would reduce the difficulty of evaluators to judge the compliance of usability principles of the target portals. These criteria were developed based on a combination of: (1) the description and explanation of each principle (Fajar Marta, 2016; Nielsen, 1994b); (2) detailed examples of each principle (Budd, 2007; Neil, 2009; Tognazzini, 2014) in the field of web applications; and (3) interpretation of relevant studies of the usability of OGD portals (Bogdanović-Dinić et al., 2014; OpenDataBarometer, 2016; OpenDataIndex, 2017; OpenDataMonitor, 2017; SunlightFoundation, 2010; Thorsby et al., 2017). In total, 54 criteria were devised for the selected usability principles, as shown in Table 6.3.

Table 6-2 Selection of usability principles

No.	Principles	Description	Include	Reason for removal	Ref.
1	Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	Yes		Nielsen.1994a
2	Match between system and the real world	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	Yes		
3	User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.	Yes		
4	Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	Yes		
5	Error prevention	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Yes		
6	Recognition rather than recall	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	Yes		
7	Flexibility and efficiency of use	Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	Yes		
8	Aesthetic and minimalist design	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	Yes		
9	Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	Yes		
10	Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.	Yes		

11	OGD Portal Design Principles-01	Provide the mechanism for users to request datasets and to track whether their request has been approved or rejected.	Yes		Fajar Marta, 2016
12	OGD Portal Design Principles-02	Provide the mechanism for users to give feedback to datasets.	Yes		
13	OGD Portal Design Principles-03	Stimulate knowledge exchange between users by providing space for interaction between users.	Yes		
14	OGD Portal Design Principles-04	Assist users in understanding and using datasets as well as using the open data portal by providing FAQ, channel for users to ask questions, tutorials, and examples of dataset usage.	No	Covered by No.10	
15	OGD Portal Design Principles-05	Tag datasets with relevant keywords to allow faster searching of datasets.	Yes		
16	OGD Portal Design Principles-06	Allow users to download and request datasets without having to register first.	Yes		
17	OGD Portal Design Principles-07	Assist users in searching datasets by providing filtering and suggestion feature on the search function.	Yes		
18	OGD Portal Design Principles-08	Provide API when users require to access datasets through external applications.	Yes		
19	OGD Portal Design Principles-09	Provide datasets in open, machine-processable format and allow users to download datasets in the format that they prefer.	Yes		
20	OGD Portal Design Principles-10	Provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets.	Yes		
21	OGD Portal Design Principles-11	Provide the function to visualize datasets to allow quick viewing of datasets.	Yes		
22	OGD Portal Design Principles-12	Use user-friendly interface to ease navigation by categorizing datasets into different topics, using universal icons, highlighting important datasets, using consistent design between pages, and providing the user interface in the language that the users need.	No	Covered by No.2, 4, 6	
23	OGD Portal Design Principles-13	Deal with privacy issues before dataset is uploaded to the portal.	No	It's hard to examine the privacy issues of all datasets on a portal	
24	OGD Portal Design Principles-14	Take into account the context of regulation, target users, and purpose in the implementation of the portal's functionalities.	Yes		
25	OGD Portal Design Principles-15	Involve users in early and later development of the portal to match portal design with users' need.	No	Could not know whether the portal has included users in the early design of the portal	

Table 6-3 Criteria for usability principles

No.	Principles	Criteria	Reference
1	Visibility of system status	1. A feedback message is displayed when an action is performed	Budd, A. (2007), Neil, T. (2009), Nielsen, J. (1994a) Tognazzini, B. (2014)
		2. Progress message and indicator shows while data is loading and downloading	
		3. Highlight the current menu items that resulted in the particular panel and links of different layers	
2	Match between system and the real world	4. The words used on the portal belong to "Lexicon of Common Words in Contemporary Chinese"	
		5. Data categories used by the portal belong to "Lexicon of Common Words in Contemporary Chinese"	
		6. Labels and buttons are clear, meaningful and understandable and are not misleading	
3	User control and freedom	7. Search can be directly opened, entered info, executed or cancelled on each page	
		8. Clearly marks where the person is and where they can go by showing the selection in each menu	
4	Consistency and standards	9. The design and structure of the portal follow general online database/platform conventions	
		10. Pages of the same level have similar layouts	
		11. Place items in standard locations like search boxes at the top right of the screen	
5	Error prevention	12. Disables the button after it is clicked, so the person cannot click twice by accident	
		13. Make the primary action prominent with a larger click area. Cancel and secondary actions are just shown as links	
		14. The auto recommending feature cuts down on misspellings	
		15. Auto focus on input	
6	Recognition rather than recall	16. Users of the portal can get instructions for use easily	
		17. Menu and button labels on the portal should have the key words describing their function, forming unique labels	
		18. Provide a clear site name and purpose	
		19. Provide a breadcrumb trail	
7	Flexibility and efficiency of use	20. The portal supports common shortcuts	
		21. Provide quick links to common features/functions	
		22. Provide advanced features	
		23. Pre-check common options	
		24. Allow defaults to be changed, cancelled or overridden	
8	Aesthetic and minimalist design	25. Text for reading on the portal has high contrast	
		26. Font size is large enough to read	
		27. Colour used on the portal repeats certain information	
		28. Strong left alignment of text is used on the portal	
		29. A light rule is used to separate tags from the other options	
		30. Use progressive disclosure to hide advanced features	
		31. Priorities using size, shape, colour, alignment and proximity	

9	Help users recognize, diagnose, and recover from errors	32. Visually highlight errors	Budd, A. (2007), Neil, T. (2009), Nielsen, J. (1994a) Tognazzini, B. (2014)
		33. Provides immediate feedback with specific instructions on user's errors	
		34. Provide feedback close to where the error occurred	
		35. Use clear messages and avoid technical jargon in helping messages	
		36. Disable irrelevant options	
10	Help and documentation	37. Embedded videos is used to showcase features as well as get people started using the product	
		38. Help tips are displayed on hover, answering the most likely questions about a field or instructions	
		39. Offer a help button or link, which opens a new browser window/tab with a full set of help resources: search, FAQ, video tutorials, customer forums	
		40. Provide examples and contextual help	
11	OGD Portal Design Principle-01	41. Provide the mechanism for users to request datasets that could not be found on the portal at present	Fajar Marta, (2016)
		42. Users could track whether their request has been approved or rejected	
12	OGD Portal Design Principle-02	43. Provide the mechanism for users to give feedback to datasets	
13	OGD Portal Design Principle-03	44. Provide the mechanism for users to interact with each other	
14	OGD Portal Design Principle-04	45. Tag datasets with relevant keywords to allow faster searching of datasets	
15	OGD Portal Design Principle-05	46. Allow users to download datasets without having to register first	Fajar Marta, (2016), Sunlight Foundation. (2010); Open Data Index. (2017);
		47. Allow users to request datasets without having to register first	
16	OGD Portal Design Principle-06	48. Provide filtering and suggestion feature on the search function	Fajar Marta, (2016)
17	OGD Portal Design Principle-07	49. Provide application programming interface when users require to access datasets through external applications	Fajar Marta, (2016), Sunlight Foundation. (2010); Open Data Index. (2017)
18	OGD Portal Design Principle-08	50. Provide datasets in open, machine-processable format	Fajar Marta, (2016) Sunlight Foundation. (2010); Open Data Index. (2017)
		51. Allow users to download datasets in the format that they prefer	Fajar Marta, (2016) OpenDataBarometer. (2016)
19	OGD Portal Design Principle-09	52. Provide context on datasets through metadata, including data owner, semantic rules, update frequency, and history of changes to datasets	Fajar Marta, (2016) Sunlight Foundation. (2010); Bogdanović-Dinić et al., (2014); OpenDataMonitor, (2017)
20	OGD Portal Design Principle-10	53. Provide the function to visualize datasets to allow quick viewing of datasets	Fajar Marta, (2016) Thorsby et al., (2017)
21	OGD Portal Design Principle-11	54.Ask or identify the type of users and offer recommendations and priorities accordingly	Fajar Marta, (2016)

6.3.4 Evaluators

Heuristic evaluation is performed by having each individual evaluator inspect the targeted OGD portals alone. Therefore, before carrying out the evaluation, one important step is to choose the evaluators. Considering cost–benefit analysis and the possibilities of finding most of the usability problems, the recommended number of evaluators according to Nielsen is 3 to 5 (Nielsen, 1994b). His study also demonstrated a significant differences in evaluators' performances between experts and non-experts (Nielsen, 1992). Thus, we choose to recruit experts in the field of computing as our evaluators. By invitation through email, we finally recruited 4 PhD students studying computer science who had experience with web development and had never used the targeted 8 portals before as the evaluators for the heuristic evaluation. Because this was an exploratory investigation, the limited number of evaluators would not greatly affect recognition of the usability problems (Z. Huang & Benyoucef, 2014). Also, because of the development of specific criteria for each principle, the heuristic evaluation of the compliance of each criterion is thus less dependent on the experience and interpretation of evaluators but more dependent on their observation and testing of the target OGD portals according to each criterion. Therefore, the recruited PhD students with experiences in web development could be qualified to be the evaluators of this study.

6.3.5 Evaluation procedure

Before the start of the evaluation process, each evaluator was provided with a clear explanation of the background and purpose of this study together with the table for all 8 targeted OGD portals listing all the criteria and principles. The evaluators were not given any training of the criteria and principles in case the training will limit their observation and testing of the portal when carrying out the evaluation. Each evaluator needed to complete heuristic evaluation of all 8 portals in the same order following the same process. Firstly, they roughly reviewed the portal for 5 minutes to become familiar with the target. Secondly, the evaluators went through the portal several times and inspected various elements and functions and compared them with the 54 criteria listed in Table 6.3. If the statement of a criterion fit the portal, the evaluators marked that criterion with a value of 1, otherwise with 0. After they had completed evaluating the usability of one OGD portal, they moved onto the next selected one. Evaluators were not allowed to communicate and did not have their findings aggregated until all had completed the evaluation of all selected OGD portals (Nielsen, 1994b). The results of the evaluation of each evaluator were recorded in a separate electronic table and aggregated to form the final report.

6.3.6 Data collection and analysis

The evaluation was carried out from 12 June to 20 July 2018. The results were separately recorded in electronic tables by each evaluator and collected through email.

After collecting all 4 evaluators' results for the 8 selected portals, a Fleiss's fixed-marginal multi-rater kappa (Fleiss, 1971) was calculated for the results to test the agreement of evaluators on each evaluation criterion. We choose to use Fleiss's kappa because it fit our circumstances that evaluators were asked to judge all 8 cases (Brennan & Prediger, 1981). Then the average score of each criterion and principle was calculated and compared to draw conclusions. We choose to use the average score because it is the most commonly used and important value in a heuristic evaluation (Nielsen, 1994b).

6.4 Results and discussion

In this section, we first present the results of Fleiss's kappa for each criterion and principle. Then comparisons of the results by principle and by portal are given separately.

6.4.1 Fleiss's kappa

Results of the calculation of Fleiss's kappa are shown in Table 6.4. According to Fleiss, kappa values less than 0.4 are "poor" in agreement, values from 0.4 to 0.75 are "intermediate to good" and values above 0.75 are "excellent" (Fleiss, Levin, & Paik, 2013). We can see from the results that 57.14% of the principles have the agreement level of "intermediate to good" and the other 42.86% have "excellent". For the Fleiss's kappa of the criteria, the percentage of "poor" is 7.41%, "intermediate to good" is 53.7% and "excellent" is 38.89%. The total average Fleiss's kappa also shows the agreement of the framework to be "intermediate to good". Among all 432 judges of the compliance of criteria for 8 target OGD portals, all four evaluators agreed on 194 of them (44.91%), one evaluators disagreed with the other three in 142 of them (32.87%), two evaluators disagreed with the other two in 96 of them (22.22%). Disagreements focus on principles 1, visibility of system status, 7, flexibility and efficiency of use, and 9, help users recognize, diagnose and recover from errors. We think this disagreement is due to the different observations of the related functions on the portal. Because the evaluation of each criterion is based on the experiences, interpretations, and testing of the targeted portals of evaluators, some of the evaluators may have noticed a function while others did not. Also, for principle 7, evaluators may have different understanding of what are common shortcuts, options and features. Since most of the criteria have medium to high rates of agreement which indicated the acceptability of the results of this evaluation, we continued our analysis of the evaluation results.

Table 6-4 Fleiss's kappa for principles and criteria

Guideline No.	Fleiss's Kappa	Criteria No.	Fleiss's Kappa	Guideline No.	Fleiss's Kappa	Criteria No.	Fleiss's Kappa
1	0.479	1	0.479	8	0.735	25	0.688
		2	0.354			26	0.854
		3	0.604			27	0.938
2	0.917	4	0.938			28	1.000
		5	0.938			29	0.750
		6	0.875			30	0.458
3	0.792	7	0.708			31	0.458
		8	0.875	9	0.488	32	0.375
4	0.806	9	0.813			33	0.458
		10	0.854			34	0.354
		11	0.750			35	0.708
5	0.734	12	0.417			36	0.542
		13	0.521	10	0.609	37	0.938
		14	1.000			38	0.479
		15	1.000			39	0.542
6	0.844	16	0.625			40	0.479
		17	1.000	11	0.667	41	0.604
		18	0.938			42	0.729
		19	0.813	12	0.792	43	0.792
7	0.508	20	0.417	13	0.792	44	0.792
		21	0.625	14	0.708	45	0.708
		22	0.583	15	0.792	46	0.938
		23	0.542			47	0.646
		24	0.375	16	0.646	48	0.646
19	0.813	52	0.813	17	0.708	49	0.708
20	0.583	53	0.583	18	0.667	50	0.458
21	0.938	54	0.938			51	0.875
Total average Fleiss's Kappa = 0.691							

6.4.2 Comparisons by principle and criterion

Results of the calculation of the average value for each principle and criterion are shown in Table 6.5.

Table 6-5 Average values for principles and criteria

Guideline No.	Average Value	Criterion No.	Average Value	Guideline No.	Average Value	Criterion No.	Average Value
1	2.583	1	2.875	8	3.268	25	3.375
		2	1.875			26	3.625
		3	3.000			27	3.875
2	3.833	4	3.875			28	4.000
		5	3.875			29	3.500
		6	3.750			30	1.750
3	3.375	7	3.250			31	2.750
		8	3.500	9	2.175	32	2.000
4	3.333	9	3.375			33	2.750
		10	3.625			34	1.875
		11	3.000			35	3.250
5	0.656	12	1.500			36	1.000
		13	1.125	10	1.406	37	0.125
		14	0.000			38	1.625
		15	0.000			39	2.000
6	3.313	16	2.250			40	1.875
		17	4.000	11	1.375	41	2.125
		18	3.875			42	0.625
		19	3.125	12	2.250	43	2.250
7	2.000	20	2.250	13	1.000	44	1.000
		21	3.000	14	1.250	45	1.250
		22	1.750	15	3.125	46	3.375
		23	1.250			47	2.875
		24	1.750	16	1.375	48	1.375
19	3.125	52	3.125	17	2.500	49	2.500
20	1.500	53	1.500	18	1.500	50	2.750
21	0.125	54	0.125			51	0.250
Average value for all principles = 2.146 for all criteria = 2.340							

For the principles, we can see that the OGD portals' performances varied greatly in relation to different principles. They performed the best with principle 2, matching between system and the real world, and the worst in principle 21, asking or identifying the type of users and offering recommendations and priorities accordingly. Principles 3, 4, 6, 8 and 15 also got very high values, while principles 5, 13, 14, 16, 18 and 20 got relatively low values. The average value (2.594) for the principles of general interface usability design (1 to 10) was much higher than the average value (1.739) for the principles of OGD portal design (11 to 21). For the criteria, we noted the highest value for criterion 17 (menu and button labels on the portal should have the key words describing their function, forming unique labels) and 28 (font size is large enough to read), with a value of 4, which means that all portals did well in relation to these two criteria. Criteria 4, 5, 6, 18 and 27 also received relative high scores. The lowest values were for criteria 14 (the auto recommending feature cuts down on misspellings) and 15 (auto focus on input) with a value of 0, which indicates that none of the portals provided these two functions. Criteria 37, 42, 51 and 54 also received relatively low scores.

6.4.3 Comparison of portals

The Kruskal–Wallis H test (Kruskal & Wallis, 1952) was carried out to compare the values for each portal by different evaluators. The results of the Kruskal–Wallis H test are shown in Table 6.6. The p-value = 0.051, indicating that the differences in the values of usability evaluation among different portals were not significant. Shanghai got the highest mean rank, indicating it to be the best in the usability evaluation, followed by Hong Kong and Taiwan, while Sichuan got the lowest mean rank.

Comparing the rankings of usability with the results of the OGD portal evaluation in Chapter 3, we noted a significant difference. In Chapter 3, the best performing OGD portal was Taiwan, followed by Hong Kong. Shanghai ranked the fifth among all 9 OGD portals. Guangdong performed the best of all mainland Chinese OGD portals. But in the results of the usability evaluation, Shanghai was the best-performing one among all 8 portals. This result indicates that at present, the usability of Chinese OGD portals does not match with the performance and quality of the data on the portal. This further validates the importance of monitoring and enhancing the usability of OGD portals (Scott, 2005) in order to realize the targeted aims of OGD programs.

Table 6-6 Kruskal–Wallis H test of OGD portals

OGD Portal	Mean Rank
Beijing	14.13
Guangdong	17.00
Guizhou	19.88
Hong Kong	21.88
Shanghai	23.13
Sichuan	2.50
Taiwan	20.00
Zhejiang	13.50
Chi-square	**14.028**
df	**7.000**
Sig.	**0.051**

We compared the scores for each principle by portal with one-way ANOVA. Results are shown in Table 6.7. The score for each principle is the average score of the criteria belonging to this principle. We can see that there were significant differences in scores among the different portals for 10 out of 21 principles. Four of them are principles for the general design of websites, while 6 are principles specifically for OGD portals. This indicates that at present, the usability development of these OGD portals is not equal, especially in relation to those principles specific for OGD portals.

Post-hoc pairwise comparison showed that for principles 3 (user control and freedom), 6 (recognition rather than recall) and 7 (flexibility and efficiency of use) the score for Sichuan was significantly lower than for other portals. For principle 4 (consistency and standards) Guizhou received a significantly lower score compared with Beijing, Guangdong, Hong Kong, Shanghai and Zhejiang. For principle 12 about providing feedback to datasets, the scores for Sichuan and Zhejiang were significantly lower than for other portals. For principle 13 about having interaction with other users, Guangdong and Taiwan received significantly higher scores than the other portals. For principle 14 about tagging datasets, Shanghai and Taiwan received significantly higher scores compared with the other portals. For principle 15 about cancelling registration, Beijing received a significantly lower score than the other portals. For principles 17 about application programming interface (API) and 19 about metadata, the scores for Sichuan were significantly lower than for other portals.

Table 6-7 Portals' average scores by principle and one-way ANOVA results

Principle No.	Beijing	Guang-dong	Guizhou	Hong Kong	Shang-hai	Sichuan	Taiwan	Zhe-jiang	Std.	F	p
1	2.33	3.00	2.67	3.00	3.00	1.67	2.33	2.67	0.46	0.514	0.815
2	3.67	4.00	4.00	4.00	3.67	4.00	3.33	4.00	0.25	0.762	0.624
3	3.50	4.00	3.00	4.00	3.50	1.50	4.00	3.50	0.83	2.786	**0.029***
4	4.00	4.00	2.33	4.00	4.00	1.33	3.33	3.67	0.99	5.061	**0.001****
5	0.25	0.50	1.00	0.75	1.00	0.50	0.75	0.50	0.27	0.730	0.649
6	3.00	3.25	4.00	4.00	3.50	1.75	3.50	3.50	0.72	5.798	**0.001****
7	2.20	1.40	2.60	2.60	2.40	0.60	2.40	1.80	0.70	2.782	**0.029***
8	3.14	3.29	3.29	3.43	3.43	3.00	3.14	3.43	0.16	0.126	0.996
9	2.40	2.40	2.20	2.60	2.40	1.20	2.40	1.80	0.46	0.332	0.932
10	1.75	0.75	1.25	2.50	1.75	0.00	1.75	1.50	0.76	1.531	0.204
11	1.00	1.50	2.00	1.00	2.00	0.00	2.50	1.00	0.79	1.429	0.240
12	4.00	2.00	3.00	1.00	4.00	0.00	4.00	0.00	1.75	7.371	**0.000****
13	0.00	3.00	0.00	0.00	2.00	0.00	3.00	0.00	1.41	4.800	**0.002****
14	0.00	0.00	2.00	2.00	3.00	0.00	3.00	0.00	1.39	3.306	**0.013***
15	0.50	4.00	4.00	3.50	4.00	3.00	2.50	3.50	1.19	5.643	**0.001****
16	1.00	0.00	2.00	3.00	3.00	0.00	2.00	0.00	1.30	2.395	0.052
17	3.00	4.00	4.00	1.00	4.00	0.00	2.00	2.00	1.51	3.918	**0.006****
18	1.50	1.50	1.50	2.00	1.50	1.00	2.00	1.00	0.38	0.381	0.904
19	3.00	4.00	4.00	3.00	4.00	0.00	4.00	3.00	1.36	4.905	**0.002****
20	3.00	3.00	2.00	2.00	1.00	0.00	1.00	0.00	1.20	1.714	0.153
21	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.35	1.000	0.455

We further compared each portal's scores for each criterion by one-way ANOVA. Results are shown in Table 6.8. The scores for each portal for each criterion are the sum of the scores from the 4 different evaluators. The one-way ANOVA shows that 11 out of 54 criteria had significantly different values among different portals, 6 of which are specific criteria for OGD portals, 5 general criteria for all websites.

Table 6-8 Portals' average scores by criterion and one-way ANOVA results

Criterion No.	Beijing	Guang-dong	Guizhou	Hong Kong	Shang-hai	Sichuan	Taiwan	Zhe-jiang	F	p	Criterion No.	Beijing	Guang-dong	Guizhou	Hong Kong	Shang-hai	Sichuan	Taiwan	Zhe-jiang	F	p
1	3	3	3	3	3	2	3	3	0.120	0.996	28	4	4	4	4	4	4	4	4	—	—
2	2	2	2	2	2	1	2	2	0.097	0.998	29	4	4	3	3	3	4	3	4	0.571	0.772
3	2	4	3	4	4	2	2	3	1.143	0.371	30	2	1	2	3	3	1	1	1	0.725	0.652
4	4	4	4	4	4	4	3	4	1.000	0.455	31	3	2	3	3	3	2	3	3	0.198	0.983
5	4	4	4	4	4	4	3	4	1.000	0.455	32	3	2	2	2	2	1	2	2	0.229	0.974
6	3	4	4	4	3	4	4	4	0.857	0.553	33	3	3	3	3	3	2	3	2	0.198	0.983
7	3	4	2	4	3	2	4	4	1.347	0.272	34	2	2	2	2	2	1	2	2	0.097	0.998
8	4	4	4	4	4	1	4	3	4.571	0.002**	35	3	4	3	4	4	2	4	2	1.347	0.272
9	4	4	3	4	4	1	3	4	3.000	0.021*	36	1	1	1	2	1	0	1	1	0.312	0.942
10	4	4	3	4	4	2	4	4	1.898	0.114	37	0	0	0	1	0	0	0	0	1.000	0.455
11	4	4	1	4	4	1	3	3	3.429	0.011*	38	2	1	1	3	2	0	2	2	0.806	0.591
12	1	1	2	2	2	1	2	1	0.245	0.969	39	3	1	1	3	3	0	3	2	1.558	0.196
13	0	1	2	1	2	1	1	1	0.429	0.875	40	2	1	3	3	2	0	2	2	0.943	0.493
14	0	0	0	0	0	0	0	0	—	—	41	1	3	4	1	3	0	3	2	2.323	0.058
15	0	0	0	0	0	0	0	0	—	—	42	1	0	0	1	1	0	2	0	1.022	0.441
16	1	2	4	4	3	0	2	2	2.571	0.040*	43	4	2	3	1	4	0	4	0	7.371	0.000**
17	4	4	4	4	4	4	4	4	—	—	44	0	3	0	0	2	0	3	0	4.800	0.002**
18	4	4	4	4	4	3	4	4	1.000	0.455	45	0	0	2	2	3	0	3	0	3.306	0.013*
19	3	3	4	4	3	0	4	4	4.905	0.002**	46	0	4	4	4	4	4	3	4	15.857	0.000**
20	3	2	2	3	2	1	3	2	0.435	0.870	47	1	4	4	3	4	2	2	3	1.790	0.136
21	3	4	4	3	3	1	3	3	1.143	0.371	48	1	0	2	3	3	0	2	0	2.395	0.052
22	2	0	3	3	3	0	2	1	1.971	0.102	49	3	4	4	1	4	0	2	2	3.918	0.006**
23	1	0	2	2	2	0	2	1	0.857	0.553	50	3	3	3	3	3	2	3	2	0.198	0.983
24	2	1	2	2	2	1	2	2	0.171	0.989	51	0	0	0	1	0	0	1	0	0.857	0.553
25	3	4	3	3	4	3	3	4	0.429	0.875	52	3	4	4	3	4	0	4	3	4.905	0.002**
26	2	4	4	4	3	4	4	4	1.898	0.114	53	3	3	2	2	1	0	1	0	1.714	0.153
27	4	4	4	4	4	3	4	4	1.000	0.455	54	0	0	0	0	0	0	1	0	1.000	0.455

For criteria 8 about the navigation bar, 9 about following the general online platform conventions, 16 about easily getting instructions and 19 about providing a breadcrumb trail, Sichuan received significantly lower scores than the other portals. This was because the Sichuan portal did not provide functions to show where the user is and where they can go, nor could the portal record the trail of user's visits. Also, the interface of Sichuan portal looked more like lists of data/datasets, instead of a common interface for a database with a search bar. For criterion 11 about standard locations, Sichuan and Guizhou received significantly lower scores compared with the other portals. However, evaluators show different opinions on the performance of these two portals concerning criterion 11, since two of them disagreed with the other two. For criterion 43 about providing feedback functions, Beijing and Taiwan received significantly higher scores compared with the other portals by providing feedback functions in a prominent position, and Sichuan received a significantly lower score. For criterion 44 about providing the function of interaction with other users, Guangdong and Taiwan received significantly higher scores than the other portals by providing functions for users to rank and leave comments for each dataset. For criterion 45 about tagging datasets, the scores for Shanghai and Taiwan were significantly higher compared with the other portals by marking each dataset with several different tags and keywords. For criterion 46 about downloading without registration, Beijing received a significantly lower score than the other portals since users have to register first before downloading any datasets. For criteria 49 about API and 52 about metadata, the scores for Sichuan were significantly lower than for the other portals since it did not provide metadata for each dataset nor did it provide API functions.

Based on the above analysis, we can deduce that there are strong signs of disadvantages with the usability of the Sichuan portal, while the advantages for those of Guangdong, Taiwan and Shanghai in usability are obvious.

6.5 Conclusion

A good interface design can generate more added value for a system than hardware manufacturing (Rappaport & Halevi, 1991). A usability heuristic evaluation can find the advantages and disadvantages of an interface and help direct future design processes (Nielsen, 1994b). Since OGD portals offer users a direct interface for accessing OGD, a usability evaluation of these portals can help discover current problems in portal design besides the quality and quantity of data on the portal. Thus, in this chapter we have focused on the usability design for an OGD portal with an empirical study of province-level OGD portals in China. We first extended usability principles to building OGD portals, then carried out a heuristic evaluation of the usability of the 8 selected province-level OGD portals in China. The extended usability principles together with the devised criteria have given the answer to our research question *RQ4.1 How to evaluate the usability of current OGD portals?*

We draw the following conclusions based on the results of applying the usability framework to a case study of Chinese province-level OGD portals. For the present usability performances of Chinese province-level OGD portals:

(1) The portals performed better in adhering to general usability design principles compared with principles specifically for OGD portals.

(2) The portals performed best in matching between the system and the real world, and the worst in identifying user types for offering recommendations accordingly.

(3) The portals also did not perform well in error prevention, user interaction, tagging datasets, providing filtering and suggestion features, offering different kinds of open format data or dataset visualization.

(4) The Shanghai portal had the best usability design among all the Chinese province-level portals.

(5) The Sichuan portal demonstrated significant disadvantages in usability compared with the other portals, while the advantages for the Guangdong, Taiwan and Shanghai portals in usability were apparent.

Results of this heuristic evaluation of the usability of OGD portals are only limited to the targeted portals in China. The evaluation also depended on the interpretations, experiences and testing of evaluators on the targeted portals, which may lead to disagreements of evaluation results among different evaluators. However, the development and application of specific criteria besides higher level usability principles in the evaluation not only reduced the dependence of heuristic evaluation on evaluators' experiences to find usability problems, but also helped to fill the defect of heuristic evaluation in recognizing problems that need testing of the targeted websites (Hasan et al., 2013). On the other hand, problems recognized of the usability of OGD portals in China are limited to the scope of selected principles and criteria.

Regarding the help functions provided by the OGD portals, the average scores for principles relating to help functions (9 and 10) in the usability evaluation were relatively low compared with the other principles. This indicates the OGD portals' disadvantages in offering users help, which is supported by other research noting OGD portals' failure to provide help functions such as FAQ, help topics or tutorials (Thorsby et al., 2017). However, studies have shown OGD users' great need to receive help from the portal (D. Wang, Richards, & Chen, 2018). OGD portals' limitation in sufficient help functions may be a cause of end-users' low utilization of these portals.

Comparing the usability evaluation results of OGD portals in China with the similar evaluations in the US by applying the 10 Nielsen's principles (Tang et al., 2019), similarities and differences are found in the performance of OGD portals on different usability principles. OGD portals in both China and the US have problems in complying principles No. 7 (Flexibility and efficiency of use), 9

(Help users recognize, diagnose and recover from errors), and 10 (Help and documentation). Customization and help function therefore could be seen as common problems in the present development of OGD portals in both China and the US. On the other hand, OGD portals in both countries performed well in principles No. 4 (Consistency and standards), 6 (Recognition rather than recall) and 8 (Aesthetic and minimalist design), which are more related to the design of the website's appearance. However, great differences have been recognized in the compliance of principle No. 5 (Error prevention) and No. 2 (Match between system and the real world). Principle No.5 received the lowest score in the heuristic evaluation in Chinese OGD portals, indicating the deficiency of functions to help users preventing making mistakes. However, only 4 problems have been recognized in the heuristic evaluation of the US portals. Principle No. 2 received the highest score among all principles in the heuristic evaluation of Chinese OGD portals. But for the US, 5 problems have been recognized with this principle. The differences of these two principles in the evaluation results of China and the US may be due to the different methods in the evaluation process. Further studies could be carried out to compare the usability performance of OGD portals in these two countries by applying the evaluation framework proposed in this study which included characteristics of OGD.

Thus, we choose to focus on help functions for the design of the experiment, as presented in the next chapter. In addition, because Shanghai obtained the highest mean rank among all the OGD portals in China, we decided to build the imitated portal for the experiment according to the design of the present Shanghai OGD portal, which screenshot is provided in Figure 6.1.

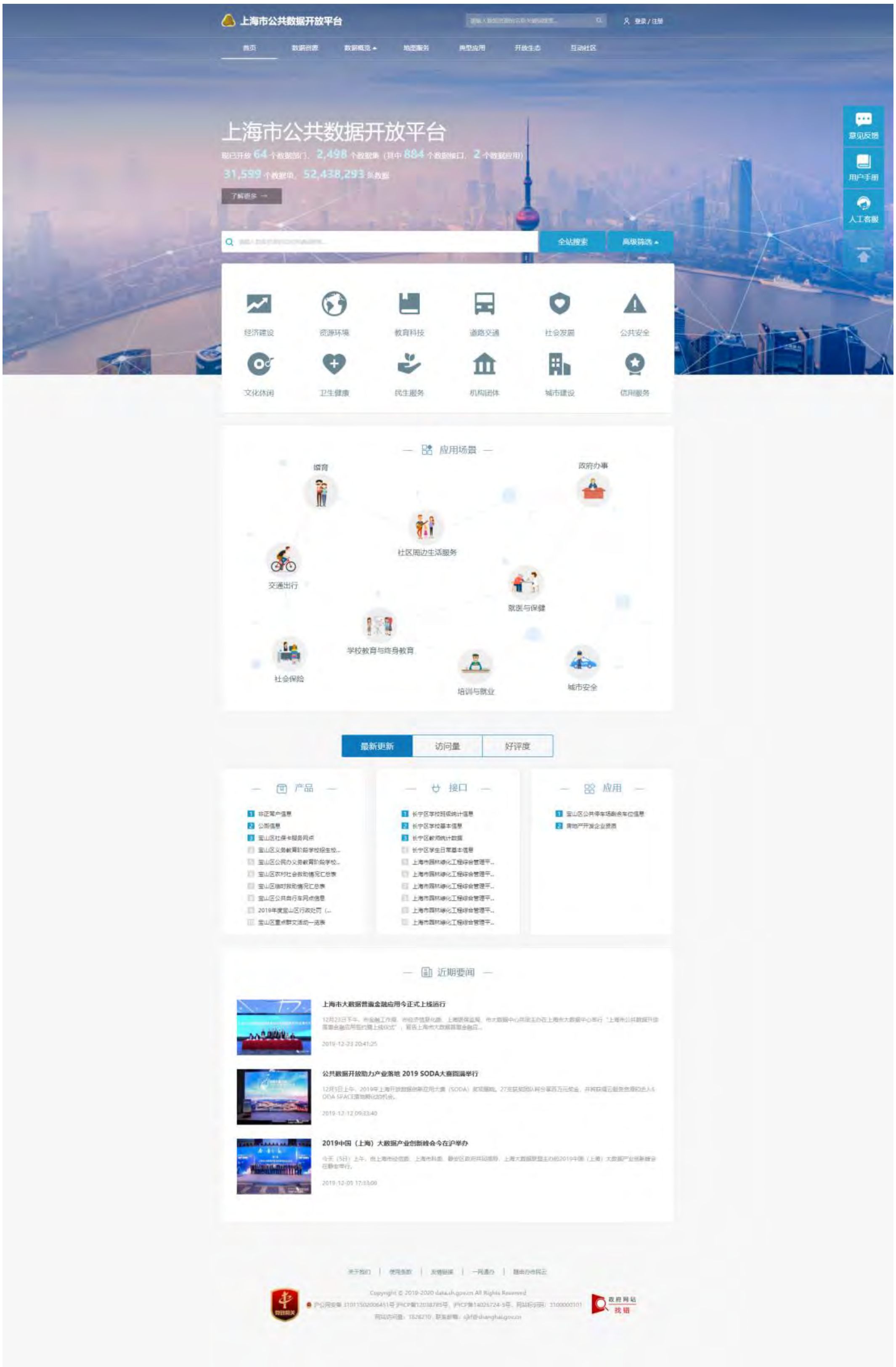
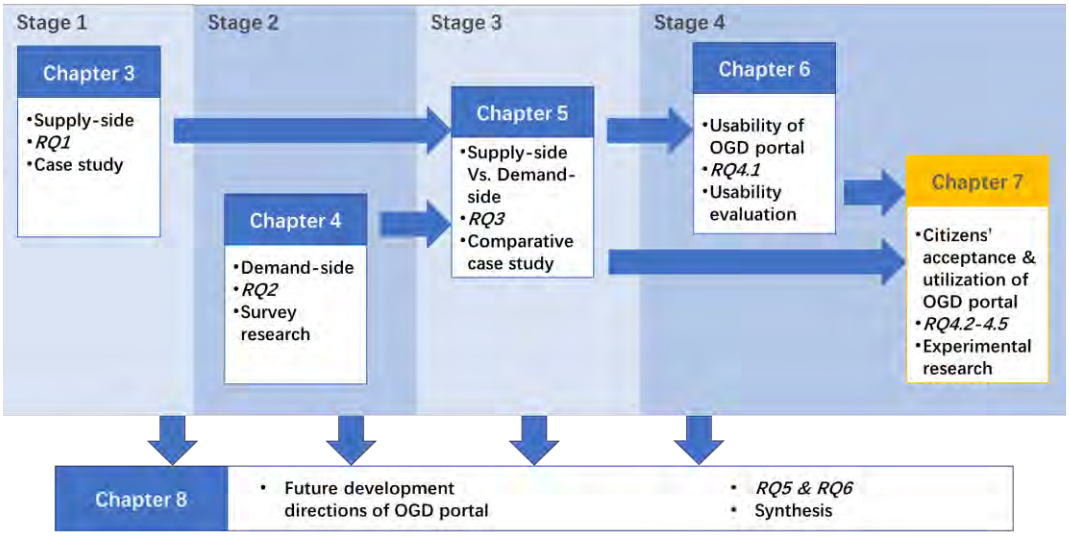


Figure 6-1 Screenshot of the Shanghai OGD portal

7 CHAPTER 7: CITIZENS' ACCEPTANCE AND UTILIZATION OF OGD PORTALS: AN EXPERIMENT USING A VIRTUAL ASSISTANT



7.1 Introduction

In this chapter we drill down deeply into possible causes and solutions to address the underutilization of OGD portals based on the results from the previous chapters. Following on from the past chapter, we want to ensure that the OGD portal with the best usability is actually usable by citizens. We thus posed the research question:

RQ4.2 Can citizens use the OGD portal with the best usability from the usability evaluation?

After ensuring that the government launches a usable OGD portal to the public, the next consideration is whether it will be accepted and utilized by the citizens. We tackle this from the perspective of technology acceptance, which has not been considered in previous chapters or stages of the thesis project.

This led to two further research questions for this chapter:

RQ4.3 What are the factors affecting citizens' acceptance of the OGD portal with the best usability from the usability evaluation?

RQ4.4 What are the factors affecting citizens' actual utilization of the OGD portal with the best usability from the usability evaluation?

According to the results of Chapter 5, in order to solve the present conflicts between the supply-side and demand-side of OGD, we need to: (1) improve online visibility; (2) keep the data up to date and perform regular updates; and/or (3) continually release more datasets to the public. But since we are not a government organization, these three aspects could not be tested in our experiments. However, it was deemed feasible for us to evaluate recommendation (4): improve help functions and provide online smart agents to assist usage. This we sought to achieve by adding help functions and creating a virtual assistant to provide support. Investigating the value of adding these functions, in particular the virtual assistant, led to our final research question for this chapter:

RQ4.5 What is the effect of using traditional help functions compared to using a virtual assistant?

To answer RQ4.2-4.5, we conducted an experiment involving two versions of a simulated portal based on the portal with the highest usability from our heuristic evaluation. One version (the control) included all the current features identified in the top-performing OGD portal from the heuristic evaluation. The second version (the experimental condition) additionally included a virtual assistant to provide support for the user. We wanted to see if the virtual assistant would increase acceptance. More information about virtual agents could be found in Section 2.8. The experiment details are reported in Section 7.2, followed by results and analysis in Section 7.3 and discussions in Section 7.4. Chapter conclusions appear in Section 7.5.

7.2 Methodology

To answer the research questions that are specified in the start of this chapter, an experiment has been designed and carried out. This section presents in detail the methodology used in this experiment, starting with the theoretical model in Section 7.2.1 and the experimental design in Section 7.2.2, followed by the experimental materials, instruments and procedures in Sections 7.2.3, 7.2.4 and 7.2.5. We then describe the pilot study and its results in Section 7.2.6, together with refining of the experiment based on the results of the pilot study. The recruitment for the formal experiment is explained in Section 7.2.7.

7.2.1 Theoretical model

An integrated research model has been built, as shown in Figure 7.1. The research model is based on four theories relating to the adoption of a technology: the Technology Acceptance Model (TAM) (Venkatesh & Davis, 2000); Diffusion of Innovation (DOI) (Rogers, 2010); trust (Carter & Bélanger, 2005; McKnight et al., 2002); and rapport (Tickle-Degnen & Rosenthal, 1990). Each of these theories and concepts were individually presented in the literature review (Chapter 2).

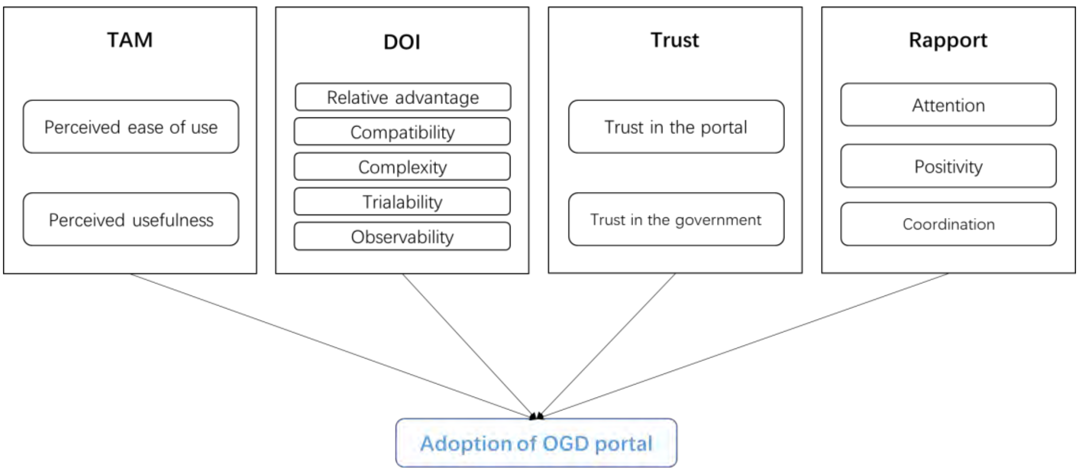


Figure 7-1 An integrated research model

In the previous chapter we presented a research model based on DOI that connected supply and demand. As shown in Figure 7.1, we have here included all five factors from DOI: relative advantage, compatibility, complexity, trialability and observability, Using that model, it was found that supply and demand did not match and that, despite the existence of OGD portals and even when citizens were aware of them, they chose not to use them. Thus, we want to understand what factors influence the (lack of) acceptance of this technology. For this purpose, we draw on a well-accepted theory known as the TAM. As shown in Figure 7.1, we included PEOU (perceived ease of use) and PU (perceived usefulness) as supplements to DOI. These two aspects of TAM are included for analyzing specific functions rather than the whole OGD portal. PEOU refers to the degree of effort required in using a certain function and PU refers to the degree to which the function can enhance the user's task performance.

In our model, we also consider the role of (lack of) trust because citizens' perception of trustworthiness can impact on their intention to use e-government services (Carter & Bélanger, 2005). Our model considers both trust in the portal and trust in the government, based on the trust model of McKnight et al. (2002). According to their trust model, because institution-based trust and disposition to trust are premises for users' trust intentions, including citizens' trust in the government as the institution-based trust and trust in the portal as the disposition to trust could thus help us understand citizens' trust intentions towards the OGD portal, which could further affect their intention to use the portal.

Finally, because RQ4.5 concerns the impact of adding a virtual assistant, we include rapport. Rapport is one of the important factors in a good conversational partnership (WEI-ERN, 2012). It is proposed by scholars that virtual agents can help establish a good sense of rapport through face-to-face interaction with the user, which helps to develop trust, liking and respect between the participants (Gratch et al., 2006). Thus, rapport is a key factor for a conversational virtual agent. According to Tickle-Degnen and Rosenthal (1990), rapport is complex and can be broken down further into three factors: attention, positivity and coordination. See Section 2.4 in the literature review for further explanation of each of the model elements.

7.2.2 Experimental design

The aims of the experiment were threefold: firstly, to analyze end-users' acceptance, trust and rapport with the present OGD portal; secondly, to analyze end-users' acceptance, trust and rapport with the OGD portal with a virtual agent; and thirdly, to compare the effect of the present portal and the virtual agent on end-users. A between-subjects design was used in our experiment with one factor (virtual agent/no virtual agent). This experimental study has been approved by the Macquarie University Human Research Ethics Committee (see Appendix A for the ethics approval letter; approval number 5201800056).

Within-subjects design and between-subjects design are commonly employed by empirical studies (Keren & Raaijmakers, 1988). The decision about the choice of design types usually considering three aspects: psychological, statistical (Greenwald, 1976), and the particular scientific problems that are supposed to be answered by the experiment (Grice, 1966). In between-subjects design, the same subject is exposed to different conditions and "there is a substantial difference between stimuli employed in different experimental conditions" (Keren & Raaijmakers, 1988, p. 234). While the same subject in within-subjects design may experience the same (or very similar) stimulus more than once. Despite the statistical efficiency afforded by removing subject variance from error terms used to test treatment effects, within-subjects designs are often faulted for being subject to context effects which may limit interpretation of results (Greenwald, 1976). While between-subjects designs are not devoid of context effects (Greenwald, 1976), but the economy in number of subjects is higher than within-subjects designs. In addition, within-subjects designs are more suitable for investigating learning effects due to repeated trials (Keren & Raaijmakers, 1988), which is not the situation of our study. Considering one of the key purposes of this experiment is to find citizens' preferences of traditional help functions versus virtual assistants, the nature of this experiment is more related to the choice of behavior of users, which, indicated by previous studies, is suitable for using the between-subjects designs to test the reflection effect of the treatment.

The structure of the experimental design is shown in Figure 7.2. In the experiment, the participants were randomly assigned by the Qualtrics survey software into one of two different groups: Group 1, which was the control group, used the simulated portal derived from Chapter 6 comprising conventional help functions and ways of obtaining help (i.e. reading start instructions and FAQs and filling in request forms online). Group 2, which was the experimental group, used the simulated portal derived from Chapter 6 but with the help functions replaced with a virtual agent to provide support through conversations. Participants in each group were given the same tasks in the same order to complete by using the portal provided to them. For experimental control purposes, a minimum time limit was set for the participants to use the portal to solve the tasks based on the results from the pilot study. Data was collected before and after the participants completed the problem tasks through instruments, which are further explained in Section 7.2.5.

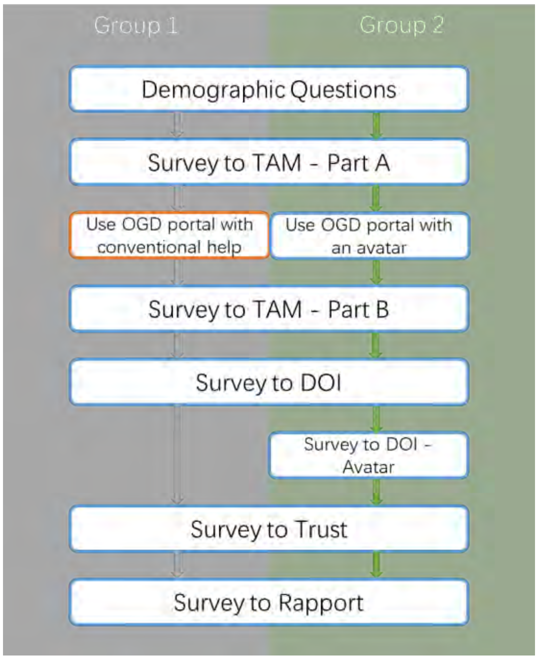


Figure 7-2 Experimental design

7.2.3 Experimental materials

The experiment required the use of three main materials: a simulated OGD portal, a virtual agent offering help functions and an online survey tool to direct participants to complete the experiment. In this section, we introduce how we built our experimental environment through these three main materials.

7.2.3.1 Imitated OGD portal

Before building the imitated portal, we first needed to choose a target to imitate. Since our previous studies were carried out in China, we decided to select one OGD portal in China as our imitation target. It would not be a fair comparison to imitate a less developed portal if there were already better ones in existence. Choosing the best-performing OGD portal ensured that our experiment reflected the present capabilities of OGD portals. According to the usability evaluation results presented in Chapter 6, among all province-level OGD portals in China the Shanghai portal had the topmost ranking, covering 46 out of 54 usability criteria. Thus, we chose the Shanghai portal as our target OGD portal to imitate for the experiment.

After deciding on the imitation target, we then developed the OGD portal (see Figure 7.3) for the experiment in Eclipse using Java, HTML and CSS. All functions in the Shanghai portal relating to the usability criteria were built into the experiment portal. We then added 100 datasets of 10 different data categories (local statistics, health, education, cultural activity, transportation, public safety, environment quality, registration, budget and spend, and credit records) and 10 data interfaces to the experiment portal. All the datasets and interfaces were selected from the Shanghai portal. We randomly selected 10 datasets from each of the data categories on the Shanghai portal that would not show the area information to the experiment participants. This was to ensure that if a participant was from Shanghai, they would not already know the answer. In other words, we sought to ensure that participants did not know that the questions were about

Shanghai, therefore they could not answer any question, even if they knew the answer, without searching the portal.

Figure 7-3 Home page of the imitated portal

7.2.3.2 Conversational virtual agent

A conversational virtual agent called Xiao Zheng was built for the experimental group to offer help functions on the imitated OGD portal. Firstly, we used Fuse to build a 3D character. According to the results in Chapter 4, citizens show preference for female virtual agents rather than males. Thus, we decided to build an Asian-looking female figure as our virtual agent (see Figure 7.4). Then we used Mixamo to rig and add body and facial animation to the figure. The rigged figure is imported into Unity 3D to complete further development.

The second part for the virtual agent was the dialogue files. We transferred four different kinds of help functions on the imitated OGD portal into dialogues for the virtual agent: introduction for new users, FAQs, leaving feedback and applying new datasets (details of these dialogues can be found in Appendix C). In the control group, these four functions were offered by pictures (introduction for new users), written paragraphs (FAQ) or online tabular forms (leaving feedback and applying new datasets). We changed the information provided by these functions into conversational dialogues for the virtual agent. After the virtual agent speaks with the user, several responses were designed from which the user cand choose to continue the dialogue. A separate dialogue file was created for each function. We made sure that no additional information was given through the dialogues besides the content provided by the original help functions of the imitated portal. The dialogues were reviewed by three university reference librarians to make sure they were friendly and logical. After confirming the dialogues for the virtual agent, we then recorded automated Text to Speech (TTS) voices for each continuous speech of the virtual agent. The Chinese language voice pack of Office 365 was used as the voice source for all the dialogues.

Finally, the character together with her animation were combined with the dialogue files, voices and background image in Unity 3D. Four separate WebGL programs were created in Unity 3D for the four different help functions. The interface of these programs is shown in Figure 7.4. The workflow of these programs is as follows: the user clicks the start button to start a conversation with the virtual agent. After the virtual agent gives her first response, several selections are given to the user according to the dialogue file. Then the virtual agent gives her next responses according to the different choices of the user, until she reaches the end of the conversation. The user's selection for each choice during the conversation is sent to the MySQL database via a PHP command. Capture of interaction data with the virtual agent allowed us to potentially re-create the interactions and analyze conversation patterns. After creating the conversational virtual agent, we then created the portal for the experimental group by deleting the four kinds of help functions from the portal and adding the four WebGL programs to the portal accordingly. When the users clicked the link on the page for help functions, a new window would pop out with the WebGL program running on it showing the virtual agent. After they clicked the "Start" button on the window, they could start their conversation with the virtual agent.

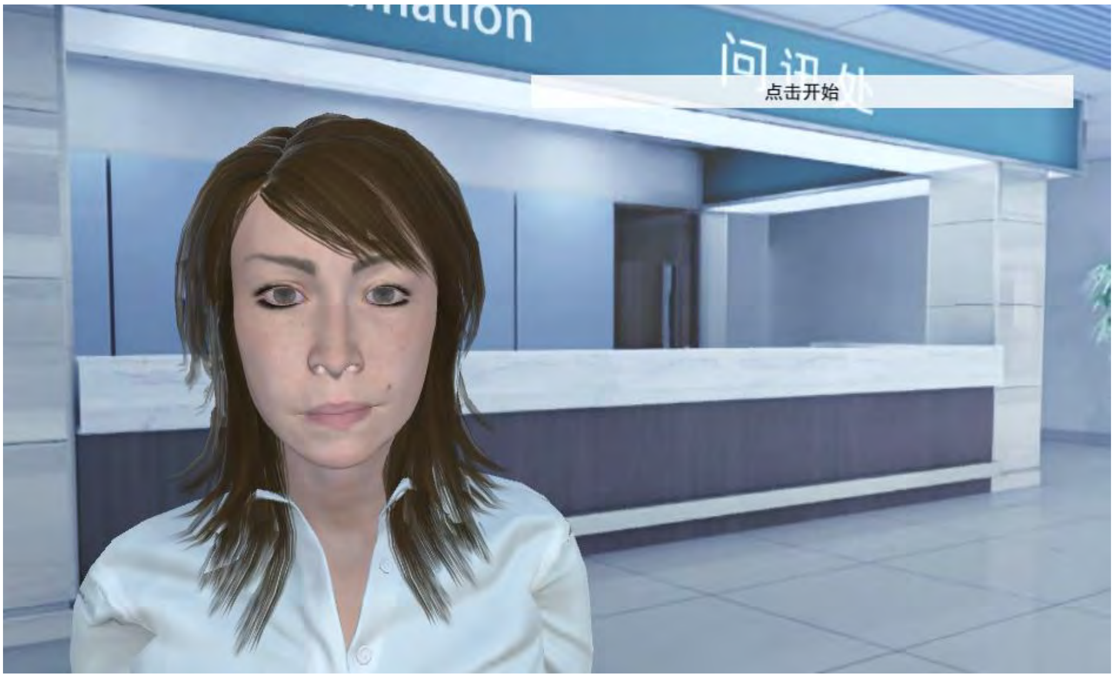


Figure 7-4 Interface for the conversational virtual agent

7.2.3.3 Experimental environment

Qualtrics, which is research data collection software, was used to build and control the whole experimental environment as well as to collect data for both the control group and the experimental group. We first created all the survey instruments for the experiment in Qualtrics, as further explained in Section 7.2.4. Then the block for guiding the participants to the imitated portals for each group was designed in Qualtrics. We further designed the block for the tasks the participants needed to complete on the imitated portal. The experiment was first designed with four different tasks for participants to complete. All of these tasks were about finding related datasets on the portal and using the information in the datasets to answer certain questions. These questions were selected because participants could not answer them according to their knowledge and thus they had to search and use the datasets on the portal. These questions were:

- What's the top first cause of death in 2016?
- How many kindergartens are there in this area?

- How many fishing locations are there in this area?
- What's the number of embassy certifications in the first half year of 2016?

Three out of these four questions had answers that participants could find on the portal. But the last one was designed to be a question for which participants could not find any related datasets on the portal. As part of the instructions for the task, if the participants could not find the answer to a question, they were required to submit a data request on the portal asking for related datasets. A timer for 10 minutes was set in this block to enforce a minimum time the participants needed to spend on the portal. The participants could not move onto the next step of the experiment until the timer counted down to zero. The randomizer in Qualtrics was used for randomly allocating the participants to the control group or the experimental group evenly.

7.2.4 Instruments

We used four different survey instruments in this experiment according to our research model, namely TAM, DOI, trustworthiness and rapport. The experiment also included a demographic questionnaire for collecting background information on each participant. All the instruments were created in Qualtrics (the complete instruments can be found in Appendix D). Detailed information about each instrument is explained below.

7.2.4.1 Demographic questionnaire

The demographic questionnaire included five parts. In the first part, we asked the participants for their gender, age, educational qualification, occupation and monthly personal income (in China individuals usually report income before tax). The second part was about participants' knowledge of OGD and OGD portals. We asked the participants whether they had ever used OGD and OGD portals before. If they had, we further asked them about how they knew about OGD and OGD portal, what types of data they had used, the utilization frequency and their aim of utilization. In the third part of the questionnaire, we asked the participants whether they played computer games or not and their average time of playing computer games per week. We wanted to be able to measure whether their familiarity with games technology influenced their responses to virtual agents.

In the fourth part of the questionnaire, we included the ten-item personality inventory (TIPI) of the Big Five personality dimensions (Gosling, Rentfrow, & Swann Jr, 2003), which is a commonly accepted instrument for collecting data on participants' personalities. Personality is a common marker of individual differences and we wanted to be able to capture personality data in case there was a relationship between personality and the responses given. The final part of the demographic questionnaire contained three questions to test participants' trust intentions (McKnight et al., 2002). For the fourth and final part of the questionnaire, the randomizer function of Qualtrics was used for presenting the questions.

7.2.4.2 TAM questionnaire

TAM states that a user's attitude towards the utilization of a system is influenced by PU and PEOU (Davis, 1989). To test participants' attitudes towards the imitated portal, we included questions asking about their PU and PEOU for 12 different functions of the portal after the participants had used the portal to complete tasks including browsing, choosing different formats, downloading without registration, filtering search results, help functions, keyword search, metadata, open and machine-readable format, ranking, requesting new datasets, giving feedback and visualizing for quick viewing. In order to maintain consistency for this thesis, we choose to use the functions that were included in the usability principles discussed in Chapter 6. For participants of the experimental group, "help functions" were replaced by "avatar". In order to make comparison of participants' attitudes towards these functions before and after their use of the portal, we also added a group of questions asking their feeling about the importance of these functions before their use of the imitated portal as a supplement to the two original TAM questions.

All of these questions used 5-point Likert scales from “Not at all useful” (1) to “Extremely useful” (5), from “Extremely difficult” (1) to “Extremely easy” (5) and from “Not at all important” (1) to “Extremely important” (5). All of these three groups of questions were presented to the participants in an in-group random order by using the Qualtrics randomizer function.

7.2.4.3 DOI questionnaire

The DOI theory explains users’ adoption of new technologies from five perspectives: relative advantage, compatibility, complexity, trialability and observability. The objects treated as new technologies in our experiment were twofold: the imitated OGD portal, and the conversational virtual agent. For both the control group and the experimental group, we included a DOI questionnaire for the OGD portal. For the experimental group, an additional DOI questionnaire for the conversational virtual agent was included. Both of these DOI questionnaires had questions covering the five aspects of DOI developed from the study of Atkinson (2007) by changing the objects to OGD portals and the virtual agent. The questionnaire for the OGD portal had 27 questions, excluding one question for trialability and three questions for observability because they did not fit this experiment, and adding one question on complexity regarding participants’ difficulty in understanding what OGD is. Similarly, the questionnaire for the virtual agent had 26 questions, excluding one question for trialability and three questions for observability because of unsuitability for this experiment. All the questions in these two questionnaires used a 6-point Likert scale from “Strongly disagree” (1) to “Strongly agree” (6). Questions were presented to the participants in an in-group random order by using the Qualtrics randomizer function.

7.2.4.4 Trustworthiness questionnaire

To test participants' perceptions of the trustworthiness of OGD and OGD portals, we included a trust questionnaire in the experiment based on the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1977). The questions in the questionnaire were developed from the study of Carter and Bélanger (2005) to fit the circumstance of OGD and OGD portals. All the questions used a 6-point Likert scale from "Strongly disagree" (1) to "Strongly agree" (6). Questions were presented to the participants in an in-group random order by using the Qualtrics randomizer function.

7.2.4.5 Rapport questionnaire

To evaluate participants' rapport with the virtual agent in the experimental group and with the help functions in the control group, we included two versions of the rapport questionnaire in the experiment for each group. Both questionnaires had 22 questions developed from the study of Ranjbartabar (2016) which reflected the three components of rapport including attention, positivity and coordination. For the version for the experimental group, questions were asked about participants' rapport with the virtual agent. For the version for the control group, questions were asked about participants' rapport with the help pages. All the questions used a 6-point Likert scale from "Strongly disagree" (1) to "Strongly agree" (6). Questions were presented to the participants in an in-group random order by using the Qualtrics randomizer function.

7.2.5 Experimental procedure

The experimental procedure is shown in Figure 7.5. After participants read the participant information and consent form and decided to participate, they first answered the demographic questionnaire. Then by using the randomizer in Qualtrics, the participants were allocated randomly and evenly to either the control group or the experimental group. For the control group, the participants answered the first part of survey 2 about the importance of functions on the portal, then they were informed of the instructions for completing tasks on the portal and guided to the imitated OGD portal with traditional help functions after the participants clicked that they were ready to complete the tasks. Each participant was given the same four tasks in the same order for them to solve on the portal. A timer of 10 minutes started to count down once the tasks were shown to the participants. They could not go onto the next session until the timer counted to zero. After 10 minutes and completion of the tasks, participants went on answering the second part of survey 2 about the PU and PEOU of TAM, survey 3 about the DOI of the OGD portal, survey 4 about the trustworthiness and survey 5 about the rapport with the help functions.

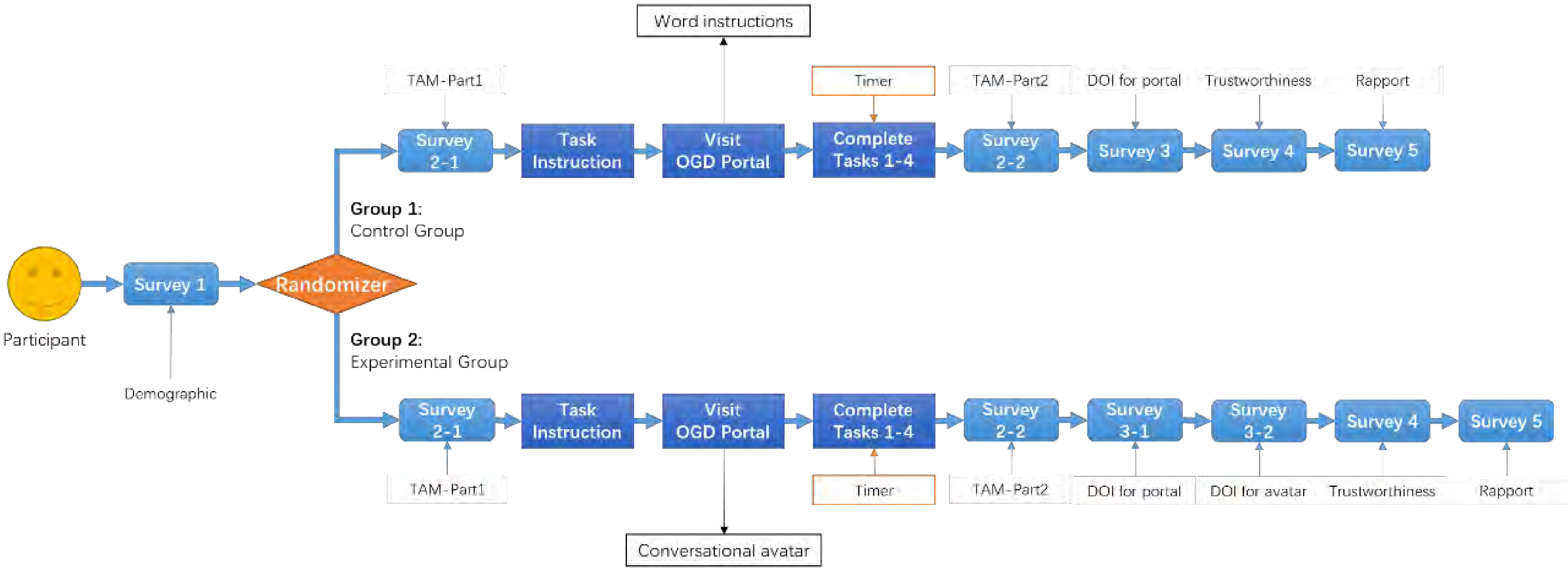


Figure 7-5 Experimental procedure

For the experimental group, the participants also first answered the first part of survey 2 about the importance of functions on the portal. Then they also got the instructions for how to complete the tasks on the portal. After they selected that they were ready to complete the tasks, they were guided to the OGD portal with a virtual agent offering help. All participants were given the same four tasks in the same order as the control group. After the participants spent 10 minutes and completed the tasks given to them, they went on answering the second part of survey 2, survey 3-1 about the DOI of the OGD portal, survey 3-2 about the DOI of the virtual agent, survey 4 about the trustworthiness and survey 5 about the rapport with the virtual agent.

7.2.6 Pilot study

To test the experimental design and effect, we carried out two pilot studies before the formal experiment. After the pilot studies, some problems were recognized with the original experimental design. Changes were made to the experimental design and process for the formal experiment. Details of the pilot studies and changes made to the experiment are explained in this section.

7.2.6.1 Pilot study I

Pilot study I was carried out according to the original experimental design explained in the previous sections. In total, 8 participants volunteered for this pilot study. The participants were all Chinese-speaking PhD students recruited from the Department of Computing at Macquarie University. We required them to complete the procedures for both the control group and the experimental group. We wanted to ensure that there were observable differences between the delivery of help via the OGD portals with and without the virtual agent. This was an opportunity to gain within-subject data useful for initial comparison of the two treatments without requiring participants in the main study to experience both conditions, which could also have learning and order effects that biased the findings. 5 of them completed the parts for the control group first and then the experimental group, while the other three completed them in the other sequence. For the 5 participants who completed the control group first, 2 of them withdrew after completing the parts for the control group before moving onto the experimental group's part. Thus, finally, 6 of them completed the whole study including 4 females and 2 males.

For the 6 participants who completed the whole pilot study, we added 10 questions for them to compare the two imitated OGD portals: one with the ordinary help functions, and the other with the conversational virtual agent. Three of them preferred the ordinary help functions to the conversational virtual agent, two showed no preference and only one preferred the conversational virtual agent to the ordinary help functions. Their reasons for not preferring the virtual agent were that she spoke too slowly and wasted their time. They preferred to quickly read the textual dialogue on the screen. They also felt embarrassed when having conversations with the virtual agent. The reasons for the one participant preferring the virtual agent was that she was more helpful. 4 out of 6 of these participants felt the OGD portal with ordinary help functions was much easier to use, because it was much faster than having conversations with a virtual agent. One of them felt the portal with the virtual agent was easier to use because it was easier to get help from the virtual agent. We also asked the participants which one they would prefer to use in future; 3 of them preferred the one with ordinary help functions because it saved time and did not have sound. 2 showed no preference. One preferred the portal with the virtual agent because the virtual agent was more helpful. Finally, we asked the participants about the advantages and disadvantages of these two portals. The portal with ordinary help functions had advantages in saving time and being quiet but had little interaction with the users, while the portal with the virtual agent had an advantage of interaction with the users but took a longer time.

The major issues with the experimental design according to the participants were threefold. Firstly, the experiment needed a very long period of time to complete, due to the inclusion of many questionnaires. This problem was greatly amplified because the pilot participants needed to complete most of the questionnaires twice in order to test both the control group and the experimental group. Secondly, the participants thought 10 minutes was too long for them to complete the tasks. According to the records of their clicking behavior, most of them spent around 5 minutes to complete the four tasks given to them. Thus, they had to wait another 5 minutes before moving on to the next section. But when they came to the second portal for the other experimental group, they did not search the datasets again on the portal. Instead, they simply clicked the answers according to their memory. Thirdly, the experiment needed the participants to use specific browsers (Chrome, Firefox or Safari) because the virtual agent runs in WebGL which is supported by only a few browsers. We also examined the reliability of all the scales in the experiment. The results for Cronbach's alpha were all above 0.55. The trustworthiness questionnaire and rapport questionnaire had lower values which may be due to the limited number of participants. The records of participants' behaviour also showed that they did not use the help functions like the introduction for new users and the FAQ. Some of them also did not submit data applications even though they could not find related data.

Therefore, we made changes to the experimental design accordingly before carrying out Pilot study II. Firstly, we changed the timer from 10 minutes to 5 minutes. Secondly, when the participants were guided to the portal, the first page shown to them was the introduction for new users, instead of the homepage of the portal. Thirdly, we changed the tasks by adding questions related to the FAQ. Because the issue of the length of the experiment was amplified by having to complete questionnaires twice, we believe all questions were relevant to the study according to our research model and the timer set for the experiment previously also extended the length of the experiment, we chose not to delete any questions.

WebGL is the best way we could find to combine a virtual agent with a website and Chrome, Firefox and Safari are very popular and commonly accepted browsers, thus we chose to keep the virtual agent running in WebGL. But we added special instructions in the experiment to remind the participants to use the proper browser for this experiment.

7.2.6.2 Pilot study II

To test the changes we made to the experiment after Pilot study I, we carried out Pilot study II. In this study, we focused on the section about using the portal to complete the tasks. Thus, participants only needed to follow the instructions to the portal (either the one for the control group or the one for the experimental group) and use the portal to complete three tasks given to them (one they could find the answer to by referring to the datasets on the portal, one they could not find the answer to and needed to submit a data application and the third one about information given in the FAQ). After they completed the first three tasks, they were guided to the second portal (either the one for the experimental group or the one for the control group) and completed three different tasks (one they could find the answer to by referring to the datasets on the portal, one they could not find the answer to and needed to submit a data application and the third one about information given in the FAQ). After they completed these two parts of the experiment, they were asked 10 questions about comparing these two portals. In total, 7 participants volunteered in this study, who were Chinese-speaking PhD students recruited from the Department of Computing, Macquarie University. 2 of them were volunteers from the previous study. 4 of them completed the whole study, 2 of them using the portal with the virtual agent first and then the portal with ordinary help functions, and the other 2 using the portals in the reverse order.

For this pilot study, participants spent at least 6 minutes to complete the tasks given to them. The longest time they spent on the tasks was 11 minutes. They were asked their preference between the two portals. 2 of them preferred the portal with the virtual agent because it was more attractive, helpful and friendly, while the other 2 preferred the portal with ordinary help functions because it was faster and more convenient. When asked about which one was easier to use, 3 of them chose the portal with ordinary help functions because it was faster. One of them preferred to use the portal with the virtual agent because it was easier to find answers to the tasks with the help of the virtual agent. We also asked the participants which portal they would like to use in the future; 2 of them chose the portal with ordinary help functions because it was faster. The other 2 chose the portal with the virtual agent because it was more modern and they believed virtual agents could become more intelligent in the future. The advantages of the portal with ordinary help functions were that it was faster, more direct and had all information on one page. The disadvantages were the information could be hard to read and thus it would take more time to find the answers. The advantages of the portal with the virtual agent were that it was more fun and friendly and had a speaking voice. The disadvantages were that the virtual agent would take more time for the participants to have conversations with it and required headphones to hear the voice.

7.2.6.3 Changes to the experimental design

After the two pilot studies, we made five changes to the experimental procedure according to the results of the pilot studies. Firstly, we changed the timer for the tasks from 10 minutes to 5 minutes. Secondly, the participants in both groups were guided to the introduction for new users page, instead of the homepage of the portal. Thirdly, we added icons to the portal with the virtual agent to mark the functions that were provided by the virtual agent (see Figure 7.6). Fourthly, we made some changes to the dialogues of the virtual agent to make the conversations smoother. Lastly, we redesigned the tasks for the experiment. We changed the previous 4 tasks to 5 tasks, 2 of which related to the FAQ, 2 of which the participants could find related datasets for on the portal and the last one that they could not find any relevant datasets for and needed to submit a data request.

All tasks were presented to the participants of the two groups in the same order. The order and content of these five tasks were as follows:

- Please find through “FAQ”/“Ask the agent” in “Help” if the following statement is true or not: **Fees will be claimed for business use of the data on the portal.**
- How many kindergartens are there in this area?
- Please find through “FAQ”/“Ask the agent” in “Help” if the following statement is true or not: **It is legal to resell the data downloaded from this portal.**
- What’s the top first cause for death in 2016?
- How many swimming pools are there in this area?

Figure 7-6 Homepage of the imitated portal with the virtual agent

7.2.7 Recruitment

We planned to recruit at least 50 participants for this study, with 25 for each group. Due to the focus of this thesis on the development and utilization of OGD portals in China, and that our previous studies involved data collection in China and the imitated portal was made in Chinese, we decided to recruit both male and female Chinese citizens of age 18 years and above. We sought to recruit participants from different backgrounds, such as students, company employees and government staff, to strengthen the generalizability of our study.

Since the experiment was carried out online through Qualtrics, we recruited participants indirectly by putting advertisements and announcements on the main social media sites in China including Wechat and Weibo to invite participants to join the study. A weblink to the experiment on Qualtrics was included in the advertisement. The participants could start the experiment by simply clicking the link in the advertisement. The first recruitment lasted from 18 March to 28 April 2019. Due to the fact that some participants of the experimental group did not use the virtual agent although they finished the experiment, we carried out a second recruitment to find more participants for the experimental group. The second recruitment lasted from 1 to 2 July 2019. The advertisement for the recruitment can be found in Appendix E. No payment or reward was offered for participation.

7.3 Results and analysis

In the following sections, we describe the participants of the experiment including their demographics and give the analytical results for the control group and the experimental group in relation to the TAM, DOI, trust and rapport.

7.3.1 Participants and demographics

In total, 160 participants volunteered to join the experiment. Two of them refused to continue the experiment after reading the participant information and consent form. 20 (12.5%) of them withdrew from the study before completing Survey 1, the demographic questionnaire, while 45 (28.12%) of them proceeded until they reached the point of visiting the portal and completing the tasks. Two of them (1.2%) withdrew after the completing tasks on the portal. So 91 (56.9%) of them completed the whole experiment. In the following analysis of the control group and the experimental group, we excluded the participants who did not complete the whole experiment and the participants in the experimental group who did not use the virtual agent at all.

7.3.1.1 Socio-demographic background

We analyzed the socio-demographics of all the 160 participants, as shown in Table 7.1. Participants included 71 males and 81 females with various backgrounds in relation to age, educational qualification, occupation and income. Most of the participants were 26 to 40 years old. Most of them had bachelor's degrees or master's degrees. Student, technician and teacher were the three main occupations of the respondents. Their monthly income was mainly in the range above 7000 RMB. We also analyzed the demographic distribution of the control group, the experimental group and those being excluded. Chi-square tests showed no significant differences in the distribution of gender, age, educational qualification, occupation and income among the control group, the experimental group and those being excluded. The proportions of gender and educational qualification distributions in each group were very similar. There were more participants of age 18–25 years in the experimental group and more of age 31–40 years in the control group. For the income, there were more participants with monthly income less than 2000 RMB and above 7000 RMB in the experimental group, and more participants with monthly income of 3001 to 7000 RMB in the control group.

Table 7-1 Socio-demographics of the participants

Topic	Dimension	Total	Percent	Ctrl Grp.	Percent	Exp Grp.	Percent	Excluded	Percent
Gender	Male	71	44.4%	15	39.5%	19	46.3%	37	45.7%
	Female	81	50.6%	23	60.5%	21	51.2%	37	45.7%
	Other	1	0.6%	0	0.0%	1	2.4%	0	0.0%
	Missing	7	4.4%	0	0.0%	0	0.0%	7	8.6%
Age	18 - 25	22	13.8%	3	7.9%	10	24.4%	9	11.1%
	26 - 30	41	25.6%	13	34.2%	12	29.3%	16	19.8%
	31 - 40	37	23.1%	11	28.9%	8	19.5%	18	22.2%
	41 - 50	25	15.6%	8	21.1%	6	14.6%	11	13.6%
	51 - 60	28	17.5%	3	7.9%	5	12.2%	20	24.7%
	Missing	7	4.4%	0	0.0%	0	0.0%	7	8.6%
Educational Qualification	Junior high	1	0.6%	0	0.0%	0	0.0%	1	1.2%
	Senior high	10	6.3%	4	10.5%	5	12.2%	4	4.9%
	Bachelor's degree	40	25.0%	22	57.9%	24	58.5%	41	50.6%
	Master's degree	21	13.1%	10	26.3%	10	24.4%	24	29.6%
	PhD.	4	2.5%	2	5.3%	2	4.9%	4	4.9%
	Missing	7	4.4%	0	0.0%	0	0.0%	7	8.6%
Occupation	Student	26	16.3%	5	13.2%	7	17.1%	14	17.3%
	Production worker	5	3.1%	2	5.3%	1	2.4%	2	2.5%
	Marketing/Salesman	11	6.9%	4	10.5%	2	4.9%	5	6.2%
	Customer service	1	0.6%	1	2.6%	0	0.0%	0	0.0%
	Logistics	12	7.5%	4	10.5%	3	7.3%	5	6.2%
	Human resources	7	4.4%	2	5.3%	1	2.4%	4	4.9%
	Financial/auditor	7	4.4%	4	10.5%	1	2.4%	2	2.5%
	Civilian post	11	6.9%	1	2.6%	5	12.2%	5	6.2%
	Technician	19	11.9%	2	5.3%	6	14.6%	11	13.6%
	Manager	12	7.5%	1	2.6%	5	12.2%	6	7.4%
	Teacher	16	10.0%	3	7.9%	5	12.2%	8	9.9%
	Consultant	5	3.1%	1	2.6%	3	7.3%	1	1.2%
	Specialist	8	5.0%	4	10.5%	1	2.4%	3	3.7%
	Other	13	8.1%	4	10.5%	1	2.4%	8	9.9%
	Missing	7	4.4%	0	0.0%	0	0.0%	7	8.6%
Montly Income	Below 2000RMB	17	10.6%	1	2.6%	8	19.5%	8	9.9%
	2001-3000RMB	17	10.6%	5	13.2%	3	7.3%	9	11.1%
	3001-5000RMB	33	20.6%	10	26.3%	6	14.6%	17	21.0%
	5001-7000RMB	23	14.4%	9	23.7%	3	7.3%	11	13.6%
	Above 7000RMB	63	39.4%	13	34.2%	21	51.2%	29	35.8%
	Missing	7	4.4%	0	0.0%	0	0.0%	7	8.6%
Total		160	100.0%	38	23.8%	41	25.6%	81	50.6%

7.3.1.2 Knowledge of OGD and OGD portals

In the demographic questionnaire, we asked the participants about their knowledge of OGD and OGD portals. Results in Table 7.2 show that more than half of the participants had never used OGD or OGD portals before. More participants had used OGD before than had used OGD portals before. There were more participants in the control group who had used OGD and OGD portals before than in the experimental group, but the difference was not statistically significant with chi-square tests.

Table 7-2 Participants' knowledge of OGD and OGD portals

Topic	*Dimension*	*Freq.*	*Percent*	*Ctrl Grp.*	*Percent*	*Exp Grp.*	*Percent*	*Excluded*	*Percent*
Ever use OGD before	Yes	58	37.9%	18	47.4%	17	41.5%	23	31.1%
	No	95	62.1%	20	52.6%	24	58.5%	51	68.9%
Total		153	100.0%	38	100%	41	100%	74	100%
Ever use OGD Portal before	Yes	45	30.4%	16	42.1%	9	22.0%	20	29.0%
	No	103	69.6%	22	57.9%	32	78.0%	49	71.0%
Total		148	100.0%	38	100%	41	100%	69	100%

For the channels of the participants knowing of OGD and OGD portals, we can see from the results in Figure 7.7 that most of them had obtained the information through searching online, followed by hearing from the news and friends.

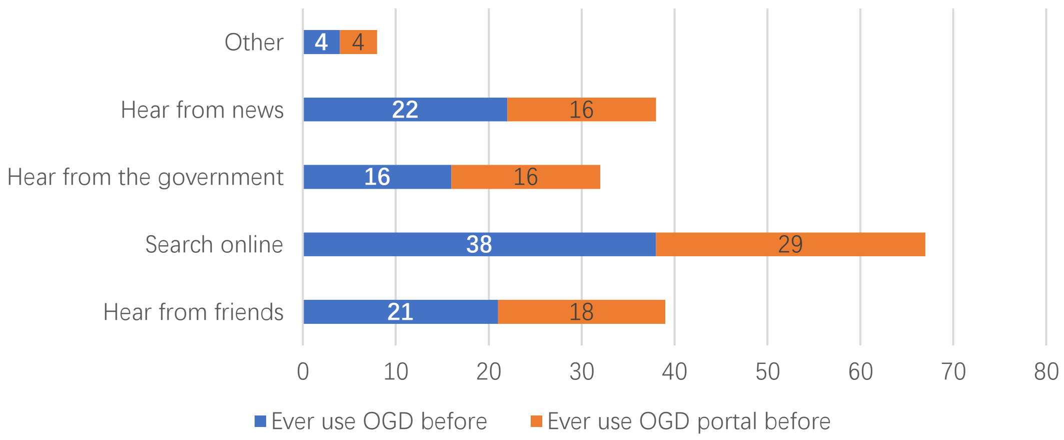
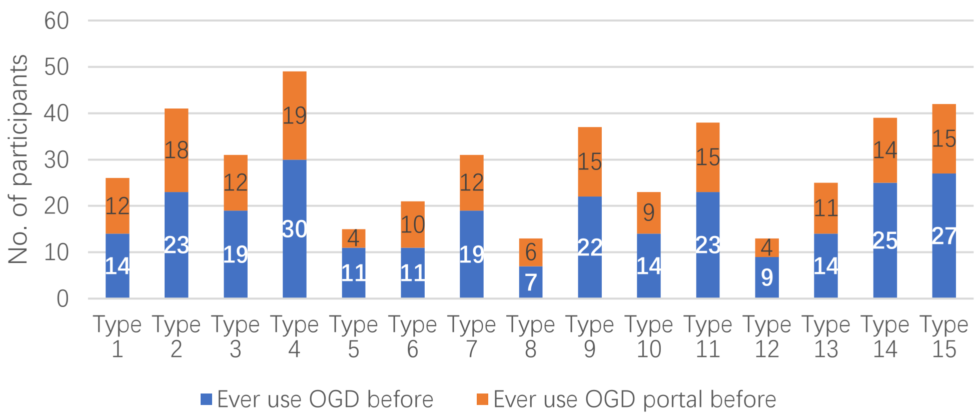


Figure 7-7 Methods of knowing about OGD and OGD portals

For those participants who had used OGD or OGD portals before, we asked them the types of data they had used. According to the results in Figure 7.8, education data had been used by most participants, followed by weather and credit records. International trade, public safety and environment quality were the three types that received least attention.



1=Budget & Spend, 2=Credit Records, 3=Cultural Activity, 4=Education, 5=Environment Quality, 6=Government Bid, 7=Health, 8=International Trade, 9=Local Statistics, 10=Maps, 11=Policies & Legislation, 12=Public Safety, 13=Registration, 14=Transportation, 15=Weather

Figure 7-8 Types of data used

We also asked the participants how often they had used OGD and OGD portals. As we can see from Figure 7.9, the usage frequency for OGD was higher than for OGD portals. Most participants had used OGD and OGD portals more than once but their usage frequency was very low, mainly "More than once but rarely use".

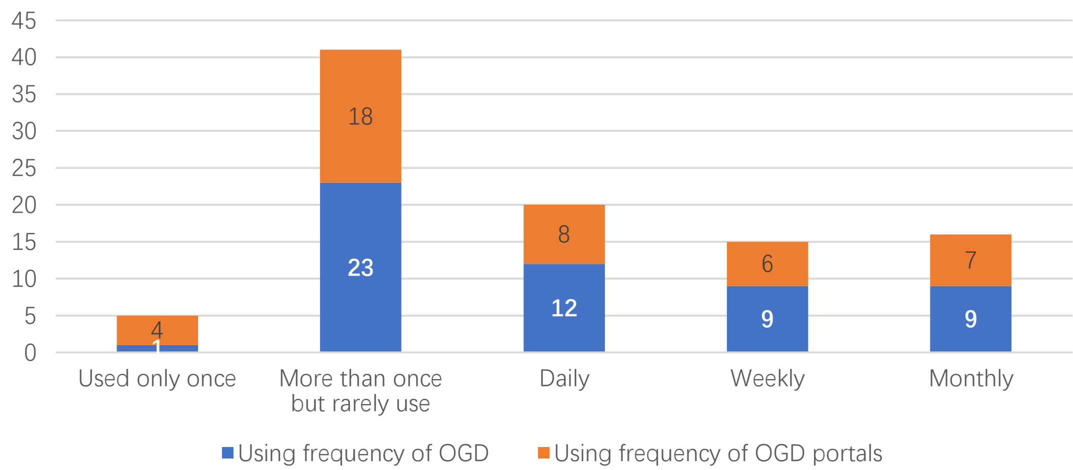


Figure 7-9 Usage frequency of OGD and OGD portals

Finally, we asked participants for their purpose of using OGD. As we can see from the results in Figure 7.10, most participants used OGD for daily life, followed by scientific research and business decision(s). Software development was the least used purpose for OGD. Other purposes of use included teaching and study.

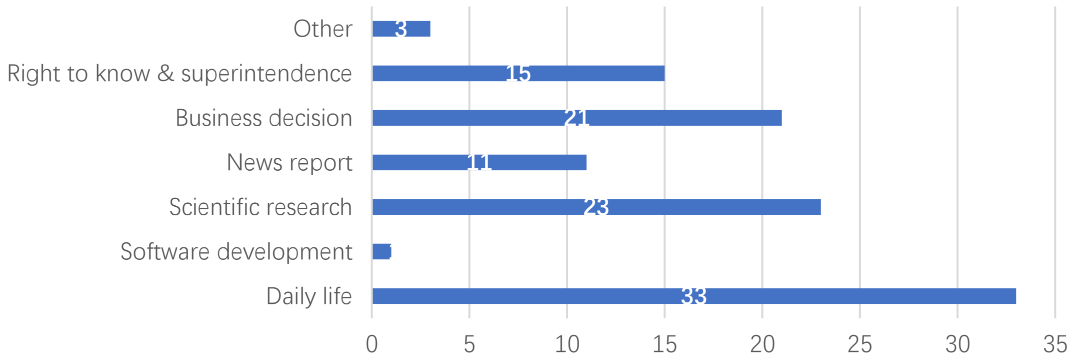


Figure 7-10 Purposes of use of OGD

7.3.1.3 Playing of computer games

Participants' experience of playing computer games might have affected their adoption of OGD portals and virtual agents. Thus, we asked participants whether they played computer games or not and their average hours spent on playing computer games per week. Figure 7.11 shows the frequency distribution of the time participants spent on computer games per week; 74 (48.75%) of the participants chose not to play computer games. For those who played computer games, more than half of them (59.8%) played for less than 7 hours per week. The average time spent on computer games was 12.95 hours per week. For the control group, there were 18 participants who played computer games and 20 who did not. For the experimental group, there were 26 participants who played computer games and 15 who did not. We noted that there were more participants who played computer games than those who did not in the experimental group. But the difference between the control group and the experimental group was not statistically significant with a chi-square test.

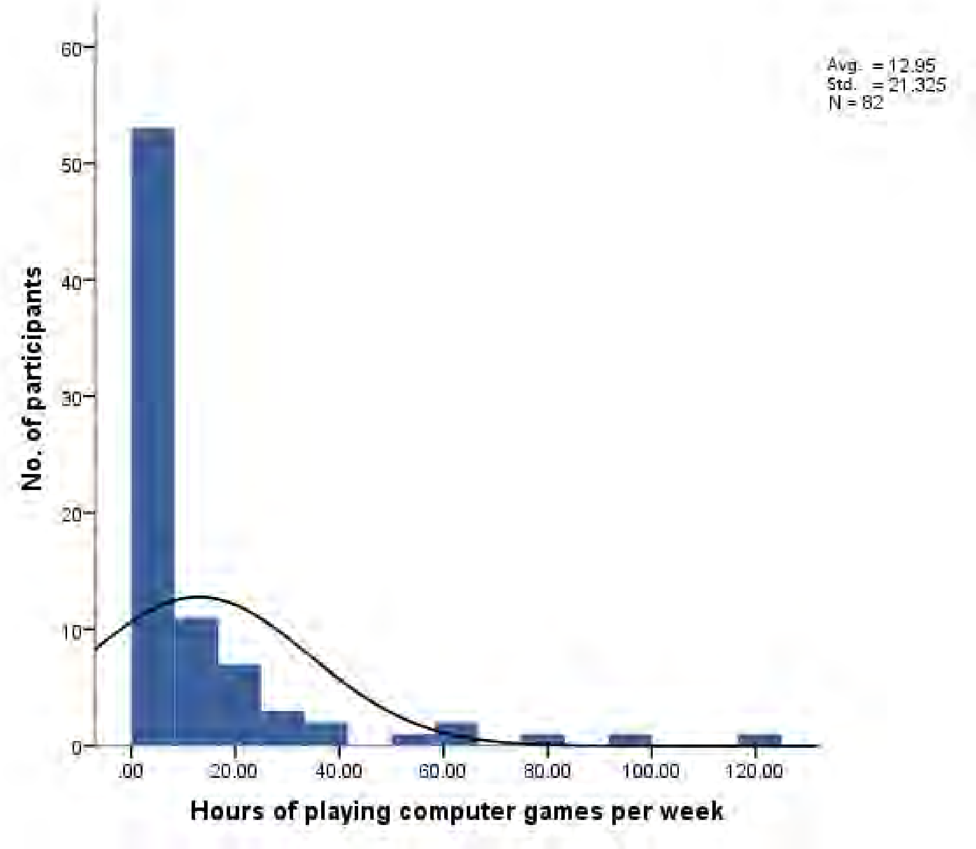


Figure 7-11 Frequency distribution histogram of playing computer games

7.3.1.4 Big Five personality dimensions and trust tendency

We calculated the score for each Big Five personality dimension by calculating the average of two items belonging to the dimension. We then calculated the descriptive values for these personality dimensions, as shown in Table 7.3. The means of these dimensions showed neutral for extraversion, emotional stability and openness to experience, and positive for agreeableness and conscientiousness. We compared the values of the Big Five personality dimensions of the control group, the experimental group and those who were excluded with one-way ANOVA. Results show there were significant differences in the values for emotional stability and openness to experience among these three groups. The experimental group had a higher average value for emotional stability than the control group. The experimental group also had a higher average value for openness to experience than the control group and those who were excluded. We calculated the average score for the three trust tendency questions as the final value for participants' willingness to trust. The results show participants' neutral tendency to trust, with a mean of 4 and standard deviation of 1.04. The maximum and minimum values for participants' trust tendency were 6 and 1. The control group had a higher minimum value for trust tendency, but the average value was the highest for the experimental group.

Table 7-3 Description of Big Five personality dimensions and trust tendency

		Extraversion	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experiences*	*Tend to Trust*
Total	Mean	4.15	5.32	5.42	4.73	4.83	4
	Std. Deviation	1.56	1.03	1.22	1.22	1.37	1.04
	Minimum	1	2.5	2	1.5	1.5	1
	Macimum	7	7	7	7	7	6
Ctrl Grp.	Mean	3.91	5.36	5.3	4.33	4.61	3.91
	Std. Deviation	1.41	1.05	1.12	1.07	1.24	0.88
	Minimum	1	2.5	2	2	2	2
	Macimum	7	7	7	6.5	7	6
Exp Grp.	Mean	4.4	5.22	5.41	5.04	5.27	4.11
	Std. Deviation	1.53	1.02	1.2	1.36	1.2	1.07
	Minimum	1	2.5	2.5	1.5	2.5	1
	Macimum	6.5	7	7	7	7	6
Excluded	Mean	4.13	5.37	5.5	4.78	4.68	3.97
	Std. Deviation	1.65	1.02	1.31	1.16	1.5	1.13
	Minimum	1	3.5	2	2.5	1.5	1
	Macimum	7	7	7	7	7	6

7.3.2 Reliability and validity

We examined the reliability of all the scales in the experiment with Cronbach's alpha (Cronbach, 1951), as shown in Table 7.4. The commonly accepted range for alpha is greater than 0.7 (Tavakol & Dennick, 2011). Thus, all the scales in the experiment showed high reliability. We also examined all the scales with the Kaiser–Meyer–Olkin (Bickmore & Cassell) measure of sampling adequacy. All values were above 0.6, indicating the scales to be suitable for factor analysis. The significance of Bartlett's test of sphericity was less than 0.05, which also indicates the high validity of all the scales. These two tests showed the scales in the experiment to be fit for the following analysis because of their high reliability and validity.

Table 7-4 Reliability and adequacy tests of scales

Group	*Scale*	*Aspect*	*Variable No.*	*Valid*	*%*	*Cronbach's Alpha*	*Bartlett's Test of Sphericity*			*Kaiser-Meyer-Olkin Measure of Sampling Adequacy.*
							Approx. Chi-Square	*df*	*Sig.*	
Control Group	TAM	Importance	12	38	100%	0.853	187.255	66	0	0.734
		PU	12	38	100%	0.872	247.987	66	0	0.630
		PEOU	12	38	100%	0.873	218.687	66	0	0.705
	Rapport	Attention	4	38	100%	0.705	30.844	6	0	0.626
		Positivity	6	38	100%	0.724	75.132	15	0	0.695
		Coordination	12	38	100%	0.919	483.731	66	0	0.850
Experimental Group	TAM	Importance	12	41	100%	0.888	265.513	66	0	0.730
		PU	12	41	100%	0.915	286.501	66	0	0.807
		PEOU	12	41	100%	0.898	332.104	66	0	0.790
	Rapport	Attention	4	41	100%	0.851	82.616	6	0	0.764
		Positivity	6	41	100%	0.775	93.848	15	0	0.658
		Coordination	12	41	100%	0.920	462.375	66	0	0.826
	DOI-Avatar	Relative advantage	6	41	100%	0.956	284.628	15	0	0.860
		Compatibility	6	41	100%	0.959	300.257	15	0	0.829
		Complexity	7	41	100%	0.881	164.992	21	0	0.766
		Trialability	4	41	100%	0.874	122.025	6	0	0.763
		Observability	3	41	100%	0.879	76.535	3	0	0.672
Both Groups	DOI-Portal	Relative advantage	6	79	100%	0.866	232.203	15	0	0.800
		Compatibility	6	79	100%	0.893	259.37	15	0	0.893
		Complexity	8	79	100%	0.909	374.617	28	0	0.886
		Trialability	4	79	100%	0.852	142.151	6	0	0.772
		Observability	3	79	100%	0.908	155.682	3	0	0.743
	Trustworthiness		8	79	100%	0.928	507.274	28	0	0.859

7.3.3 Normal distribution detection

Because the number of participants of the experiment was 160, we chose to use the Shapiro–Wilk normality test for detection of the distribution for all variables. Results of the Shapiro–Wilk normality test are shown in Table 7.5. We can see that all the variables of TAM and rapport have normal distributions. But some variables of DOI and trustworthiness do not seem to be normally distributed. None of the variables of the timing records and task accuracy seem to have normal distributions.

Table 7-5 Results of Shapiro–Wilk normality test

Group	*Scale*	*Aspect*	*Statistic*	*df*	*Sig.*
Control Group	TAM	Importance	0.971	38	0.412
		PU	0.953	38	0.112
		PEOU	0.965	38	0.268
	Rapport	Attention	0.966	38	0.301
		Positivity	0.978	38	0.651
		Coordination	0.965	38	0.272
Experimental Group	TAM	Importance	0.963	41	0.207
		PU	0.969	41	0.309
		PEOU	0.953	41	0.092
	Rapport	Attention	0.965	41	0.227
		Positivity	0.96	41	0.161
		Coordination	0.95	41	0.069
	DOI-Avatar	Relative advantage	0.926	41	**0.011***
		Compatibility	0.925	41	**0.010****
		Complexity	0.971	41	0.360
		Trialability	0.865	41	**0.000****
		Observability	0.873	41	**0.000****
Both Groups	DOI-Portal	Relative advantage	0.971	79	0.069
		Compatibility	0.963	79	0.023
		Complexity	0.982	79	0.346
		Trialability	0.94	79	**0.001****
		Observability	0.924	79	**0.000****
	Timing	First click	0.463	79	**0.000****
		Last click	0.790	79	**0.000****
		Submit	0.74	79	**0.000****
		Click count	0.767	79	**0.000****
	Task Accuracy	1	0.463	79	**0.000****
		2	0.586	79	**0.000****
		3	0.447	79	**0.000****
		4	0.634	79	**0.000****
		5	0.369	79	**0.000****
		Total	0.855	79	**0.000****
	Trustworthiness		0.941	79	**0.001****

7.3.4 Completion of tasks

In this section, results for the participants' completion of the 5 tasks given to them in the experiment are discussed. We compared the results of the control group and the experimental group for their first click, last click, time to submit, click count and accuracy of completing each task. The relationships of the demographic variables and the completion accuracy are also discussed.

As we can see in Table 7.6, for participants of both groups there were great differences in the time of their last click and page submission. Generally, the experimental group spent longer time on completing the tasks than the control group. This was because having conversation with the virtual agent needed more time than reading texts on a webpage. For the control group, their average time of first click was 36.94 seconds and for last click was about 7 minutes. On average, they spent 8.56 minutes on completing the 5 tasks. The quickest completion time was about 5 minutes and the longest time spent on the tasks was 21 minutes. For the experimental group, their average time of first click was 58.54 seconds and for last click was about 11 minutes. On average, they spent 12.98 minutes on completing all 5 tasks. The quickest completion time was about 5 minutes and the longest time spent on the tasks was 44.3 minutes. Because the time variables did not follow a normal distribution, we chose to carry out Mann–Whitney U tests to compare the results for the control group and the experimental group. Results are shown in Table 7.7. We can see that there are statistically significant differences between these two groups in last click, submit and click count.

Table 7-6 Descriptive analysis of completion of tasks

	Timing-First click		*Timing-Last click*		*Timing-Submit*		*Click Count*		*Accuracy-All*	
	Ctrl Grp.	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*
No.	38	41	38	41	38	41	38	41	38	41
Average	36.94	58.54	420.59	660.46	513.37	778.75	13.50	16.95	3.76	3.78
Median	25.13	13.46	342.70	465.41	425.18	634.72	9.50	14.00	4.00	4.00
Std.	42.10	126.82	295.26	568.56	262.65	516.35	9.57	12.02	1.22	1.24
Min.	0.92	1.97	15.01	32.54	301.96	305.26	5.00	6.00	0.00	1.00
Max.	170.10	712.06	1262.33	2655.61	1263.46	2657.96	44.00	64.00	5.00	5.00
	Accuracy-T1		*Accuracy-T2*		*Accuracy-T3*		*Accuracy-T4*		*Accuracy-T5*	
	Ctrl Grp.	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*	*Ctrl Grp.*	*Exp Grp.*
Right No.	33	32	26	28	30	36	20	23	34	36
Right %	86.8%	78.0%	68.4%	68.3%	78.9%	87.8%	52.6%	56.1%	89.5%	87.8%
Wrong No.	5	9	12	13	8	5	18	18	4	5
Wrong %	13.2%	22.0%	34.6%	31.7%	21.1%	12.2%	47.4%	43.9%	10.5%	12.2%

We also compared the accuracy of completing the tasks given to the participants during the experiment. We can see from Table 7.6 that on average the participants of these two groups completed 4 out of 5 tasks correctly. The highest accuracy rate lay in Task 5, while the lowest lay in Task 4. Both tasks were about finding data on the portal. For the control group, 14 (36.8%) of the participants completed all 5 tasks correctly, while one (2.6%) participant answered all questions incorrectly. For the experimental group, 16 (39%) of the participants completed all 5 tasks correctly, while no participants answered all 5 tasks incorrectly. Although the Mann–Whitney U test results, in Table 7.7, show no statistically significant differences in the accuracy rates of these two groups, we can see that the experimental group performed better on completing the tasks.

Table 7-7 Mann–Whitney U test of completion of tasks between the two groups

	Timing				*Accuracy -All*
	First click	*Last click*	*Submit*	*Click Count*	
Mann-Whitney U	726	565	475	561	770.5
Z	-0.52	-2.10	-2.98	-2.14	-0.09
Sig.	0.60	**0.036***	**0.003****	**0.032***	0.93
	Accuracy				
	Task 1	*Task 2*	*Task 3*	*Task 4*	*Task 5*
Mann-Whitney U	710.5	778	710	752	766
Z	-1.02	-0.01	-1.05	-0.31	-0.23
Sig.	0.31	0.99	0.29	0.76	0.82

7.3.5 Analysis of TAM

In this section, results for the TAM variables of the control group and the experimental group are discussed. We compared the values of importance, PEOU and PU for these two groups. Relationships between the demographic variables and TAM variables are also discussed. Pearson's correlation analysis has been carried out for discussing the relationships between importance, PU and PEOU.

7.3.5.1 Importance scale

For the importance of each function, the results in Table 7.8 for the descriptive analysis show that on average, participants of both groups showed positive attitudes to all the functions. Keyword search received the highest score for importance, followed by filtering of search results. Each of these two functions was related to the search function on the portal. For the control group, metadata received the lowest score for importance, followed by ranking. The largest deviation lay in downloading without registration, showing participants' different opinions of the importance of this function. The smallest deviation lay in filtering of search results, followed by keyword search, showing participants' agreement on the high importance of these two functions. For the experimental group, the virtual agent received the lowest score for importance, followed by ranking. The largest deviation lay in the virtual agent, showing participants' different opinions on the importance of the virtual agent. The smallest deviation lay in keyword search, followed by browsing. In comparison of the average scores of the importance of all 12 functions, we can see that the experimental group ranked the help functions as more important than the control group. The control group also scored the help functions as more important than the experimental group scoring the virtual agent. But the two independent sample T-tests showed there was no statistically significant difference between these two groups' average scores of the importance of each.

Table 7-8 Descriptive analysis of TAM importance scale

	Browse		*Various format*		*Download without registration*		*Filter*		*Help/Avatar*		*Keyword search*		*Metadata*	
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*
No.	38	41	38	41	38	41	38	41	38	41	38	41	38	41
Average	3.74	3.88	3.37	3.61	3.45	3.73	4.05	3.95	3.39	3.15	4.37	4.39	3.11	3.51
Median	4.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00	3.00	3.00	5.00	5.00	3.00	4.00
Std.	0.92	0.93	0.97	1.00	1.29	1.12	0.84	1.00	1.05	1.17	0.85	0.80	1.06	1.14
Min.	2.00	2.00	1.00	2.00	1.00	1.00	2.00	2.00	1.00	1.00	2.00	2.00	1.00	1.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	Machine-processable		*Ranking*		*Data request*		*Feedback*		*Visualize*		*Average*			
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*		
No.	38	41	38	41	38	41	38	41	38	41	38	41		
Average	3.58	3.85	3.34	3.39	3.45	3.49	3.58	3.90	3.74	3.83	3.60	3.72		
Median	4.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.67	3.75		
Std.	1.06	0.94	0.99	1.02	0.98	0.98	1.06	1.02	0.98	0.97	0.62	0.68		
Min.	2.00	2.00	2.00	1.00	2.00	1.00	1.00	2.00	1.00	2.00	2.00	2.00		
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.92		

7.3.5.2 Perceived ease of use scale

For the PEOU of each function, results, in Table 7.9, for the descriptive analysis show that on average, participants of both groups showed positive attitudes to all the functions, since the average scores were all above 3. For the control group, keyword search received the highest score for PEOU, followed by downloading without registration. Metadata received the lowest score for PEOU, followed by ranking. The smallest standard deviation lay in keyword search, showing participants' agreement on its PEOU. The largest standard deviation lay in metadata. For the experimental group, downloading without registration received the highest average score, followed by browsing. Metadata received the lowest average score, followed by filtering. The largest standard deviation lay in keyword search and the smallest in ranking. We noted that the scores for the PEOU of the control and experimental groups were quite different. On average, the control group scored PEOU higher than the experimental group, with a smaller standard deviation. But the two independent sample T-tests showed there was no statistically significant difference between these two groups' average scores of the PEOU of each function.

Table 7-9 Descriptive analysis of TAM perceived ease of use scale

	Browse		*Various format*		*Download without registration*		*Filter*		*Help/Avatar*		*Keyword search*		*Metadata*	
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*
No.	38	41	38	41	38	41	38	41	38	41	38	41	38	41
Average	4.08	3.93	3.66	3.68	4.21	4.22	4.08	3.59	3.68	3.66	4.34	3.71	3.34	3.44
Median	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	3.00	4.00
Std.	0.88	0.88	0.85	0.96	0.84	0.82	0.88	1.10	0.84	1.15	0.82	1.21	0.99	0.98
Min.	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	2.00	1.00	1.00	1.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00

	Machine-processable		*Ranking*		*Data request*		*Feedback*		*Visualize*		*Average*	
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*
No.	38	41	38	41	38	41	38	41	38	41	38	41
Average	3.89	3.83	3.66	3.63	3.74	3.80	3.74	3.61	3.89	3.73	3.86	3.74
Median	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.96	3.83
Std.	0.89	0.83	0.97	0.73	0.89	0.93	0.83	0.95	0.86	0.98	0.57	0.66
Min.	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	2.00	2.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.83

We tested the relationships between participants' socio-demographic background and their scores for PEOU. The two independent sample T-tests of the average PEOU by gender (Table 7.10) showed that there was a statistically significant difference in average PEOU between genders. On average, females ranked the functions as easier to use than males. For providing datasets of different formats and keyword search, on average females gave a statistically significantly higher PEOU score than males. One-way ANOVA showed statistically significant differences in average PEOU scores among participants of different occupations and monthly income. For participants of different occupations, they ranked the PEOU of providing datasets in various format and visualization differently. For participants with monthly income below 2000 RMB, on average they had lower PEOU scores on filtering and visualization compared to the participants of other monthly income. On average, they also had lower PEOU scores for metadata and open, machine-readable format compared to the participants of monthly income of 2001 to 3000 RMB and 5001 to 7000 RMB.

Table 7-10 Analysis of PEOU and socio-demographic variables

Aspect	*Gender*	*N*	*Avg.*	*Std.*	*T-test of Gender*				*Occupation*		*Income*	
					F	*P-*	*t*	*p*	*F*	*P-*	*F*	*P-*
Browse	Male Female	34 44	3.85 4.07	0.925 0.737	1.709	0.195	-1.145	0.256	0.798	0.660	0.907	0.465
Various format	Male Female	34 44	3.44 3.93	0.927 0.873	1.692	0.197	-2.395	**0.019***	1.973	**0.037***	2.399	0.058
Download without registration	Male Female	34 44	4.12 4.43	0.844 0.661	0.089	0.766	-1.844	0.069	0.241	0.997	0.953	0.438
Filter	Male Female	34 44	3.53 3.91	1.107 0.936	3.439	0.068	-1.64	0.105	1.691	0.084	4.498	**0.003****
Help/Avatar	Male Female	34 44	3.47 3.80	1.051 1.002	0.920	0.340	-1.39	0.169	1.100	0.375	1.597	0.184
Keyword search	Male Female	34 44	3.53 4.16	1.237 1.010	2.983	0.088	-2.475	**0.016***	1.632	0.099	2.173	0.080
Metadata	Male Female	34 44	3.41 3.48	0.988 0.792	1.884	0.174	-0.325	0.746	0.346	0.981	3.085	**0.021***
Machine-processable	Male Female	34 44	3.59 3.95	1.019 0.776	7.256	**0.009****	-1.742	0.087	1.506	0.139	2.609	**0.042***
Ranking	Male Female	34 44	3.62 3.70	0.853 0.795	0.323	0.572	-0.464	0.644	1.574	0.116	1.684	0.163
Data request	Male Female	34 44	3.56 3.93	1.021 0.759	8.593	**0.004****	-1.783	0.080	0.688	0.768	0.723	0.579
Feedback	Male Female	34 44	3.56 3.64	0.894 0.917	0.022	0.883	-0.374	0.709	1.281	0.248	1.273	0.288
Visualize	Male Female	34 44	3.65 3.82	1.070 0.947	1.630	0.206	-0.748	0.457	2.772	**0.003****	2.808	**0.031***
Average	Male Female	34 44	3.61 3.90	0.699 0.528	4.046	**0.048***	-2.036	**0.046***	1.624	0.101	3.726	**0.008****

We tested the relationships between participants' past experiences of OGD and OGD portals and their scores for PEOU of each function, including whether they had ever used OGD or OGD portals before and their usage frequency of OGD and OGD portals. Results are shown in Table 7.11. The two independent sample T-tests showed that on average, compared with those who had not used OGD before, participants who had used OGD ranked the PEOU dimension of downloading without registration significantly higher. One-way ANOVA showed that there were statistically significant differences in average scores of these

PEOU dimensions: browsing, downloading without registration and data request, among participants of different use frequency of OGD. Compared with the participants who had never used OGD and those who had used OGD more than once but very rarely, those who had used it daily gave browsing, on average, a higher PEOU score. Compared with those who had never used OGD, participants who had used OGD daily, weekly or monthly gave downloading without registration, on average, a higher PEOU score. Compared with participants who had used OGD more than once but very rarely, those who had used it daily, weekly or monthly gave requesting new datasets, on average, a higher PEOU score.

Table 7-11 Analysis of PEOU and OGD & OGD portal experience variables

Aspect	*Use OGD*	*N*	*Avg.*	*Std.*	*T-test of OGD Experience*				*OGD Frequency*		*Portal Frequency*	
					F	*P-*	*t*	*p*	*F*	*P-*	*F*	*P-*
Browse	Yes No	35 44	4.06 3.88	0.838 0.823	0.000	0.994	0.918	0.361	3.067	**0.022***	1.233	0.303
Various format	Yes No	35 44	3.69 3.70	1.105 0.795	5.800	0.018	-0.085	0.933	0.418	0.795	1.614	0.167
Download without registration	Yes No	35 44	4.51 4.11	0.702 0.754	0.035	0.853	2.419	**0.018***	2.506	**0.049***	0.652	0.661
Filter	Yes No	35 44	3.77 3.70	1.060 1.002	0.000	0.983	0.287	0.775	0.621	0.649	0.384	0.859
Help/Avatar	Yes No	35 44	3.74 3.59	0.980 1.064	0.610	0.437	0.653	0.516	0.730	0.574	0.786	0.563
Keyword search	Yes No	35 44	3.94 3.84	1.305 1.010	1.940	0.168	0.392	0.696	2.015	0.101	0.419	0.834
Metadata	Yes No	35 44	3.31 3.50	0.900 0.928	0.078	0.780	-0.896	0.373	0.311	0.870	2.378	**0.047***
Machine-processable	Yes No	35 44	3.69 3.89	0.963 0.841	2.679	0.106	-0.987	0.327	0.265	0.900	1.050	0.395
Ranking	Yes No	35 44	3.71 3.64	0.825 0.810	0.014	0.905	0.421	0.675	0.416	0.796	0.922	0.472
Data request	Yes No	35 44	3.80 3.70	0.964 0.878	0.079	0.779	0.460	0.647	3.117	**0.020***	1.189	0.323
Feedback	Yes No	35 44	3.57 3.57	0.778 1.065	4.412	0.039	0.016	0.988	0.874	0.484	0.629	0.678
Visualize	Yes No	35 44	3.69 3.77	1.132 0.886	2.305	0.133	-0.384	0.702	0.381	0.822	1.210	0.313
Average	Yes No	35 44	3.79 3.74	0.605 0.647	0.095	0.759	0.324	0.747	1.205	0.316	1.134	0.350

7.3.5.3 Perceived usefulness scale

For the PU of each function, results, in Table 7.12, for the descriptive analysis show that on average, participants of both groups showed positive attitudes to all the functions, since all average scores were above 3. For the control group, keyword search received the highest score for PU, followed by downloading without registration. Metadata received the lowest score for PU, followed by different formats to choose and ranking. The smallest standard deviation lay in keyword search, showing participants' agreement on its PU. The largest standard deviation lay in metadata. For the experimental group, downloading without registration received the highest score for PU, followed by keyword search. Virtual agent received the lowest score for PU, followed by feedback and visualization. The smallest standard deviation lay in ranking, while the largest standard deviation lay in virtual agent. Generally, the control group gave a higher PU score than the experimental group, with a smaller standard deviation. Virtual agent received the highest standard deviation and the lowest average PU, indicating participants' different opinion of this function. The two independent sample T-tests showed that there were no statistically significant differences between these two groups in the average scores of the PU of each function. We tested the relationships between participants' socio-demographic background and their scores for PU. No statistically significant relationships have been found between participants' demographic background and their scores for PU.

We also analyzed the relationships between participants' past experience of OGD and OGD portals and their scores for PU of each function, including whether they had ever used OGD or OGD portals before and their usage frequency of OGD and OGD portals. The two independent sample T-tests showed there was no statistically significant difference in the average PU scores between participants who had used OGD or OGD portals before and those who had not. But the one-way ANOVA showed there were statistically significant differences in the average PU scores for browsing ($F=2.834$, $P=0.03$), help ($F=2.57$, $P=0.045$) and metadata ($F=3.204$, $P=0.018$) among participants of different usage frequency of OGD. Participants who had used OGD daily and weekly gave on average a higher score for the PU of browsing than those who had not used OGD before and who had used OGD more than once but very rarely. For the help functions, participants who had used OGD daily gave on average a higher PU score than those who had used OGD more than once but very rarely and those who had used OGD monthly.

For metadata, participants who had used OGD more than once but very rarely gave on average a lower PU score than those who had never used OGD and those who had used OGD daily and weekly.

Table 7-12 Descriptive analysis of TAM perceived usefulness scale

	Browse		*Various format*		*Download without registration*		*Filter*		*Help/Avatar*		*Keyword search*		*Metadata*	
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*
No.	38	41	38	41	38	41	38	41	38	41	38	41	38	41
Average	4.08	4.07	3.66	3.83	4.21	4.32	4.08	3.83	3.68	3.41	4.34	4.12	3.34	3.68
Median	4.00	4.00	4.00	4.00	4.00	5.00	4.00	4.00	4.00	4.00	5.00	4.00	3.00	4.00
Std.	0.88	0.82	0.85	0.95	0.84	0.99	0.88	1.02	0.84	1.14	0.82	0.98	0.99	0.99
Min.	2.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	2.00	1.00	2.00	1.00	1.00	1.00
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	Machine-processable		*Ranking*		*Data request*		*Feedback*		*Visualize*		*Average*			
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*		
No.	38	41	38	41	38	41	38	41	38	41	38	41		
Average	3.89	4.05	3.66	3.66	3.74	3.90	3.74	3.61	3.89	3.61	3.86	3.84		
Median	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.96	4.00		
Std.	0.89	0.81	0.97	0.79	0.89	0.89	0.83	1.00	0.86	0.95	0.57	0.68		
Min.	2.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	2.00	1.00	2.00	2.00		
Max.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00		

7.3.5.4 Correlation analysis of importance, PEOU and PU

We carried out Pearson's correlation analysis of the importance, PEOU and PU of different functions. The correlations between importance and PEOU are shown in Table 7.13 following. We can see that there were statistically significant weak positive correlations between the importance scores and the PEOU scores for functions including providing different data formats, filtering, help/virtual agent, metadata and machine-readable data, as well as for the average scores.

The correlations between importance and PU are shown in Table 7.14 following. We can see that except for visualization, there were statistically significant positive correlations between the scores for importance and PU for each function. The correlations were weak for browsing, downloading without registration, filtering, keyword search and data request, and were moderate for various formats, help/virtual agent, metadata, machine-readable data, ranking, feedback and the average scores.

The correlations between PEOU and PU are shown in Table 7.15 following. We can see that there were statistically significant positive correlations between the scores for PEOU and PU for each function and the average scores. The correlations were weak for browsing, filtering, machine-readable data and ranking, and moderate for various formats, downloading without registration, help/virtual agent, keyword search, metadata, data request, feedback and visualization, and strong for the average scores.

7.3.6 DOI for the OGD portal

In this section, results of the DOI variables for participants' acceptance of the OGD portals of the control group and the experimental group are presented. We first calculated the average scores for each participant for these five aspects. We compared the values for relative advantage, compatibility, complexity, trialability and observability of these two groups. Relationships between the demographic variables and DOI variables are also discussed.

The descriptive analysis of the five variables of DOI is displayed in Table 7.16 following. On average, participants of both groups show positive attitudes to all five aspects of DOI, since all average scores were above 4 (of the 6-point Likert scale). For the control group, trialability received the highest average score among all five aspects of DOI, while complexity received the lowest score. Trialability also showed the smallest standard deviation, indicating participants' agreement on this aspect. The largest standard deviation lay in observability, which had the lowest minimum value. For the experimental group, the largest average score lay in compatibility, with the smallest standard deviation, and the smallest average score lay in complexity, the same as for the control group. The largest standard deviation lay in observability, which also had the smallest minimum value. Mann–Whitney U tests showed there were no statistically significant differences in the scores of these five aspects between the control group and the experimental group.

Table 7-13 Pearson's correlation analysis of importance and PEOU

Importance	*Correlation*	*PEOU*												
		Browse	Various Format	Download without registration	Filter	Help/Avatar	Keyowrd search	Metadata	Machine-processable	Ranking	Data request	Feedback	Visualize	Average
Browse	Pearson's r	0.007	0.102	0.204	-0.021	0.085	0.013	0.048	0.071	0.007	0.226*	0.070	0.007	0.095
	p-	0.954	0.371	0.072	0.855	0.458	0.909	0.678	0.533	0.951	0.046	0.542	0.953	0.404
Various format	Pearson's r	0.323**	0.330**	0.237*	0.191	0.126	0.112	0.326**	0.329**	0.233*	0.295**	0.187	0.223*	0.351**
	p-	0.004	0.003	0.036	0.091	0.268	0.327	0.003	0.003	0.039	0.008	0.098	0.048	0.002
Download without registration	Pearson's r	0.084	0.033	0.212	0.007	-0.147	-0.047	-0.037	0.078	-0.083	0.249*	0.093	-0.027	0.043
	p-	0.464	0.774	0.061	0.952	0.196	0.680	0.746	0.494	0.468	0.027	0.416	0.812	0.707
Filter	Pearson's r	0.271*	0.282*	0.444**	0.232*	0.082	0.158	0.351**	0.233*	0.000	0.428**	0.207	0.126	0.342**
	p-	0.016	0.012	0.000	0.040	0.474	0.163	0.002	0.039	1.000	0.000	0.067	0.269	0.002
Help/Avatar	Pearson's r	0.305**	0.277*	0.180	0.247*	0.228*	0.162	0.297**	0.292**	0.099	0.159	0.095	0.148	0.308**
	p-	0.007	0.014	0.112	0.029	0.044	0.153	0.008	0.009	0.387	0.161	0.405	0.192	0.006
Keyword search	Pearson's r	0.288*	0.201	0.503**	0.137	0.034	0.156	0.264*	0.228*	0.113	0.369**	0.148	0.047	0.295**
	p-	0.011	0.075	0.000	0.228	0.764	0.170	0.019	0.044	0.322	0.001	0.195	0.683	0.008
Metadata	Pearson's r	0.279*	0.326**	0.255*	0.221	0.264*	0.119	0.321**	0.449**	0.088	0.168	0.241*	0.180	0.359**
	p-	0.013	0.003	0.023	0.050	0.019	0.296	0.004	0.000	0.44	0.139	0.033	0.111	0.001
Machine-processable	Pearson's r	0.147	0.196	0.211	0.027	0.069	-0.028	0.213	0.280*	0.154	0.245*	0.062	0.002	0.181
	p-	0.200	0.084	0.062	0.813	0.548	0.806	0.059	0.013	0.175	0.029	0.589	0.986	0.110
Ranking	Pearson's r	0.175	0.188	0.247*	0.134	0.136	0.026	0.138	0.055	0.135	0.299**	0.373**	0.125	0.243*
	p-	0.125	0.097	0.028	0.239	0.231	0.822	0.224	0.629	0.237	0.007	0.001	0.274	0.031
Data request	Pearson's r	0.154	0.270*	0.197	-0.015	0.227*	-0.055	0.224*	0.154	0.182	0.164	0.251*	0.064	0.213
	p-	0.179	0.016	0.083	0.895	0.044	0.629	0.047	0.174	0.109	0.148	0.026	0.575	0.059
Feedback	Pearson's r	0.287*	0.300**	0.388**	0.201	0.086	0.158	0.300**	0.287*	-0.084	0.295**	0.096	0.008	0.282*
	p-	0.011	0.007	0.000	0.076	0.451	0.164	0.007	0.010	0.459	0.008	0.398	0.941	0.012
Visualize	Pearson's r	0.122	0.096	0.157	0.058	-0.036	0.012	0.146	0.126	0.039	0.184	0.136	-0.020	0.116
	p-	0.288	0.399	0.167	0.612	0.751	0.914	0.199	0.268	0.732	0.105	0.233	0.860	0.307
Average	Pearson's r	0.315**	0.326**	0.425**	0.186	0.163	0.111	0.304**	0.322**	0.127	0.388	0.260*	0.105	0.364**
	p-	0.005	0.003	0.000	0.100	0.152	0.329	0.006	0.004	0.266	0.000	0.021	0.356	0.001

Table 7-14 Pearson's correlation analysis of importance and PU

Importance	*Correlation*	*PU*												
		Browse	Various Format	Download without registration	Filter	Help/Avatar	Keyowrd search	Metadata	Machine-processable	Ranking	Data request	Feedback	Visualize	Average
Browse	Pearson's r	0.387**	0.229*	0.146	0.163	0.225*	0.184	0.194	0.357**	0.147	0.378**	0.144	0.041	0.313**
	p-	0.000	0.042	0.200	0.152	0.046	0.105	0.087	0.001	0.196	0.001	0.206	0.717	0.005
Various format	Pearson's r	0.316**	0.493**	0.181	0.265*	0.202	0.279*	0.258*	0.363**	0.224*	0.348**	0.411**	0.192	0.427**
	p-	0.005	0.000	0.111	0.018	0.075	0.013	0.022	0.001	0.047	0.002	0.000	0.090	0.000
Download without registration	Pearson's r	0.119	0.113	0.254*	0.128	0.046	0.170	0.131	0.234*	0.154	0.253*	-0.049	0.049	0.193
	p-	0.296	0.320	0.024	0.261	0.688	0.135	0.251	0.038	0.177	0.024	0.671	0.668	0.089
Filter	Pearson's r	0.281*	0.155	0.122	0.363**	0.304**	0.185	0.182	0.362**	0.159	0.377**	0.259*	0.260*	0.366**
	p-	0.012	0.172	0.285	0.001	0.007	0.103	0.109	0.001	0.161	0.001	0.021	0.021	0.001
Help/Avatar	Pearson's r	0.132	0.316**	0.167	0.373**	0.403**	0.254**	0.344**	0.307**	0.345**	0.250*	0.416**	0.311**	0.444**
	p-	0.246	0.005	0.142	0.001	0.000	0.024	0.002	0.006	0.002	0.026	0.000	0.005	0.000
Keyword search	Pearson's r	0.328**	0.219	0.290**	0.350**	0.242*	0.348**	0.194	0.346**	0.219	0.357**	0.305**	0.113	0.401**
	p-	0.003	0.053	0.009	0.002	0.032	0.002	0.086	0.002	0.053	0.001	0.006	0.322	0.000
Metadata	Pearson's r	0.219	0.426**	0.130	0.327**	0.323**	0.296**	0.438**	0.416**	0.230*	0.368**	0.354**	0.231*	0.459**
	p-	0.052	0.000	0.254	0.003	0.004	0.008	0.000	0.000	0.041	0.001	0.001	0.041	0.000
Machine-processable	Pearson's r	0.314**	0.349**	0.124	0.159	0.178	0.142	0.237*	0.416**	0.154	0.420**	0.179	-0.008	0.321**
	p-	0.005	0.002	0.276	0.162	0.118	0.212	0.036	0.000	0.176	0.000	0.115	0.944	0.004
Ranking	Pearson's r	0.103	0.062	0.269*	0.353**	0.205	0.246*	0.192	0.117	0.481**	0.204	0.329**	0.159	0.333**
	p-	0.366	0.588	0.016	0.001	0.069	0.029	0.091	0.305	0.000	0.072	0.003	0.162	0.003
Data request	Pearson's r	0.112	0.138	0.118	0.108	0.194	0.096	0.235*	0.170	0.266*	0.305**	0.305**	0.135	0.266*
	p-	0.324	0.227	0.302	0.342	0.087	0.401	0.037	0.133	0.018	0.006	0.006	0.234	0.018
Feedback	Pearson's r	0.168	0.355**	0.353**	0.384**	0.169	0.374**	0.312**	0.312**	0.255*	0.172	0.408**	0.080	0.408**
	p-	0.140	0.001	0.001	0.000	0.137	0.001	0.005	0.005	0.023	0.129	0.000	0.484	0.000
Visualize	Pearson's r	0.083	0.099	0.109	0.264*	0.173	0.130	0.157	0.243*	0.245*	0.253*	0.222*	0.126	0.256*
	p-	0.468	0.387	0.341	0.019	0.126	0.255	0.168	0.031	0.030	0.025	0.049	0.268	0.023
Average	Pearson's r	0.327**	0.371**	0.330**	0.395**	0.343**	0.350**	0.358**	0.443**	0.368**	0.475**	0.422**	0.203	0.533**
	p-	0.003	0.001	0.003	0.000	0.002	0.002	0.001	0.000	0.001	0.000	0.000	0.073	0.000

Table 7-15 Pearson's correlation analysis of PEOU and PU

PEOU	*Correlation*	*PU*												
		Browse	Various Format	Download without registration	Filter	Help/Avatar	Keyowrd search	Metadata	Machine-processable	Ranking	Data request	Feedback	Visualize	Average
Browse	Pearson's r	0.375**	0.299**	0.354**	0.438**	0.318**	0.374**	0.524**	0.239*	0.183	0.464**	0.349**	0.487**	0.537**
	p-	0.001	0.008	0.001	0.000	0.005	0.001	0.000	0.035	0.109	0.000	0.002	0.000	0.000
Various format	Pearson's r	0.240*	0.440**	0.185	0.239*	0.217	0.339**	0.376**	0.168	0.278*	0.304**	0.330**	0.343**	0.422**
	p-	0.033	0.000	0.103	0.034	0.054	0.002	0.001	0.140	0.013	0.007	0.003	0.002	0.000
Download without registration	Pearson's r	0.408**	0.243*	0.518**	0.446**	0.328**	0.541**	0.376**	0.333**	0.094	0.480**	0.141	0.351**	0.520**
	p-	0.000	0.031	0.000	0.000	0.003	0.000	0.001	0.003	0.408	0.000	0.217	0.002	0.000
Filter	Pearson's r	0.425**	0.359**	0.172	0.392**	0.390**	0.496**	0.351**	0.200	0.241*	0.357**	0.316**	0.532**	0.517**
	p-	0.000	0.001	0.129	0.000	0.000	0.000	0.002	0.078	0.032	0.001	0.005	0.000	0.000
Help/Avatar	Pearson's r	0.283*	0.309**	0.125	0.269*	0.566**	0.334**	0.402**	-0.010	0.269*	0.356**	0.302**	0.428**	0.450**
	p-	0.012	0.006	0.271	0.016	0.000	0.003	0.000	0.930	0.017	0.001	0.007	0.000	0.000
Keyword search	Pearson's r	0.487**	0.334**	0.323**	0.392**	0.421**	0.521**	0.322**	0.222*	0.153	0.346**	0.233*	0.513**	0.521**
	p-	0.000	0.003	0.004	0.000	0.000	0.000	0.004	0.049	0.179	0.002	0.039	0.000	0.000
Metadata	Pearson's r	0.191	0.255*	-0.073	0.317**	0.125	0.085	0.574**	0.246*	0.117	0.329**	0.396**	0.190	0.339**
	p-	0.092	0.023	0.522	0.004	0.271	0.457	0.000	0.029	0.306	0.003	0.000	0.094	0.002
Machine-processable	Pearson's r	0.309**	0.508**	0.207	0.286*	0.251*	0.389**	0.520**	0.381**	0.188	0.325**	0.199	0.328**	0.475**
	p-	0.006	0.000	0.067	0.011	0.026	0.000	0.000	0.001	0.097	0.004	0.079	0.003	0.000
Ranking	Pearson's r	0.243*	0.341**	0.050	0.011	0.221	0.156	0.245*	0.044	0.381**	0.167	0.180	0.301**	0.285*
	p-	0.031	0.002	0.660	0.922	0.050	0.171	0.300	0.702	0.001	0.141	0.112	0.007	0.011
Data request	Pearson's r	0.325**	0.187	0.358**	0.351**	0.416**	0.350**	0.428**	0.158	0.131	0.545**	0.221	0.245*	0.457**
	p-	0.003	0.099	0.001	0.002	0.000	0.002	0.000	0.166	0.250	0.000	0.050	0.029	0.000
Feedback	Pearson's r	0.138	0.142	0.134	0.316**	0.276*	0.206	0.377**	0.002	0.239*	0.428**	0.442**	0.304**	0.371**
	p-	0.224	0.211	0.238	0.005	0.014	0.068	0.001	0.984	0.034	0.000	0.000	0.006	0.001
Visualize	Pearson's r	0.314**	0.354**	0.149	0.241*	0.439**	0.281*	0.334**	0.068	0.380**	0.294**	0.268*	0.560**	0.452**
	p-	0.005	0.001	0.191	0.033	0.000	0.012	0.003	0.552	0.001	0.009	0.017	0.000	0.000
Average	Pearson's r	0.470**	0.471**	0.305**	0.462**	0.503**	0.511**	0.597**	0.251*	0.327**	0.543**	0.419**	0.580**	0.667**
	p-	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.026	0.003	0.000	0.000	0.000	0.000

Table 7-16 Descriptive analysis of DOI for OGD portal

	Relative Advantage		Compatibility		Complexity		Trialability		Observability	
	Ctrl	Exp	Ctrl	Exp	Ctrl	Exp	Ctrl	Exp	Ctrl	Exp
No.	38	41	38	41	38	41	38	41	38	41
Average	4.71	4.70	4.81	4.84	4.48	4.35	4.94	4.80	4.68	4.57
Median	5.00	4.67	5.00	4.83	4.50	4.38	5.00	5.00	5.00	5.00
Std.	0.79	0.72	0.79	0.70	0.77	0.78	0.76	0.79	0.86	0.89
Min.	3.17	2.67	3.00	3.50	3.13	2.75	3.00	2.75	2.67	1.67
Max.	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00

We tested the relationships between the socio-demographic backgrounds and participants' experience with OGD and OGD portals and the scores for DOI for the portal. Results are shown in Table 7.17. We can see from the results of Kruskal–Wallis one-way analysis that participants' occupation had a statistically significant relationship with the score for relative advantage. Logistics, financial staffs, civilian posts, teachers and consultants gave higher scores for relative advantage than students, production workers, human resource workers, technicians and specialists. Kruskal–Wallis one-way analysis also showed that there were statistically significant differences in the score for complexity among participants of different monthly income. Participants with monthly income of 2001 to 7000 RMB gave higher scores for complexity than participants with monthly income below 2000 RMB or above 7000 RMB. Mann–Whitney U tests showed a statistically significant difference in the scores for compatibility between participants who had used OGD and those who had not. Participants who had used OGD before gave higher scores for compatibility compared with those who had never used OGD.

Table 7-17 Analysis of demographic background and DOI

Aspect		Relative Advantage	Compatibility	Complexity	Trialability	Observability
Gender	Mann-Whitney U	611.000	669.500	723.000	735.000	738.000
	Z	-1.388	-0.795	-0.253	-0.133	-0.103
	P-	0.165	0.427	0.800	0.894	0.918
Age	Chi-squre	6.007	2.882	3.320	0.235	3.311
	df	4.000	4.000	4.000	4.000	4.000
	P-	0.199	0.578	0.506	0.994	0.507
Education	Chi-squre	0.169	1.392	0.989	2.970	0.938
	df	3.000	3.000	3.000	3.000	3.000
	P-	0.982	0.707	0.804	0.396	0.816
Occupation	Chi-squre	27.062	16.198	19.164	12.473	13.573
	df	13.000	13.000	13.000	13.000	13.000
	P-	**0.012***	0.239	0.118	0.489	0.405
Monthly Income	Chi-squre	4.449	3.289	10.197	1.100	4.218
	df	4.000	4.000	4.000	4.000	4.000
	P-	0.349	0.511	**0.037***	0.894	0.377
Use OGD	Mann-Whitney U	691.500	561.500	639.500	593.000	701.000
	Z	-0.779	-2.067	-1.292	-1.778	-0.700
	P-	0.436	**0.039***	0.196	0.075	0.484
Use Portal	Mann-Whitney U	569.000	604.500	560.000	610.000	594.500
	Z	-1.123	-0.747	-1.216	-0.698	-0.873
	P-	0.261	0.455	0.224	0.485	0.383

7.3.7 DOI for the virtual agent

In this section, we discuss the results for the scales of DOI for the virtual agent, including relative advantage, compatibility, complexity, trialability and observability. We first calculated the average scores for each participant for these five aspects. The descriptive analysis results for these five aspects are shown in Table 7.18. We can see from the results that on average, participants showed positive attitudes towards the five aspects of DOI for the virtual agent, with all average scores above 4. The largest average score lay in trialability and the smallest in compatibility. Compatibility also had the largest standard deviation. The smallest standard deviation lay in complexity, which also had the highest minimum score among these five aspects.

Table 7-18 Descriptive analysis of DOI for the virtual agent

	Relative Advantage	*Compatibility*	*Complexity*	*Trialability*	*Observability*
No.	41	41	41	41	41
Average	4.24	4.18	4.48	4.60	4.51
Median	4.50	4.67	4.43	5.00	4.67
Std. deviation	1.19	1.21	0.79	0.91	1.02
Min.	1.17	1.00	2.29	1.00	1.00
Max.	6.00	6.00	6.00	6.00	6.00

We tested the relationships of participants' socio-demographic background and their knowledge of OGD and OGD portals with the five aspects of DOI for the virtual agent. Results are shown in Table 7.19. Kruskal–Wallis one-way analysis showed that there were statistically significant relationships between participants' education background with the scores for compatibility and complexity, and between their occupation with the scores for relative advantage and compatibility. Participants with bachelor's degrees gave higher compatibility and complexity scores than participants with senior high school degrees and master's degrees. Participants with doctoral degrees gave the lowest compatibility and complexity scores. Consultants, civilian posts, logistic workers and managers gave higher relative advantage and compatibility scores than students, salespeople, human resources workers, technicians and teachers. Mann–Whitney U tests showed a statistically significant difference in the score for trialability between participants who had used OGD and those who had not. Compared with those who had never used OGD before, participants who had used OGD gave a higher score for the virtual agent's trialability.

Table 7-19 Analysis of demographic background and DOI

Aspect		Relative Advantage	Compatibility	Complexity	Trialability	Observability
Gender	Mann-Whitney U	186.500	194.500	174.500	186.000	187.500
	Z	-0.353	-0.136	-0.680	-0.375	-0.331
	P-	0.724	0.892	0.496	0.707	0.741
Age	Chi-squre	4.904	5.112	5.374	2.529	2.929
	df	4.000	4.000	4.000	4.000	4.000
	P-	0.297	0.276	0.251	0.639	0.570
Education	Chi-squre	6.431	8.342	10.486	7.076	4.604
	df	3.000	3.000	3.000	3.000	3.000
	P-	0.092	**0.039***	**0.015***	0.070	0.203
Occupation	Chi-squre	24.519	21.326	20.607	12.826	14.931
	df	12.000	12.000	12.000	12.000	12.000
	P-	**0.017***	**0.046***	0.056	0.382	0.245
Monthly Income	Chi-squre	7.026	6.490	8.815	5.308	4.056
	df	4.000	4.000	4.000	4.000	4.000
	P-	0.135	0.165	0.066	0.257	0.398
Use OGD	Mann-Whitney U	172.500	155.500	157.000	128.500	167.000
	Z	-0.836	-1.289	-1.249	-2.049	-0.997
	P-	0.403	0.197	0.212	**0.040***	0.319
Frequency	Chi-squre	5.913	6.469	5.095	5.202	4.112
	df	3.000	3.000	3.000	3.000	3.000
	P-	0.116	0.091	0.165	0.158	0.250
Use Portal	Mann-Whitney U	132.000	132.500	108.500	116.500	143.500
	Z	-0.379	-0.364	-1.123	-0.888	-0.016
	P-	0.705	0.716	0.262	0.374	0.987
Frequency	Chi-squre	2.230	1.778	0.431	0.073	0.088
	df	2.000	2.000	2.000	2.000	2.000
	P-	0.328	0.411	0.806	0.964	0.957

7.3.8 Trustworthiness

In this section, we discuss the results for the scale of trustworthiness for the control and experimental groups. We first calculated the average scores for all items in the scale for each participant, then carried out descriptive analysis of the average scores of trustworthiness for both groups. For the control group, the average score for trustworthiness was 4.81 and the median value was 5, both showing participants' positive attitudes towards trust in the OGD portal. The standard deviation was 0.79, with the minimum value 3 and the maximum value 6. For the experimental group, the average score for trustworthiness was 4.77 and the median value was 5, both showing their positive attitudes. The standard deviation was 0.79, with the minimum value 2.75 and the maximum value 6. Mann–Whitney U tests showed there was no statistically significant difference in the values for trustworthiness between groups (U=765.5, Z=−0.134, P=−0.893).

We tested the relationship between participants' socio-demographic background and their perceptions of the trustworthiness of the OGD portal. Mann–Whitney U tests and Kruskal–Wallis one-way analysis showed that there were no statistically significant relationships between participants' gender, age, education background, occupation or income with their scores of trustworthiness for the portal. We also tested the relationship between participants' past experience with OGD and their perceptions of the trustworthiness of the OGD portal. Results are shown in Table 7.20. Mann–Whitney U tests showed that there was a statistically significant difference in the score for trustworthiness of the OGD portal between participants who had used OGD before and those who had not. Compared with those who had never used OGD before, participants who had used OGD gave a much higher trustworthiness score for the OGD portal.

Table 7-20 Analysis of participants' past experience with OGD and trustworthiness

Aspect		*Trustworthiness*
Use OGD	Mann-Whitney U	532.500
	Z	-2.372
	P-	**0.018***
Frequency	Chi-squre	3.367
	df	3.000
	P-	0.338
Use Portal	Mann-Whitney U	595.000
	Z	-0.853
	P-	0.394
Frequency	Chi-squre	2.482
	df	4.000
	P-	0.648

7.3.9 Rapport with help pages and virtual agent

In this section, we discuss the results for the scale of rapport for the help pages for the control group and the virtual agent for the experimental group. We first calculated the average scores for attention, positivity and coordination. The average scores of all the rapport questions were also calculated. The descriptive results are shown in Table 7.21. For the control group, generally participants showed positive rapport with the help pages. The highest average score lay in coordination and the lowest in positivity, which also received the lowest minimum score. The largest standard deviation lay in attention and the smallest in positivity, showing participants' agreement on their attitudes towards the positivity of the help pages. For the experimental group, generally participants also showed positive rapport with the virtual agent. The highest average score lay in coordination and the lowest in positivity. The largest standard deviation lay in attention, which also received the lowest minimum score and the highest maximum score. The smallest standard deviation lay in positivity, which received the highest minimum score and the lowest maximum score. Although the scores for attention, positivity and coordination for rapport of the control group with the help pages was higher than the scores for the experimental group with the virtual agent, the two independent sample T-tests showed that the differences between these two groups were not statistically significant.

Table 7-21 Descriptive analysis of rapport

	Attention		*Positivity*		*Coordination*		*Average*	
	Ctrl	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*	*Ctrl*	*Exp*
No.	38	41	38	41	38	41	38	41
Average	4.21	3.94	4.12	3.93	4.25	4.07	4.21	4.00
Median	4.25	4.00	4.08	3.83	4.33	4.17	4.30	3.95
Std.	0.86	1.12	0.76	0.81	0.84	0.92	0.77	0.88
Min.	2.50	1.00	2.33	2.50	2.50	1.42	2.45	1.68
Max.	6.00	6.00	6.00	5.50	6.00	5.67	5.77	5.64

We tested the relationship between participants' socio-demographic background and rapport with the help functions of the OGD portal. Results are shown in Table 7.22. The two independent sample T-tests showed that there was a statistically significant difference in the average scores for attention between participants of different gender. Compared to males, females gave higher attention scores for the help functions, indicating that the help functions did not distract the attention of the participants when they were trying to complete the tasks in the experiment. One-way ANOVA showed that education had a statistically significant effect on participants' scores for positivity and coordination and the total average score for rapport. Compared with the others, on average participants with PhD degrees gave lower scores for positivity, coordination and the average rapport. We also tested the relationship between participants' knowledge of OGD and OGD portals and the scores for rapport including attention, positivity and coordination. Results of the two independent sample T-tests and one-way ANOVA showed there was no statistically significant relationship between participants' past experience with OGD and OGD portals and their rapport with the help functions on OGD portals.

Table 7-22 Analysis of socio-demographic background and rapport

Aspect	*Gender*	*N*	*Avg.*	*Std.*	*T-test of Gender*				*Age*		*Education*		*Occupation*		*Income*	
					F	*P-*	*t*	*p*	*F*	*P-*	*F*	*P-*	*F*	*P-*	*F*	*P-*
Attention	Male Female	34 44	3.82 4.27	1.132 0.874	1.993	0.162	-2.011	**0.048***	0.160	0.958	2.051	0.114	1.172	0.320	1.255	0.295
Positivity	Male Female	34 44	3.92 4.13	0.755 0.790	0.260	0.612	-1.22	0.226	0.446	0.775	3.275	**0.026***	1.731	0.075	1.216	0.311
Coordination	Male Female	34 44	4.06 4.24	0.929 0.854	0.027	0.870	-0.894	0.374	0.498	0.737	3.942	**0.011***	1.467	0.154	2.302	0.066
Average	Male Female	34 44	3.98 4.22	0.860 0.800	0.145	0.704	-1.275	0.206	0.394	0.813	3.680	**0.016***	1.490	0.145	1.881	0.123

We tested the correlation of participants' trustworthiness and rapport scores with the help pages/virtual agents. Analysis showed these two had significant medium positive correlations, with sig.<0.001, Pearson's r=0.448.

7.3.10 Further analysis by task accuracy

Among the five tasks in the experiment, two of them were about finding answers to the questions by referring to certain datasets on the portal. Participants needed to know how to search, download and use the datasets on the portal to correctly complete both tasks. By calculating the number of participants who correctly answered both questions, we found that only 41 out of the 79 participants got the right answers for both questions, indicating that 51.9% of the participants knew how to use the portal to solve certain problems. To find out whether there were any differences between participants who could correctly complete these two tasks and those who could not, we carried out analysis based on their two-task accuracy rates. We separated the participants into two subgroups: the first included those who did not get the correct answers to both questions, while the second included those who got the right answers to both questions.

We tested the relationship between participants belonging to the control group or the experimental group with their accuracy rate for these two tasks. The chi-square test results in Table 7.23 for the two subgroups, together with whether they belonged to the control group or the experimental group, show no statistically significant differences. Both the control group and the experimental group had 19 participants who could not correctly complete the two tasks. The control group had 19 participants who completed the tasks correctly and the experimental group had 22 such participants. We also tested the relationship between participants' socio-demographic background and their accuracy rate for these two tasks. The chi-square test results in Table 7.23 show a statistically significant difference among participants of different occupations. Compared with participants of other occupations, there were more participants working as salesmen, logistic workers and civilian posts who correctly completed both tasks. We also tested the relationship between participants' past experience with OGD and OGD portals, including whether they had used OGD or OGD portals before and their usage frequency, and their accuracy rate for these two tasks. Chi-square tests showed no statistically significant difference. Thus, whether the participants had used OGD or OGD portals before did not benefit their ability to correctly find the datasets they needed on an OGD portal.

Table 7-23 Chi-square tests of demographic background and two-task accuracy rate

Aspect	Pearson Chi-square	df	P-
Ctrl./Exp.	0.106	1.000	0.745
Gender	1.096	2.000	0.578
Age	4.140	4.000	0.387
Education	0.285	3.000	0.963
Occupation	22.905	13.000	**0.043***
Monthly Income	8.711	4.000	0.069
Use OGD	0.006	1.000	0.941
Frequency	7.203	4.000	0.126
Use OGD Portal	3.703	1.000	0.054
Frequency	7.349	5.000	0.196

We tested the relationship between participants' accuracy rate for these two tasks and their scores for the TAM scales. Results of the two independent sample T-tests are shown in Table 7.24. For the importance of each function, there were statistically significant differences in the average importance scores for help/virtual agent, metadata and feedback between participants who completed the tasks correctly and those who did not. Compared with those who did not complete the tasks correctly, participants who correctly completed the tasks, on average, gave higher importance scores for help/virtual agent, metadata and feedback. For the PEOU of each function, participants who completed the tasks correctly, on average, gave statistically significantly higher PEOU scores for providing data in various formats and open, machine-readable data than those who did not. For the PU, participants who completed the tasks correctly, on average, gave statistically significantly higher PU scores for providing data in various formats, filtering of search results, help/virtual agent, keyword search, metadata, visualization and the average PU score for all functions.

Table 7-24 T-tests of TAM scales and two-task accuracy rate

Aspect	T-test of Importance								T-test of PEOU								T-test of PU							
	Accuracy	N	Avg.	Std.	F	P-	t	p	Accuracy	N	Avg.	Std.	F	P-	t	p	Accuracy	N	Avg.	Std.	F	P-	t	p
Browse	Wrong	38	3.82	1.010	1.009	0.318	0.288	0.774	Wrong	38	3.87	0.811	0.085	0.771	-0.967	0.337	Wrong	38	4.00	0.870	0.253	0.617	-0.768	0.445
	Right	41	3.76	0.830					Right	41	4.05	0.846					Right	41	4.15	0.823				
Various format	Wrong	38	3.39	0.974	0.021	0.884	-0.996	0.323	Wrong	38	3.37	0.942	4.657	0.034	-3.141	0.002**	Wrong	38	3.42	0.976	10.424	0.002	-3.254	0.002**
	Right	41	3.61	0.945					Right	41	4.00	0.837					Right	41	4.05	0.705				
Download without registration	Wrong	38	3.76	1.195	1.519	0.222	0.507	0.614	Wrong	38	4.21	0.843	2.794	0.099	-0.914	0.363	Wrong	38	4.26	1.032	1.927	0.169	-0.025	0.980
	Right	41	3.63	1.067					Right	41	4.37	0.662					Right	41	4.27	0.807				
Filter	Wrong	38	3.87	1.018	2.417	0.124	-1.228	0.223	Wrong	38	3.53	1.133	7.579	0.007	-1.747	0.085	Wrong	38	3.71	1.088	5.540	0.021	-2.153	0.035*
	Right	41	4.12	0.812					Right	41	3.93	0.877					Right	41	4.17	0.771				
Help/Avatar	Wrong	38	2.79	1.094	0.284	0.596	-3.639	0.000**	Wrong	38	3.45	1.032	0.955	0.331	-1.787	0.078	Wrong	38	3.18	1.062	2.168	0.145	-3.229	0.002**
	Right	41	3.63	0.968					Right	41	3.85	0.989					Right	41	3.88	0.842				
Keyword search	Wrong	38	4.34	0.781	0.314	0.577	-0.390	0.698	Wrong	38	3.74	1.155	0.500	0.482	-1.119	0.267	Wrong	38	4.00	1.040	2.476	0.120	-2.207	0.030*
	Right	41	4.41	0.865					Right	41	4.02	1.129					Right	41	4.44	0.709				
Metadata	Wrong	38	2.92	1.124	0.348	0.557	-3.208	0.002**	Wrong	38	3.29	0.927	0.281	0.598	-1.204	0.232	Wrong	38	3.26	0.921	0.173	0.679	-2.249	0.027*
	Right	41	3.68	0.986					Right	41	3.54	0.897					Right	41	3.76	1.019				
Machine-processable	Wrong	38	3.61	1.104	4.747	0.032	-0.988	0.327	Wrong	38	3.55	0.921	4.210	0.044	-2.396	0.019*	Wrong	38	3.82	0.955	6.687	0.012	-1.605	0.113
	Right	41	3.83	0.892					Right	41	4.02	0.821					Right	41	4.12	0.714				
Ranking	Wrong	38	3.39	1.128	2.954	0.090	0.235	0.815	Wrong	38	3.58	0.793	0.132	0.717	-0.968	0.336	Wrong	38	3.50	1.007	7.613	0.007	-1.541	0.128
	Right	41	3.34	0.883					Right	41	3.76	0.830					Right	41	3.80	0.715				
Data request	Wrong	38	3.29	1.063	2.006	0.161	-1.590	0.116	Wrong	38	3.74	0.921	0.197	0.658	-0.093	0.926	Wrong	38	3.68	0.933	2.543	0.115	-1.342	0.184
	Right	41	3.63	0.859					Right	41	3.76	0.916					Right	41	3.95	0.835				
Feedback	Wrong	38	3.45	1.155	11.011	0.001	-2.511	0.014*	Wrong	38	3.61	0.946	0.066	0.798	0.322	0.749	Wrong	38	3.47	1.059	11.265	0.001	-1.846	0.069
	Right	41	4.02	0.851					Right	41	3.54	0.951					Right	41	3.85	0.727				
Visualize	Wrong	38	3.71	0.867	0.201	0.655	-0.653	0.516	Wrong	38	3.61	1.079	1.882	0.174	-1.109	0.271	Wrong	38	3.45	0.950	4.455	0.038	-2.923	0.005**
	Right	41	3.85	1.062					Right	41	3.85	0.910					Right	41	4.02	0.790				
Average	Wrong	38	3.53	0.663	0.283	0.596	-1.784	0.078	Wrong	38	3.63	0.644	1.892	0.173	-1.914	0.059	Wrong	38	3.65	0.655	3.946	0.051	-2.912	0.005**
	Right	41	3.79	0.623					Right	41	3.89	0.587					Right	41	4.04	0.538				

We tested the relationship between participants' accuracy rate for these two tasks and their scores for the DOI scales for the portal. Results of the Mann–Whitney U tests in Table 7.25 show statistically significant differences between participants who completed the two tasks correctly and those who did not in scores for compatibility, complexity, trialability and observability. Compared with those who did not get the right answers to the two tasks, participants who completed the tasks correctly gave higher scores for compatibility, complexity, trialability and observability. We also tested the relationship between participants' accuracy rate for these two tasks and their perceptions of trustworthiness towards the portal. Results of the Mann–Whitney U tests in Table 7.25 show statistically significant differences between participants who completed the two tasks correctly and those who did not. Compared with those who did not get the correct answers to the two tasks, participants who completed the tasks correctly gave higher scores for the trustworthiness of the OGD portal.

Table 7-25 Mann–Whitney U tests of DOI and trustworthiness scales and two-task accuracy rate

Aspect	*Mann-Whitney U test*						
	Accuracy	*N*	*Mean Rank*	*Sum of*	*U*	*Z*	*p*
Relative Advantage	Wrong	38	35.22	1338.50	597.500	-1.790	0.073
	Right	41	44.43	1821.50			
Compatibility	Wrong	38	31.96	1214.50	473.500	-3.011	**0.003****
	Right	41	47.45	1945.50			
Complexity	Wrong	38	33.18	1261.00	520.000	-2.550	**0.011***
	Right	41	46.32	1899.00			
Trialability	Wrong	38	34.80	1322.50	581.500	-1.973	**0.049***
	Right	41	44.82	1837.50			
Observability	Wrong	38	34.55	1313.00	572.000	-2.089	**0.037***
	Right	41	45.05	1847.00			
Trustworthiness	Wrong	38	34.41	1307.50	566.500	-2.110	**0.035***
	Right	41	45.18	1852.50			

We tested the relationship between participants' accuracy rate for these two tasks and their rapport with the help pages/virtual agent. Results of the two independent sample T-tests in Table 7.26 show statistically significant differences in the average rapport scores between participants who correctly completed these two tasks and those who did not. Compared with those who did not give the correct answers to the two tasks, participants who correctly completed these two tasks, on average, gave significantly higher scores for attention, positivity and coordination, as well as the average rapport score.

Table 7-26 T-tests of rapport scales and two-task accuracy rate

Rapport	*Two independent samples T-test*							
	Accuracy	*N*	*Avg.*	*Std.*	*F*	*P-*	*t*	*p*
Attention	Wrong	38	3.69	0.892	1.823	0.181	-3.431	**0.001****
	Right	41	4.42	0.991				
Positivity	Wrong	38	3.80	0.744	2.087	0.153	-2.448	**0.017***
	Right	41	4.22	0.781				
Coordination	Wrong	38	3.82	0.896	0.090	0.765	-3.430	**0.001****
	Right	41	4.46	0.765				
Average	Wrong	38	3.79	0.806	0.315	0.576	-3.401	**0.001****
	Right	41	4.39	0.754				

We finally analyzed the relationship between experimental group participants' accuracy rate for these two tasks with the DOI scales for the virtual agent. Results of the Mann–Whitney U test in Table 7.27 show statistically significant differences in the scores for relative advantage, compatibility and trialability. Compared with the participants who did not give the correct answers to the two tasks, those who correctly completed the tasks gave higher scores for relative advantage, compatibility and trialability for the virtual agent.

Table 7-27 Mann–Whitney U tests of DOI for virtual agent and two-task accuracy rate

Aspect	*Mann-Whitney U test*						
	Accuracy	*N*	*Mean Rank*	*Sum of*	*U*	*Z*	*p*
Relative Advantage	Wrong	38	15.82	300.50	110.500	-2.584	**0.010***
	Right	41	25.48	560.50			
Compatibility	Wrong	38	16.00	304.00	114.000	-2.495	**0.013***
	Right	41	25.32	557.00			
Complexity	Wrong	38	18.18	345.50	155.500	-1.404	0.160
	Right	41	23.43	515.50			
Trialability	Wrong	38	16.95	322.00	132.000	-2.065	**0.039***
	Right	41	24.50	539.00			
Observability	Wrong	38	17.50	332.50	142.500	-1.770	0.077
	Right	41	24.02	528.50			

7.4 Discussion

In this section, we further analyze the collected data for the connections among variables and predicting the possibility of utilizing the OGD portal effectively. To analyze the relationships between variables and to predict OGD users' attitudes towards usage of the portal, we chose to build decision tree models for the five variables of DOI through machine learning. To predict OGD users' possibility of using the OGD portal, we also built decision tree models through machine learning. Details of these models are explained below.

7.4.1 OGD users' attitudes towards OGD portal usage

Through the literature review, we noted that in the field of open data and e-government services, compared with the variables in TAM, there is limited research about the five variables in DOI or the relationships among these variables (Rana et al., 2015). Weerakkody et al. (2017) and Carter and Bélanger (2005) have shown support for using relative advantage, compatibility, complexity and observability to explain citizens' intentions to use open data. However, the variables that affect these aspects are not clear. Thus, we chose to use machine learning to build models for explaining the five variables of DOI in the field of OGD.

Five new variables were created as the dependent variables (targets) for building the models, which are intention–relative advantage (IRA), intention–compatibility (ICP), intention–complexity (ICL), intention–trialability (ITR) and intention–observability (IOB). To create the five dependent variables, we first separated the variables of relative advantage, compatibility, complexity and trialability into two kinds: negative and positive. Because the scales for these five aspects are 6-point Likert scales, if a score was less than or equal to 3, we treated it as negative; if a score was larger than 3, we treated it as positive. Then we calculated the 3-quantile of the positive scores to find the cutoff points. According to the results, we chose 4.5 as the 1/3 cutoff point for relative advantage, compatibility, trialability and observability, and 4 as the 1/3 cutoff point for complexity. We chose 5 as the 2/3 cutoff point. By using these two cutoff points, we separated the positive scores for these five variables into three categories: low, moderate and high. We marked these five variables to form the new variables as the targets for building the models. We can see from Table 7.28 that the negative ones were too few for building models, thus we excluded those records. We used systematic sampling to select 2/3 of each category (low, moderate and high) as the training data and 1/3 of each category as the testing data.

Table 7-28 Cutoff points and categorization of DOI variables

Aspect		Relative Advantage	Compatibility	Complexity	Trialability	Observability
Value	1/3 cutoff point	4.5	4.5	4	4.5	4.5
	2/3 cutoff point	5	5	5	5	5
N	Negative	1	1	2	3	6
	Low	28	27	27	25	24
	Moderate	29	25	25	28	35
	High	21	26	25	23	14
	Total	79	79	79	79	79
	Training	53	53	52	52	49
	Testing	25	25	25	24	23

We used an SPSS modeler for building the models for the five target variables. The process stream of building the models is shown in Figure 7.12. We first used the training datasets as the sources for the models. A filter node was added to exclude the unnecessary independent variables. In our process of building models for the DOI variables, we selected 61 input variables as the independent variables, including scenario marks for the control group and experimental group, gender, age, education background, occupation, monthly income, 5 TIPI variables, trust tendency, 4 variables about past OGD experience, habits of playing computer games, importance, PEOU and PU of TAM for 12 OGD portal functions, DOI variables except the target variable, trustworthiness of the OGD portal and the 3 rapport variables. Then we added the type node to select the target variable for building the model. Different model nodes were added representing different methods for building the decision tree models, including Chaid, Cart and Quest. An analysis node was added to calculate the accuracy rate of the model for the training data. We then tested the built model on the testing data for the accuracy rate. We repeated the process to build models for all five DOI variables.

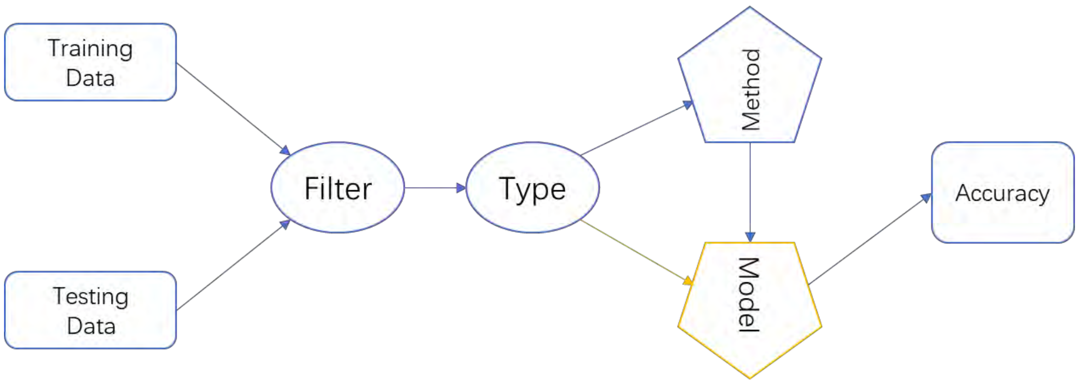


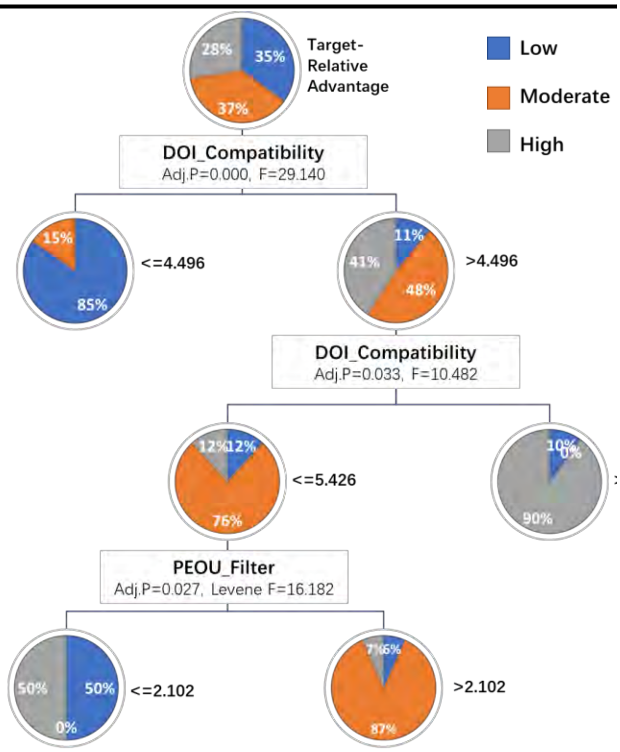
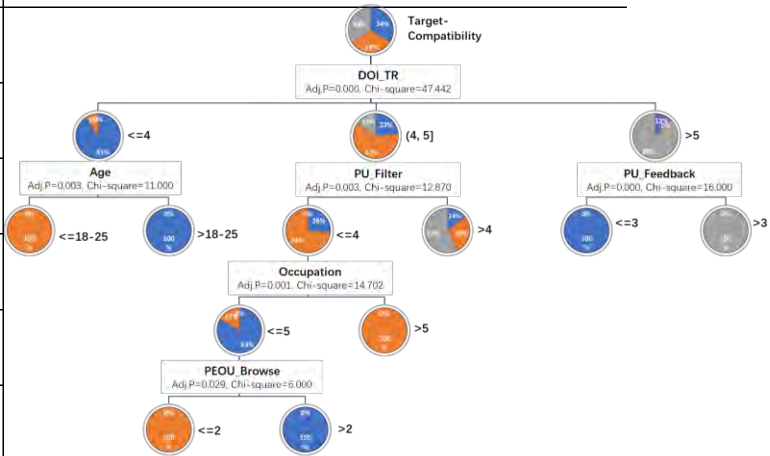
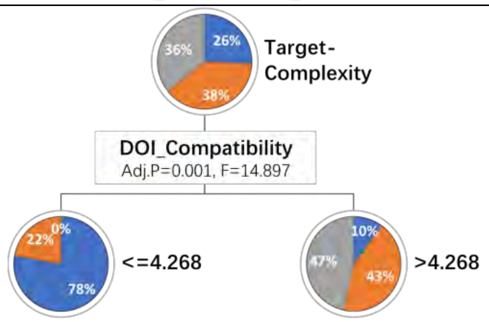
Figure 7-12 Process for building DOI models

A comparison of the accuracy of the built models is shown in Table 7.29. We listed the accuracy rates of the models built for each target variable through different methods. The best-performing models were selected based on the accuracy rate of the model for both the training dataset and the testing dataset.

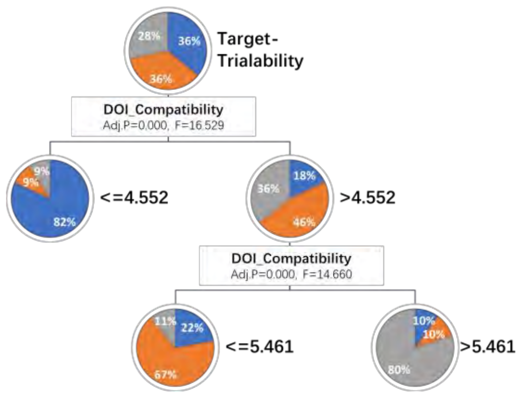
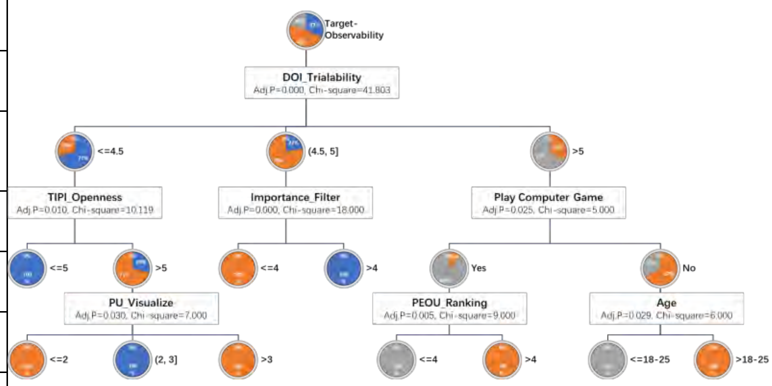
Table 7-29 Accuracy rates of models for DOI variables

Aspect	*Data Source*	*Accuracy*		
		Chaid	*Cart*	*Quest*
Relative Advantage	Training	88.68%	81.13%	**79.25%**
	Testing	60.00%	48.00%	**72.00%**
Compatibility	Training	**94.34%**	33.96%	73.58%
	Testing	**56.00%**	32.00%	56.00%
Complexity	Training	98.08%	**59.62%**	57.69%
	Testing	44.00%	**52.00%**	52.00%
Trialability	Training	96.15%	84.62%	**71.15%**
	Testing	45.83%	45.83%	**75.00%**
Observability	Training	**100.00%**	84.00%	48.00%
	Testing	**65.22%**	52.17%	47.83%

Table 7-30 Models for DOI variables[6]

Aspect	*Variable*	*Weigh*	*Model*
Relative Advantage (IRA)	Compatibility	0.71	
	PEOU-Filter	0.1	
	PEOU-Visualize	0.01	
	PEOU-Keyword Search	0.01	
	PEOU-Help	0.01	
	Importance-Request	0.01	
	PU-Request	0.01	
	Importance-No registration	0.01	
	Observability	0.01	
	PU-Help	0.01	
Compatibility (ICP)	Trialability	0.42	
	PU-Feedback	0.27	
	PU-Filter	0.23	
	Occupation	0.04	
	Age	0.03	
	PEOU-Browse	0.01	
Complexity (ICL)	Compatibility	0.33	
	Relative Advantage	0.13	
	Use OGD frequency	0.13	
	Trust	0.13	
	Trialability	0.13	
	Use OGD portal frequency	0.13	
Trialabili	Compatibility	0.45	

[6] For some decision trees not all of the important variables were included for splitting due to stopping criteria for model building.

ty (ITR)	PU-Browse	0.06	
	PEOU-Browse	0.06	
	PU-Filter	0.06	
	PU-No registration	0.06	
	Trust	0.06	
	Complexity	0.06	
	Relative Advantage	0.06	
	Use OGD frequency	0.06	
	Use OGD portal frequency	0.06	
Observability (IOB)	Trialability	0.49	
	Importance-Filter	0.16	
	TIPI-Openness to experiences	0.15	
	PEOU-Ranking	0.09	
	Age	0.08	
	Play computer game	0.02	
	PU-Visualize	0.02	

Details of the best-performing models for each target are displayed in Table 7.30. For relative advantage, the model showed the strong relationship between compatibility and relative advantage. OGD users with a strong positive attitude to the compatibility of the OGD portal were more likely to also feel the strong relative advantage of the OGD portal. For compatibility, the model showed the strong positive effect of trialability and PU of the feedback function. If the OGD users had a chance to try using the OGD portal and feel the usefulness of the feedback function, they were more likely to have a strong positive view of the compatibility of the OGD portal. If the OGD users were older than 25 years and had not had the chance to try OGD portals, they were more likely to have a negative attitude towards the compatibility of the OGD portal. For the complexity of the OGD portal, the model also showed a strong effect of compatibility. OGD users' stronger feeling for the compatibility of the OGD portal led to a stronger feeling for the ease of use of the portal. For the trialability of the OGD portal, the model showed a strong effect of compatibility. OGD users' stronger feeling for the compatibility of the OGD portal led to a stronger feeling for the trialability of the portal. For the observability of the OGD portal, the model showed the strong effect of trialability. If the OGD users had a stronger feeling for the trialability of the OGD portal, they were more likely to also have a stronger positive attitude to the observability of the OGD portal.

A relationship map of all variables in the models of DOI is shown in Figure 7.13. We used software called Gephi[7] to create the relationship map. Force Atlas 2 was used as the layout algorithm for all the nodes and the lines between them. The size of the nodes indicates the effects of the variable on other variables. A larger node means that this variable affected a higher number of other variables. The distance between nodes indicates the strength of the relationship between two variables. The thickness of lines indicates the weight of the effect of one variable on another variable. We can see from the figure that compatibility and trialability are of great importance due to their strong effects on other variables. These two variables also have strong relationships between each other. Functions' importance, PEOU and PU, which have an effect on observability, compatibility and relative advantage, are quite different. Filtering of search results is the only function of the OGD portals whose importance, PEOU and PU are all included in models for DOI. The thickness of the lines also indicates filtering's great effect on compatibility, relative advantage and observability. OGD users' usage frequency of OGD and OGD portals has effects on both the trialability and the complexity of OGD portals. OGD users' perception of trustworthiness of the OGD portal also affects their feelings of trialability and complexity for the portal. OGD users' age has an effect on their feelings for the observability and complexity of the OGD portal.

[7] See the introduction to this software at https://gephi.org/

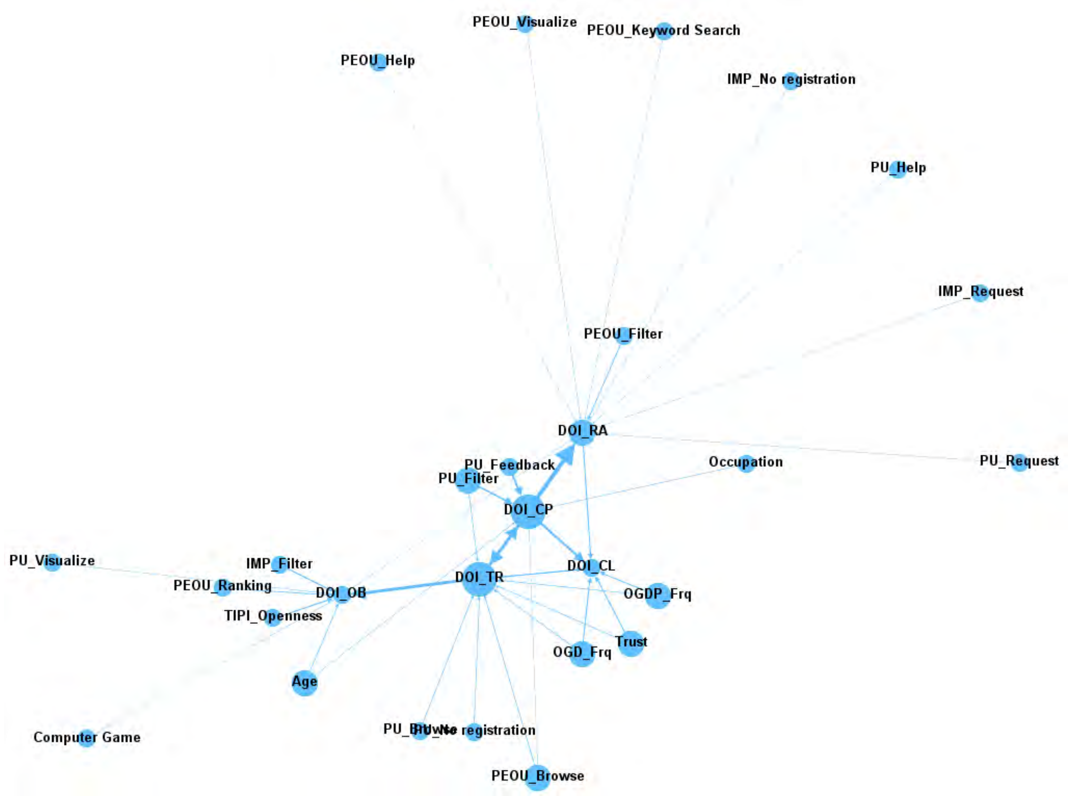


Figure 7-13 Relationship map of the variables in DOI models

In the previous meta-analysis of existing research on citizens' adoption of e-government services, Rana et al. (2015) confirmed the effect of compatibility on PU. Our analysis above shows that conversely, PU also affects the compatibility of an OGD portal. The study of Carter and Bélanger (2005) drew from previous studies to conclude that the complexity of DOI and PEOU of TAM, relative advantage of DOI and PU of TAM are all the same construct. The models we built based on the experiments have shown the effects of both PU and PEOU from TAM on the relative advantage of DOI, while no significant effect of PU or PEOU on complexity has been recognized. Instead, complexity is more related to compatibility and trustworthiness. This may be because the variables of TAM that we used in our study are about the PEOU and PU of specific functions on OGD portals rather than the PU and PEOU of the portal itself. But this also indicates the need for further analysis of the relationship between complexity and PEOU, as well as between relative advantage and PU.

7.4.2 OGD users' possibility of using OGD portals

Previous studies of open data and e-government mainly focused on understanding users' attitudes towards the adoption of e-government and their intention to use the services (Rana et al., 2015), but seldom focused on the actual usage of open data. Although studies have supported that citizens' behavioral intention to use e-government services determines their actual system usage (Carter & Belanger, 2004), no studies have built models to explain citizens' actual utilization of open data and open data portals. Thus, by combining the research data collected during our experiment, we wanted to build a research model to predict citizens' possibility of using OGD portals accurately.

To build such a model, a new variable was created as the dependent variable (target) for building the model, which is the ability to use. This is a Boolean logical variable indicating whether a user can use the OGD portal effectively. It was created by analyzing whether the participants in the experiment had correctly completed the two tasks asking them to find certain datasets on the portal. If they gave the correct answers to both tasks, the variable has a value of 1, otherwise it has a value of 0.

We also used an SPSS modeler for building this model to predict citizens' possibility of using OGD portals effectively. The process stream for building this model is shown in Figure 7.14. We first added a source node to link the dataset to the model. A filter node was added to exclude unnecessary independent variables. In this process of building the model, we only included 13 variables in the research model for the design of this study, which were the scenario marks for the control group and the experimental group, the average importance score for the 12 functions, the average PEOU and PU scores for the 12 functions, the average scores for the 5 DOI variables, trustworthiness of the portal and the 3 rapport variables. Then we added the type node to select AU as the target variable for building the model. Because AU is a Boolean logical variable, we chose C5.0 as the method for building the model. Ten-fold cross-validation was used to increase the performance of the model. Finally, an analysis node was added to calculate the accuracy rate of the model for the dataset.

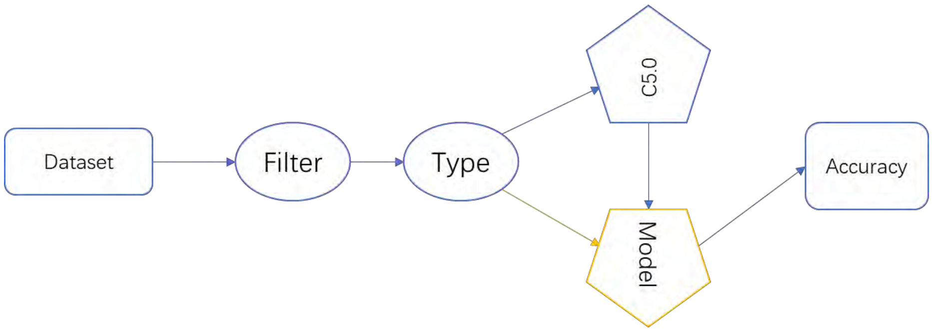
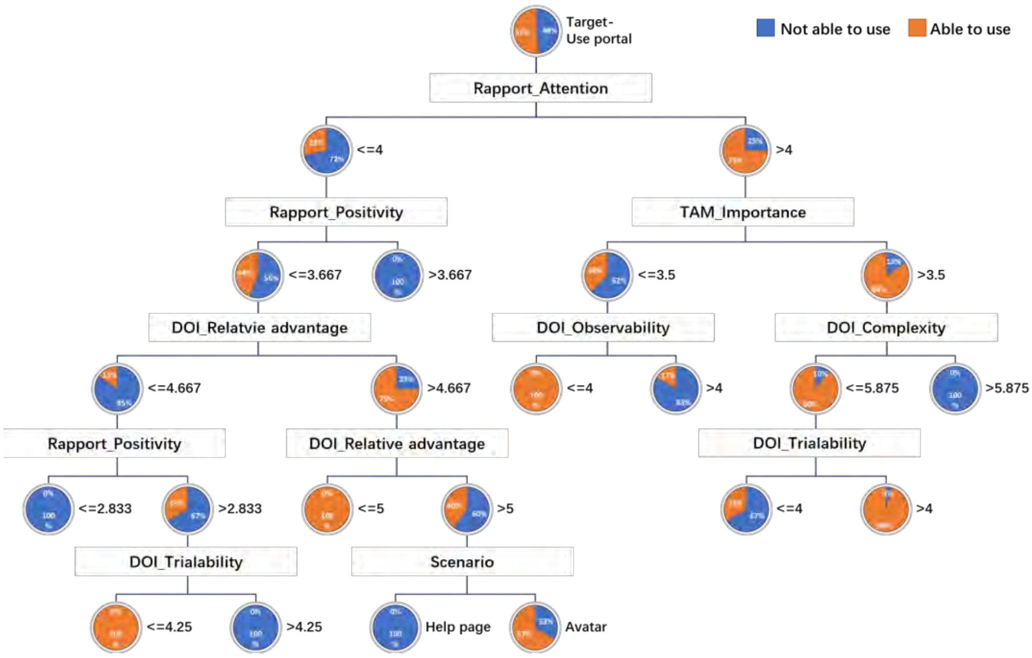


Figure 7-14 Process for building model to predict citizens' possibility of using the OGD portal

The results show the accuracy rate of the built model after cross-validation was 93.67%. Details of the research model are shown in Figure 7.15. Eight out of 14 variables were included in this research model for predicting citizens' possibility of using the OGD portal. The attention variable belonging to the rapport scale is the most important variable in this model, followed by the relative advantage and trialability of the DOI. We also noted that all DOI variables except compatibility are included in the model. But the TAM variables and trustworthiness are not included in the model. We can see from the model that high complexity and observability scores will lead to citizens failing to use the OGD portal. A moderate relative advantage score indicates citizens' possibility of using the portal. Using a virtual agent instead of the traditional help pages will increase citizens' possibility of using the portal.



Weight of variables: Rapport-Attention=0.35, DOI_RA=0.22, DOI_TR=0.15, Rapport-Positivity=0.13,

TAM_Importance=0.08, DOI_OB=0.06, DOI_CL=0.02, Scenario=0.01

Figure 7-15 Model for predicting citizens' possibility of using the OGD portal

Comparing the model for predicting citizens' possibility of using the OGD portal with related studies, we found both agreements and disagreements. In the study of Weerakkody et al. (2017), which used the DOI to explain citizens' behavioural intention to use open data, the most important variable in their validated model was relative advantage, which is also supported in our model where the relative advantage of DOI received the highest weight among all DOI variables. Weerakkody et al. (2017) also validated observability as the third most important variable to predict citizens' behavioural intention to use open data, which is supported by our model. However, the models of Weerakkody et al. (2017) and Carter and Bélanger (2005) based on DOI theory showed a significant influence of compatibility on users' behavioural intention. This variable is not included (identified as not as important by an impurity measure of the decision tree) in our model to predict citizens' possibility of using the OGD portal. This may be due to our use of participants' actual utilization of the OGD portal as the dependent variable instead of asking participants about their intention to use OGD portals.

In addition, we noted that trialability and observability, which are often excluded by other scholars (Carter & Bélanger, 2005; Rana et al., 2015) actually had great effects on participants' utilization of the OGD portal. This may be due to others' analyses being mainly based on surveys instead of experiments. Our experiment actually offered the participants a chance to try out using an OGD portal, which went further than asking participants about their intention to use the portal and allowed analysis based on actual usage.

Also, although the studies of Carter and Bélanger (2005) and Venkatesh et al. (2016) supported the significant postive effect of citizens' perception of trustworthiness on their intention to use e-government services, our model for predicting citizens' utilization of OGD portals did not select trustworthiness as an essential factor. We suggest that this may be because previous studies dealt with e-government services, while our model focused on a specific kind of portal built by governments. OGD portals have different characteristics from other e-government websites since they provide citizens with data resources and services relating to the utilization of data. Another possible reason is that our study and previous studies are based on different sizes and areas of samples.

7.5 Conclusion

Since OGD portals act as a bridge between the supply-side and demand-side of OGD, the acceptance and usage of OGD portals thus became the premise for improving OGD utilization. In this chapter, we have focused on the research questions regarding citizens' acceptance and usage of OGD portals. A between-subject experiment has been designed and carried out based on a combination of technology acceptance theories including TAM, DOI, trustworthiness and rapport.

Through analysis of the two pilot studies, we noted that some individuals preferred traditional help functions while others preferred the virtual agent. They preferred traditional help functions due to their efficiency, while virtual agents were preferred for their fun and friendly conversation functions. Different users may have different needs and use habits; thus, it would benefit users to have both traditional help functions and a virtual agent for them to choose from. A future study could be carried out to test users' choices when given both help options at the same time, rather than only one option at a time as in our pilot studies and formal experiments.

Through analysis of the experimental data, we discovered that citizens showed good performance in completing the experimental tasks by using the imitated OGD portal with the best usability, which answered in the affirmative RQ4.2 about whether citizens can use the OGD portal with the best usability from the usability evaluation. Regarding RQ4.3 about citizens' acceptance of the OGD portal, we found that the five variables of DOI have close relationships with each other. Compatibility and trialability are of great importance due to their strong effects on the other variables of our research model. For RQ4.4 about citizens' actual utilization of the OGD portal, the attention variable belonging to the rapport scale weighted the highest among all the decisive factors for predicting citizens' OGD utilization, followed by the relative advantage and trialability of the DOI. On the other hand, the TAM in relation to different functions of OGD portals and trustworthiness were shown to have little effect on citizens' actual usage of OGD portal. Finally, for RQ4.5 about the effects of traditional help functions compared to offering help through a virtual assistant, our experiment has shown no significant differences between these two different ways of help, but the accuracy rate for completing the two tasks was higher in the group with the virtual assistant. Our model predicting citizens' usage of the OGD portal also shows that providing a virtual assistant will increase the possibility of using the portal.

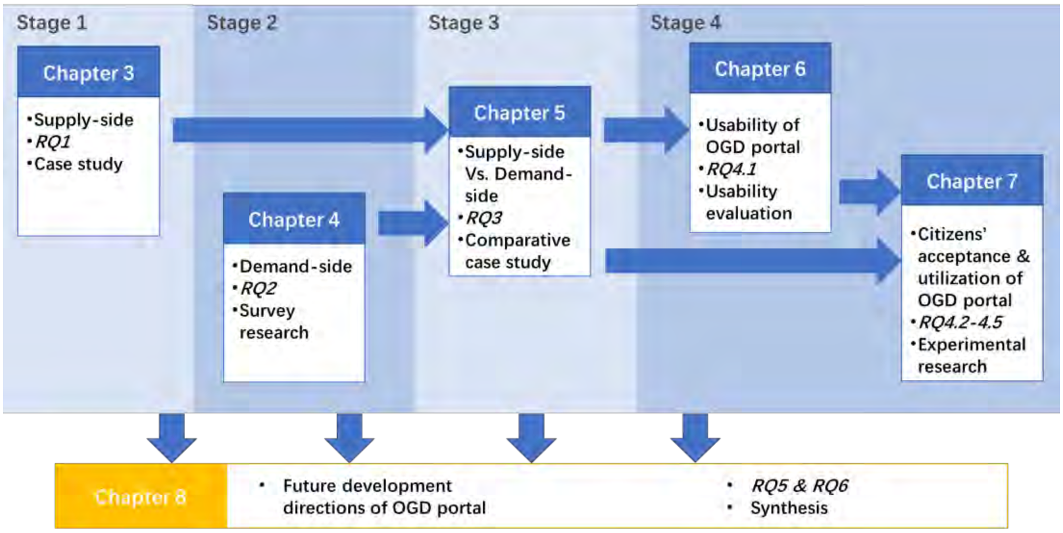
Other important conclusions include:

(1) Citizens' higher acceptance intention of the OGD portal (a higher DOI score) and higher rapport with the help functions will lead to their higher accuracy rate for completing tasks on OGD portals.

(2) Significant positive relationships are identified among the importance, the perceived ease of use and the perceived usefulness of OGD portals' functions.

(3) Citizens' demographic background has an effect on their acceptance and utilization of OGD portals. Their occupation and past experience with OGD affect their acceptance of OGD portals and their rapport with the help functions. Their socio-demographic background including gender and occupation also affects their feelings for the importance, PU and PEOU of different functions.

8 CHAPTER 8: DISCUSSION



In previous chapters, we have investigated the supply-side and demand-side of OGD separately and analyzed the present relationships between these two sides. Because OGD portals are a bridge linking these two sides where activities that utilize OGD mainly happen, we thus focused on the usability of OGD portals and the acceptance and utilization of OGD portals from the perspective of citizens as the key users. Based on different research methods, various data has been collected and different conclusions have been drawn from different aspects of the analysis. The four stages of these previous studies are not separate but related to each other. In this chapter, we address the final two research questions of this thesis:

RQ5. How to connect the supply-side and demand-side of OGD through portals?

RQ6. What are the future directions for developing OGD portals ?

In order to do this, we synthesized the results from previous chapters for their agreement and relevance and combined our discussion with related literature.

8.1 Connecting the supply-side and demand-side

We base our discussion of RQ5 about how to connect the supply-side and demand-side of OGD through portals on the key conclusions of Chapter 5, the comparison of the supply-side and the demand-side. In order to solve the conflicts between the two sides and to improve the utilization of OGD portals, we propose four methods for connecting the supply-side and demand-side of OGD by applying the results from the other chapters as well as from the literature. Details of each method are explained below.

8.1.1 Improving observability and trialability

Our evaluation of the OGD portals has shown their relatively low visibility on the internet. Relating to the number of datasets on a portal, this low online visibility reflects the requirement to provide more open data resources online (OpenDataBarometer, 2018a). Our analysis of citizens' knowledge and past experience of OGD and its utilization has also shown the limited impact of OGD on the public and wider society, in line with other studies (Zuiderwijk et al., 2013). Citizens' not knowing about the existence of OGD has resulted in low utilization of the existing OGD resources and the weak impact of OGD development (OpenDataBarometer, 2018a; Ribeiro, 2017). Our research model for predicting citizens' utilization of OGD portals shows the effect of the low observability of OGD portals on citizens' usage of the portals. Therefore, improving the observability of OGD portals would help citizens to become conscious of the data being opened up by the government, which is the first step in possible utilization.

The model for predicting the observability of OGD portals by citizens in Chapter 7 has shown the great effect of trialability on citizens' awareness of OGD and the operation of OGD portals. Trialability is an aspect belonging to DOI which is usually ignored by other scholars due to doubt about its effect on the intention to use (Carter & Bélanger, 2005). The experiment in Chapter 7 offered us a chance to collect data on trialability because most of the participants had never heard of or used OGD and OGD portals before. The strong effect of trialability on observability and citizens' utilization of the OGD portal emphasizes the importance of improving the trialability of OGD portals in order to encourage citizens' utilization.

8.1.2 Strengthening relative advantage

Different OGD principles and frameworks have emphasized the specific characteristics of OGD, including timeliness (OpenGovernmentWorkingGroup, 2007; SunlightFoundation, 2010; Ubaldi, 2013), no registration required for downloading data (Bogdanović-Dinić et al., 2014; SunlightFoundation, 2010; Ubaldi, 2013), machine-readability (Lourenço, 2015; OpenDataIndex, 2017; Thorsby et al., 2017) and ability to download in bulk (OpenDataBarometer, 2016; SunlightFoundation, 2010; Ubaldi, 2013). These features are essential for providing data of good quality (OpenDataBarometer, 2018a). Our survey on the demand-side of OGD also shows citizens' high demand for these OGD quality elements.

However, when comparing the demands from citizens with the performance of the data on portals, several mismatches have been recognized. Up-to-date and regularly updated data received the highest demands among all offered data qualities in our study of the demand-side, but the data on the OGD portals actually performed worst in these two aspects. The usability evaluation of present OGD portals and the evaluation of the supply-side both reflected poor performance in offering machine-readable data, which is an important usability criterion for OGD portals. Therefore, in order to meet the needs of the demand-side of OGD, it is important to strengthen the relative advantage of OGD from the supply-side, especially in relation to timeliness and machine-readability, in line with the study of (Bogdanović-Dinić et al., 2014).

8.1.3 Improving compatibility

Our comparative analysis of the supply-side and demand-side of OGD has shown the mismatch between the data provided and the data demands of citizens. At present, data opened on the portal that citizens are highly interested in is of low quantity or quality, which confirmed the findings of scholars that governments are not publishing the data needed by citizens (OpenDataBarometer, 2017a; Thorsby et al., 2017). This will not only result in the low utilization of the open datasets (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017) but also limit the effects and benefits of OGD on the whole of society (Attard et al., 2015). The success of open data is measured by its impact on society, especially in improving citizens' lives (OpenDataBarometer, 2018a). Therefore, it is important to improve the compatibility of OGD by publishing data according to the needs of citizens. Our studies have shed light on citizens' OGD demands for different purposes, for different data needs. For OGD development in China, data relating to weather, maps and international trade are in low quantity on OGD portals at present, which does not meet the needs of citizens. Opening up more data in these subjects in the future could help improve the compatibility of OGD for Chinese citizens.

8.1.4 Reducing complexity

Our investigation of the demand-side showed citizens feel difficulty in getting access to OGD, in line with the research of other scholars (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017) about the barriers to open data usage. Our usability evaluation showed the OGD portal's poor efforts to reduce complexity for users when using the portal, including error prevention, user interaction and offering help tips. However, the model for predicting citizens' utilization of OGD portals has shown the importance of reducing complexity, which would result in higher utilization rates. Considering the effect of complexity on citizens' utilization of OGD and OGD portals, it is thus important to reduce the complexity of using OGD portals by breaking through the barriers to citizens' OGD usage

(Ruijer, Grimmelikhuijsen, Hogan, et al., 2017) through actions such as improving filtering functions and offering more help features on the portal to enhance user experience (Thorsby et al., 2017).

8.2 Future directions for OGD portal development

In order to find answers to RQ6 about future directions for developing OGD portals, we base our discussion on the results from the four different stages of the study. The model for predicting users' utilization of OGD portals in Chapter 7 provides the main support for proposing these development directions. Results from the usability evaluation and the comparison of the supply-side and demand-side of OGD are also taken into consideration, combined with information from related literature. Details of the proposed directions for OGD portal development are explained below.

8.2.1 Identifying user types

From the usability evaluation results in Chapter 6, we noted current OGD portals did not identify different types of users when offering services like recommendations and help. However, our analysis of the demand-side of OGD showed significant differences in the demands for OGD categories and qualities, as well as the purposes of utilization, among different people. For example, those with a higher education background use OGD for scientific research and business decisions, while occupation affects perceptions of the importance, perceived ease of use and perceived usefulness of the functions provided by OGD portals.

The study of Venkatesh et al. (2014) also confirmed the role of citizens' socio-demographic background in understanding their portal usage. Our results from the survey and the experiment confirm that citizens' experience and knowledge of OGD and OGD portals also affect their data demands and utilization. The study of Graves-Fuenzalida (2013) further characterized citizens into different groups, including advocates/journalists, civil developers and data consumers. They argued that the different groups have specific purposes relating to their occupation, which agree with the results from our survey in Chapter 5. Because of the significant effects of citizens' background on their utilization behaviour, identifying different types of users and offering OGD services accordingly could thus help to reduce the data utilization barriers and consequently improve the usage of the published data.

8.2.2 Increasing online visibility

The model for predicting citizens' utilization of OGD portals in Chapter 7 has shown the observability of DOI to be one of the decisive variables. However, according to the evaluation of OGD portals in Chapter 1, due to blocking of search engines and limited numbers of datasets on the portals, at present OGD portals' online visibility in China is relatively low. Our survey and other studies (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017) show the majority of citizens do not know about the existence of OGD, which has become a great barrier to encouraging utilization of OGD. Therefore, government could call the attention of the citizens to OGD through the press and television. The model for analyzing citizens' feeling for observability in Chapter 7 showed that if they had a chance to try out using a portal, they were more likely to observe the advantages of OGD. Due to the nature of OGD portals, increasing their online visibility thus becomes a good method for improving citizens' utilization of the data on the portal. According to the annual report of Hootsuite (Hootsuite, 2019), internet users are growing by an average of more than one million every day, which indicates the great number of potential users for the OGD portals. Therefore, stopping the blocking of search engines (Amerland, 2013), increasing the numbers of datasets on the portals (OpenDataBarometer, 2018a) and linking OGD portals to other well-known government websites are reasonable ways to increase their online visibility.

8.2.3 Improving help functions

Our survey in Chapter 5 identified the great demand of citizens for OGD portals to offer help functions when they use the portals to get access to the data they need. The experiment in Chapter 7 relating to different help functions also showed that reasonable help information was useful to the participants when completing tasks on the OGD portal. In addition, the usability evaluation in Chapter 6 showed OGD portals' failure to prevent errors, provide timely feedback to users' questions or help beginners to start using the portal. The failure in the usability design of help functions of OGD portals resulted in technical and access barriers for citizens when using the portals (Ruijer, Grimmelikhuijsen, Hogan, et al., 2017). These barriers increased citizens' feeling for complexity in using the OGD portals. According to the model for predicting citizens' OGD usage in Chapter 7, the easier that citizens find an OGD portal to use, the more likely they are to make proper use of the portal and the data provided. Therefore, improving the help functions provided on the portals is a reasonable approach to enhancing the usability of the portals and consequently encouraging more citizens to accept and use the open data on the portals.

8.2.4 Strengthening user interaction

Rapport was found to be a decisive factor in citizens' actual utilization of the OGD portal according to the model built in Chapter 7. A closer rapport with the help functions would have increased the possibility of completing the tasks about finding the data needed on OGD portals. However, the usability evaluation in Chapter 6 revealed the insufficiency of present OGD portals in providing user interaction functions such as suggestions in searches and user feedback to communicate with the portal as well as with other users. Therefore, in our experiment in Chapter 7, we added a virtual agent to provide help information and to communicate with users when needed. The model for predicting citizens' utilization of OGD portals confirmed the positive effects of the virtual agent on completion of tasks. Other studies also confirmed that the visual presence of virtual agents can enhance the accessibility of websites (Chattaraman et al., 2011) by smoothing the human–computer interaction (Lester et al., 1997). In order to strengthen user interaction on OGD portals as well as improving the usability, providing virtual agents and functions like user feedback are possible methods.

8.2.5 Reinforcing OGD characteristics

OGD has its own characteristics compared with other open data or government data. These characteristics include timeliness, open licencing, accessibility, comparability and being free of charge. OpenDataBarometer (2018a) assured OGD's positive impacts on the whole of society. Governments' ownership of the unique data that can only be obtained from governments such as registration, legislation, transportation and public safety further enhances the value of OGD (Ubaldi, 2013). However, the evaluation of the data on OGD portals in Chapter 3 showed poor delivery of OGD characteristics such as open licencing, downloading in bulk, machine-readable formats and timeliness. These attributes of OGD are actually in great need by citizens, as shown in Chapter 4. The limited numbers of datasets in categories that citizens have great interest in such as maps and environmental quality also indicate that governments may not be opening enough of the data that the public needs (OpenDataBarometer, 2017a). The true value of open data derives from how the public makes use of it (OpenDataBarometer, 2018a; Safarov et al., 2017), thus sufficient resources are a premise of utilization in providing the data needed by the users of OGD. In order to encourage citizens' utilization of OGD on OGD portals, it is important to reinforce OGD characteristics in two aspects. On one hand, the operators of OGD portals, which are usually government departments, should engage more with citizens to understand their data needs. On the other hand, the needed data should be published in ways that citizens can use it and prefer to use it, such as updating the data regularly, providing it in different machine-readable formats and allowing downloading in bulk.

8.3 Conclusion

The discussion of this chapter has focused on the last two research questions of this thesis about how to connect the supply-side and demand-side of OGD through portals and future development directions for OGD portals by drawing together the findings from the previous chapters. These two questions have been answered by considering the findings from Chapters 3 to 7 and the literature review in Chapter 2. By combining the results from previous chapters, four methods are recommended for connecting the supply-side and demand-side of OGD, including improving observability and trialability, strengthening relative advantage, improving compatibility and reducing complexity. Five directions are proposed for the future development of OGD portals, which are identifying different user types, increasing OGD portals' online visibility, improving help functions, strengthening user interaction with the portals and reinforcing OGD characteristics.

9 CHAPTER 9: CONCLUSION

Considering the under-utilization problem of OGD in comparison to the fast development and launch of government data on the supply-side, this thesis tried to focus on improving citizens' utilization of OGD by connecting the supply-side and demand-side of OGD. Several research questions have been proposed to achieve the objectives of this study, as are shown in Table 9.1. In this final chapter, we present a summary of the key conclusions from the previous chapters drawing from the different aspects studied. Following the research design and roadmap described in Chapter 1, the core results from each stage are presented in Section 9.1. The theoretical contributions and pragmatic implications of this thesis are detailed in Section 9.2 for other researchers to consider. The strengths and limitations are described and discussed in Section 9.3. Based on the limitations of this thesis, we propose several possible options for future research in Section 9.4.

Table 9-1 Research objectives and research questions of this thesis

Objectives	Research Questions	
	Level 1	Level 2
O1: Evaluate the supply-side development of OGD.	**RQ1:** How to analyze the development of the supply-side of OGD?	**RQ1.1:**What framework can be used to assess local-level OGD portals? **RQ1.2:**How to obtain priorities for elements of the framework? **RQ1.3:**How to use the framework to analyze the present development of local OGD portals? **RQ1.4:**How to use the framework to guide the future development of local OGD portals?
O2: Understand the demands and motivations of citizens on the demand-side of OGD.	**RQ2:** How can citizens' demand for and utilization of OGD be characterized?	**RQ2.1:** What are citizens' demands for OGDs? **RQ2.2:** What are citizens' demands for and utilization of OGD portal services? **RQ2.3:** What are the relationships between citizens, their OGD demands and OGD utilization?
O3: Recognize the existing conflicts between the demand-side and supply-side of OGD.	**RQ3:** What is the relationship between the supply-side and demand-side of OGD?	**RQ3.1:** What research model could be used for the comparison of the supply-side and demand-side of OGD? **RQ3.2:** How to analyze the present relationship between the supply-side and demand-side of OGD by using the research model?
O4: Find possible solutions for the conflicts and discuss future development directions of OGD portals to improve OGD utilization.	**RQ4:** What are the decisive factors for citizens' utilization of OGD portals?	**RQ4.1:** How to evaluate the usability of current OGD portals? **RQ4.2:** Can citizens use the OGD portal with the best usability from the usability evaluation? **RQ4.3:** What are the factors affecting citizens' acceptance of the OGD portal with the best usability from the usability evaluation? **RQ4.4:** What are the factors affecting citizens' actual utilization of the OGD portal with the best usability from the usability evaluation? **RQ4.5:** What is the effect of using traditional help functions compared to using a virtual assistant?
	RQ5: How to connect the supply-side and demand-side of OGD through portals?	
	RQ6: What are the future directions for developing OGD portals?	

9.1 Key conclusions

As a result of OGD initiatives, OGD portals have been launched by different governments for public access to open data. Although the desired effect of OGD is for its benefits to the whole of society (Attard et al., 2015), scholars have noted lack of utilization has become a great problem in the development of OGD. Therefore, we raised the core research question of this study: *How to improve citizens' utilization of OGD by connecting the supply-side and demand-side through portals?* To find answers to this core research question, we designed a four-stage study covering different aspects of OGD development.

The first stage of the study focused on the supply-side of OGD and developed a prioritized evaluation framework for analyzing the development of the supply-side, which answered the first sub-question (RQ1). A systematic evaluation of province-level OGD portals in China has been carried out. Based on the results of the evaluation, we found that experts emphasized data accessibility and quality more than data quantity. Imbalances in the development of local OGD portals have been recognized, not only in the data categories, but also in aspects such as open license, keeping data up to date and updating data regularly. The evaluation also showed the low online visibility of several local portals, which may be due to blocking search engines from collecting their information. We did not find significant relationships between the context of an area, such as the population or the size and wealth of the local government, and the number of datasets provided on the portal.

The second stage of the study characterized citizens' demand for and utilization of OGD, which answered the second sub-question (RQ2). A survey among Chinese citizens has been designed and distributed to collect data for understanding the demand-side of OGD. Based on the data analysis from the survey, we found that at present, most citizens did not know of or use OGD or OGD portals. Some of those who knew about OGD sand OGD portals chose not to use the data and the portals. Citizens' socio-demographic background including age, occupation and education affected their preferences concerning data subjects (e.g. weather, education) and attributes (e.g. machine-readability, being up to date, being updated regularly). Daily life and anti-corruption were the two most selected purposes of OGD utilization. Results revealed that the purpose and reason for using OGD had an influence on which subject citizens expressed demand for. Citizens' education and knowledge of OGD affect their OGD utilization purposes. The survey also reflected citizens' feeling of difficulty in accessing the data on the portal and their need of help in using the portal.

The third stage of the study examined the relationships between the supply-side and demand-side of OGD, which answered the third sub-question (RQ3). A comparative analysis of the supply-side and demand-side based on the DOI theory identified several interesting results. The analysis firstly confirmed citizens' poor use of OGD and OGD portals. Conflicts in the relative advantage, compatibility and complexity of the two sides were identified, uncovering that the supply-side did not meet the needs of the demand-side. For relative advantage, providing timely datasets by updating data regularly and keeping data up to date was identified as the key issue to address on the supply-side, since this is the quality most required of OGD by the demand-side. For compatibility, the results showed that governments' greatest failure is in providing enough data in the categories that citizens are demanding most. For complexity, the greatest difference between the two sides lies in providing an online smart or virtual agent to help with using the portal.

The fourth stage of the study focused on the bridge between the supply-side and the demand-side of OGD, which is the OGD portal. We tried to find the decisive factors in citizens' utilization of OGD portals, which answered the fourth sub-question (RQ4). A heuristic evaluation of the usability of Chinese province-level OGD portals has been carried out. Results of the heuristic evaluation showed OGD portals' ineffective performance in usability guidelines specifically for OGD portals. Error prevention, user interaction, tagging datasets, providing filtering and suggestion features, offering different kinds of open format data and dataset visualization are aspects that need improvement in future usability design. In addition, OGD portals' inadequacy in providing sufficient help functions may be a cause of their low utilization.

Based on the results of the usability evaluation as well as the results of the comparison analysis of the supply-side and demand-side of OGD, an experiment on citizens' utilization of OGD portals was carried out. The analysis of the experiment data showed good performance in completing tasks when using an OGD portal with good usability. Among the five variables of DOI which could explain citizens' acceptance of OGD portals, compatibility and trialability are of great importance because of their effects on the other variables. Regarding citizens' effective use of OGD portals, their rapport with the help functions together with the relative advantage and trialability of DOI play decisive roles in predicting citizens' utilization. Rather than traditional methods of using pictures and texts on web pages, offering help through virtual assistants could improve citizens' ability to use OGD portals effectively to find the data they need.

In conclusion, the supply-side and demand-side of OGD have significant conflicts in relation to observability, relative advantage, compatibility and complexity, which are revealed by the acceptance and utilization of the bridge between these two sides: the OGD portal. In order to properly connect and balance the two sides of OGD, which is the focus of the fifth sub-question (RQ5), we suggest improving the observability and trialability of OGD portals, strengthening the relative advantage of OGD, improving the compatibility of OGD portals and reducing the complexity of using the portals. To answer the sixth sub-question (RQ6), five directions are proposed for the future development of OGD portals, which are identifying different user types and offering data and services accordingly, increasing the online visibility of OGD portals, improving help functions, strengthening user interaction and reinforcing OGD characteristics.

9.2 Contributions

The main aim of this thesis has been to find possible directions for balancing the supply-side and demand-side of OGD through OGD portals to encourage OGD utilization. This aim was sought because the ultimate goal of disclosing government data is for its utilization (Attard et al., 2015) and at present, the usage of the opened government data on the portals is limited (Safarov et al., 2017) and thus shows its insignificant impact on society and the economy (OpenDataBarometer, 2018a). Balancing the supply-side and demand-side of OGD through OGD portals could therefore encourage citizens to accept and use the data on the portals and consequently benefit the whole of society with transparency and accountability of governments and public engagement in administration as well as economic effects (Florini, 2008; Willinsky, 2005).

In this section, we review the main contributions from two aspects: the contributions to knowledge theory and the practical insights regarding the development of OGD portals.

9.2.1 Theoretical contributions

There are several different theoretical contributions of this thesis:

Firstly, for understanding the supply-side of OGD, a local-level OGD portal assessment framework was proposed, based on a comparison of related studies and principles. This assessment framework fills the present research gap by deriving priorities from AHP, showing experts' perspectives on the importance of different OGD characteristics. Also, the process of deriving priorities by AHP indicated the capability of this process in the area of OGD. The building of the prioritized framework together with the verification of AHP in OGD studies contributed to the present literature through the publication of (D. Wang, Chen, et al., 2018).

Secondly, for understanding the demand-side of OGD, the relationships among citizens as OGD users, their OGD demands and OGD utilization are revealed in this thesis. This analysis further confirms other researchers' conclusions about the effect of users' demographic background on their OGD demands and utilization aims (Graves-Fuenzalida, 2013). Moreover, it reveals the relationship between citizens' OGD utilization purposes and behaviour with their OGD demands for different subjects and attributes. The testing of the hypothesis about correlations between citizens, their OGD demands and utilizations not only filled the present research gap about connecting users' demands and utilization habits but also help forming the model for future studies about OGD users' behaviour. Related results contributed to the present literature through the publication of (D. Wang et al., 2019) and (D. Wang, Richards, et al., 2018).

Thirdly, for analyzing the relationship between the supply-side and demand-side of OGD, a model has been developed based on the theory of DOI by including elements from related literature to guide the comparison of the supply-side and demand-side of OGD. Our comparative case study demonstrates the capacity of the model for analyzing the matches and conflicts of the two sides, which could help in adjusting the development strategy for the supply-side. The contribution to the present literature not only lies in the model itself for recognizing the relationships between supply-side and demand-side of OGD, but also the process in the case study for applying the proposed model. The separate data collection process on the supply-side and demand-side as well as the data pre-processing through normalization guaranteed a direct quantitative comparison of the supply-side and demand-side of OGD.

Fourthly, a usability evaluation framework specifically for OGD portals has been built by extending usability guidelines and developing criteria for each guideline. This framework not only extends Nielsen's 10 usability guidelines (Nielsen, 1994b) by adding principles fitting the specific needs of OGD but also develops operational criteria for judging the performance of OGD portals in usability. The adding of lower level criteria for the usability evaluation helps to reduce reliance of the heuristic evaluation on the level of experience of evaluators and fills a current weakness of heuristic evaluations by making explicit what problems or features of a website need testing (Hasan et al., 2013).

Fifthly, the relationships between variables including TAM, DOI, trustworthiness and rapport are discussed in this study. Although related studies focusing on users' intention to use e-government services have analyzed possible relationships among TAM, DOI and trustworthiness (Rana et al., 2015), specific discussions in the field of OGD are limited. Our separate models for each of the five DOI variables not only reveal new relationships between TAM and DOI, but also emphasize the importance of help functions on a portal and the role of rapport building between the portal and users. It is shown how these factors affect users' acceptance and utilization of the OGD portals.

Lastly, a model for predicting citizens' effective utilization of OGD portals was built based on machine-learning with the experimental data. Previous research paid more attention to the factors affecting users' intention to use (Weerakkody et al., 2017), but our model focuses on the actual utilization of citizens instead of their intention to accept and use OGD portals. The model for predicting citizens' possibility of using the OGD portal also indicated the importance for combining the theory of DOI, rapport and TAM in predicting users' acceptance behavior, which to some extent, contributed to the present literature for synthesizing technology acceptance models.

9.2.2 Pragmatic implications

There are three main pragmatic implications of this thesis: the supply-side of OGD, the demand-side of OGD and the OGD portals.

For the supply-side of OGD, our evaluation of province-level OGD portals in China has shown the capability of the proposed framework for analyzing the present development of local OGD portals and provided an objective view of the latest development of OGD portals in China. By taking different snapshots of OGD portals, it is also possible to keep track of an evaluation over time with the help of our evaluation framework. It is also possible to adjust the priority of elements in the evaluation framework to fit the local needs of different areas by repeating the AHP process with local experts.

For the demand-side of OGD, our analysis reveals Chinese citizens' perspectives on OGD, including their needs for certain data content, requirements for data quality and demands for help functions. The relationships between citizens' characteristics, their OGD demands and their OGD utilization could help future studies to characterize different user types for OGD portals so as to provide services accordingly.

The contributions of the thesis to the development of OGD portals are mainly related to the last two sub-questions (RQ5 & RQ6) regarding balancing the supply-side and demand-side of OGD, as well as future development directions for OGD portals. Based on the data analysis from the whole thesis, we proposed four possible ways to balance the supply-side and demand-side of OGD, which are improving observability and trialability, strengthening the relative advantage of OGD, improving compatibility and reducing complexity. Discussion based on these four aspects led to five more specific suggestions for the future development of OGD portals, including identifying user types, increasing online visibility, improving help functions, strengthening user interaction and reinforcing OGD characteristics. Although these directions may not solve all the existing problems with OGD portals in usability and meeting the needs of the demand-side, they shed light on solving the most obvious ones.

9.3 Importance of findings

The importance of the findings of this thesis for the future development and the utilization of OGD is fourfold:

Firstly, the OGD ecosystem requires the supply-side to keep a balance with the demand-side (Tim Davies, 2011). Also, certain mechanisms are needed in the OGD ecosystem for the supply-side and demand-side to communicate with each other to keep the balance. Our findings in Chapter 5 provided an overview of the present relationship between the supply-side and demand-side of OGD. Moreover, the model built in Chapter 5 makes it possible for the two sides to communicate with each other based on the same “language”. This mechanism would help future development of OGD by keeping the balance of the two sides.

Secondly, the governments on the supply-side call for improved utilization of the opened data online (Ohemeng & Ofosu-Adarkwa, 2015) as well as the promotion of citizen engagement (Kassen, 2013). Facing the insufficiency of the present OGD utilization, our findings provide various approaches for OGD development and for reducing the conflicts between two sides of OGD in order to improve OGD utilization and citizens’ acceptance of OGD, as discussed separately in Chapter 8.

Thirdly, the development of OGD and OGD portals is still in its infancy (Attard et al., 2015; Zuiderwijk & Janssen, 2014). On one hand, lots of local OGD portals emerged during this fast development period. While on the other hand, local governments are facing serious tasks regarding the building of data resources in the context of limited budgets, IT infrastructure and skills (Ebrahim & Irani, 2005). The findings of this thesis in Chapter 3 offered possible priority choices for local governments when arranging the limited resources. While findings in Chapter 4, 6 and 7 will help the developers to recognize the needs of OGD users as well as tips for the design of OGD portals and its functions.

Finally, studies have shown an insignificant impact of OGD development on the whole society (OpenDataBarometer, 2018b). In order to increase the positive social and economic impact of OGD (Ribeiro, 2017) for citizens, the acceptance of OGD and OGD portals is the premise for further utilization. Our findings in Chapter 7 provided a systematic review of the constructs for citizens' decision of the acceptance of OGD as well as a model for predicting citizens' acceptance of OGD portals. Thus, the findings will enlighten the future development of OGD by increasing citizens' acceptance of OGD and OGD portals.

9.4 Limitations and future work

Besides the above contributions, our study also has several limitations.

Firstly, our study focuses on a specific area in the world, which is the largest developing country, China. For the general context of politics and economy, distinctions exist between China and other "Western" countries as well as between other well-developed countries. The supply-side analysis is based on province-level OGD portals, while the demand-side analysis includes only a sample of the Chinese population. Due to the previous stages of the study being all about China, the usability evaluation and the experiment have also been carried out with similar samples to maintain the consistency of the study. Thus, the results of this study may not be generalizable to other populations in other countries of totally different political and economic background. However, as we explain in Section 1.4, one of the reasons for carrying out this study in China is due to its various population sizes among provinces, economic contexts and political systems among its sub-areas, which may improve the applicability of the results in other areas around the world. Besides, it is difficult to include a great number of countries in one study and comparing the data from disparate geographical regions and contexts is problematic. In order to confirm the conclusions of this study as well as to perform comparative analysis, we plan to carry out similar studies in one or two countries of different social contexts, such as Australia.

Secondly, our study only focuses on one type of users of OGD portals, that is, the citizen, because they are recognized as the primary stakeholders and major beneficiaries of OGD utilization (Parycek et al., 2014). It is true that choosing citizens as the representative of the demand-side covers the largest range of users, but this has led to neglect of the differences among specific types of OGD users (Safarov et al., 2017), including researchers, journalists and developers. Therefore, the conclusions of this study, especially about the demand-side of OGD, are not particularized for different types of OGD users. Our study also did not take OGD users in the public sector (Graves-Fuenzalida, 2013) into account when analyzing the demand-side. In future studies, we will carry out analysis of the characteristics, OGD needs and usage habits of different types of OGD users, since identifying the user types is concluded by the study to be an important direction for future development of OGD portals.

Thirdly, the analysis of this study is based on a limited number of samples and limited amount of collected data. For the supply-side, the randomly selected 2636 datasets could not cover all the data on the province-level OGD portals in China. For the demand-side, the sample size of the survey could not reflect the attitudes of all Chinese citizens. Self-selection bias may also exist in the recruited sample due to online recruitment, because distributing on social medias may attract participants who are more likely to complete the online survey than others (Wright, 2005). The experiment was also carried out on limited numbers of participants. The limited number of samples may affect the validity of the research conclusions. Nevertheless, we have tried to include as many samples as possible in the given research conditions. The reliability and adequacy tests also show the validity of the collected samples. For future studies, we will carry out other evaluations on the supply-side with larger numbers of datasets to compare with the results of this study. For the demand-side, future research will be aimed at more focused samples of different types of OGD users including journalists, developers and scientists.

Fourthly, the usability evaluation was only carried out by 4 recruited experts who may not have the range of expertise required. But it is impossible to discover all usability design problems purely based of the experiences of 4 experts. However, this number is within the range recommended by Nielson and deemed to be adequate and supported by satisfactory iterator reliability scores. The evaluation also depended on the interpretations, experiences and testing of evaluators on the targeted portals, which may lead to disagreements of evaluation results among different evaluators. The heuristic evaluation also couldn’t provide a systematic way to fix the usability problems. However, in our study, we tried to cover this limitation by adding specific criteria to the principles so that the criteria could provide suggestions for fixing the existing problems.

Finally, this study focuses on the OGD portal, which is a bridge between the supply-side and demand-side, for its data, usability and functions. But because OGD portals are usually products of governments’ OGD initiatives (Kassen, 2013), their development is usually affected by many different factors, such as the extent to which there is political support and open data infrastructure (OpenDataBarometer, 2018a). In our study, although we considered the local context when analyzing the supply-side of OGD, it was not our key focus for understanding the supporting conditions for the development of OGD portals. Therefore, in future studies we would further analyze the political support and open data infrastructure that are needed by OGD portal development to cover the present imperfections in data quantity, data quality and usability.

9.5 Final remarks

With the help of information and communication technologies, OGD faces a great opportunity for development. Its utilization is believed to contribute to public administration and bring economic benefits. This thesis focuses on the problems in the present stage of OGD development by comparing and matching the supply-side and demand-side of OGD through portals. The comparison of the two sides reveals the mismatch of governments' supply of OGD and portal services with the demands of citizens, which results in not only their limited knowledge of such information resources but also their rejection of using the data. As a result of the issue in OGD utilization, little historical evidence of real benefits from OGD initiatives has been recognized so far (OpenDataBarometer, 2018a).

We anticipate that the theoretical and pragmatic outputs of this research will contribute not only to a more usable and dynamic design of OGD portals, but also to a more inclusive and interactive OGD ecosystem. In particular, we have provided direction for the development of OGD portals to better connect the supply of OGD with the demands of citizens. Encouraging citizens' involvement in the development and utilization of OGD would be a good way to realize the supposed benefits of OGD including transparency, anti-corruption, enhancing accountability and commercial achievements. Therefore, considering the nature of OGD and OGD portals, we believe that the findings of this thesis are in the direction of the common good and will help lead to a brighter future for OGD and OGD portals.

REFERENCES

Abbattista, F., Lops, P., Semeraro, G., Andersen, V., & Andersen, H. (2002). *Evaluating virtual agents for e-commerce.* Paper presented at the AAMAS.

Agrawal, D., Kettinger, W., & Zhang, C. (2014). The openness challenge: why some cities take it on and others don't.

Aichholzer, G., & Burkert, H. (2004). *Public sector information in the digital age: between markets, public management and citizens' rights*: Edward Elgar Publishing.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39): Springer.

Amerland, D. (2013). *Google Semantic Search: Search Engine Optimization (SEO) Techniques That Get Your Company More Traffic, Increase Brand Impact, and Amplify Your Online Presence*: Que Publishing.

André, E., Rist, T., & Müller, J. (1998). WebPersona: a lifelike presentation agent for the World-Wide Web. *Knowledge-Based Systems, 11*(1), 25-36.

Anthopoulos, L. G., Siozos, P., & Tsoukalas, I. A. (2007). Applying participatory design and collaboration in digital public services for discovering and re-designing e-Government services. *Government Information Quarterly, 24*(2), 353-376.

Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism, 2*(3), 394-405.

Arbel, A., & Orgler, Y. E. (1990). An application of the AHP to bank strategic planning: The mergers and acquisitions process. *European Journal of operational research, 48*(1), 27-37.

Armstrong, C. L. (2011). Providing a clearer view: An examination of transparency on local government websites. *Government Information Quarterly, 28*(1), 11-16.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., . . . Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal, 3*, 135-152.

Atkinson, N. L. (2007). Developing a questionnaire to measure perceived attributes of eHealth innovations. *American Journal of Health Behavior, 31*(6), 612-621.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), 399-418. doi:10.1016/j.giq.2015.07.006

Babbie, E. R. (2015). *The practice of social research*: Nelson Education.

Babic, Z., & Plazibat, N. (1998). Ranking of enterprises based on multicriterial analysis. *International Journal of Production Economics, 56*, 29-35.

Badri, M. A. (2001). A combined AHP–GP model for quality control systems. *International Journal of Production Economics, 72*(1), 27-40.

Bahorsky, R., Graber, J., & Mason, S. (1998). *Official Internet dictionary: A comprehensive reference for professionals*: Government Institutes.

Baiocchi, G., & Ganuza, E. (2014). Participatory budgeting as if emancipation mattered. *Politics & Society, 42*(1), 29-50.

Baker, D. L. (2009). Advancing e-government performance in the United States through enhanced usability benchmarks. *Government Information Quarterly, 26*(1), 82-88.

Bar-Ilan, J. (2001). Data collection methods on the Web for infometric purposes—A review and analysis. *Scientometrics, 50*(1), 7-32.

Barroso, J. M. (2011). *Commission Decision of 12 December 2011 on the reuse of Commission documents*. Brussels Retrieved from http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32011D0833.

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in human behavior, 25*(2), 450-457.

Baylor, A. L., Kim, S., Son, C., & Lee, M. (2005). *Designing effective nonverbal communication for pedagogical agents.* Paper presented at the Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology.

Baylor, A. L., Shen, E., Warren, D., & Park, S. (2004). *Supporting learners with math anxiety: The impact of pedagogical agent emotional and motivational support.* Paper presented at the Workshop on" Social and Emotional Intelligence in Learning Environments," held at the International Conference on Intelligent Tutoring Systems., Maceió, Brazil.

Behkamal, B., Kahani, M., Bagheri, E., & Jeremic, Z. (2014). A metrics-driven approach for quality assessment of linked open data. *Journal of theoretical and applied electronic commerce research, 9*(2), 64-79.

Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The journal of strategic Information Systems, 11*(3-4), 245-270.

Benjamin, C. O., Ehie, I. C., & Omurtag, Y. (1992). Planning facilities at the University of Missouri-Rolla. *Interfaces, 22*(4), 95-105.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*(3), 264-271. doi:https://doi.org/10.1016/j.giq.2010.03.001

Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems* (pp. 23-54): Springer.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American society for information science and technology, 55*(14), 1216-1227.

Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. (2014). How open are public government data? An assessment of seven open data portals. In *Measuring E-government Efficiency* (pp. 25-44): Springer.

Bollettino, J. O. (2002). The customer-centric digital department: e-Service in government. *2002): e-Service: New Directions in Theory and Practice, New Delhi, Prentice-Hall of India*.

Bonsón, E., Torres, L., Royo, S., & Flores, F. (2012). Local e-government 2.0: Social media and corporate transparency in municipalities. *Government Information Quarterly, 29*(2), 123-132. doi:10.1016/j.giq.2011.10.001

Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The state of open data. *Limits of current open data platforms*.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement, 41*(3), 687-699.

Bryson, S. (1996). Virtual reality in scientific visualization. *Communications of the ACM, 39*(5), 62-71.

Budd, A. (2007). Heuristics for Modern Web Application Development. Retrieved from http://www.andybudd.com/archives/2007/01/heuristics_for_modern_web_application_development/

Burawoy, M., Burton, A., Ferguson, A. A., & Fox, K. J. (1991). *Ethnography unbound: Power and resistance in the modern metropolis*: Univ of California Press.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in human behavior, 16*(6), 553-574.

Carter, L., & Belanger, F. (2004). *Citizen adoption of electronic government initiatives.* Paper presented at the 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the.

Carter, L., & Bélanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors. *Information systems journal, 15*(1), 5-25.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in human behavior, 29*(6), 2156-2160.

Cassell, J. (2000). *Embodied conversational agents*: MIT press.

Cassell, J., Gill, A. J., & Tepper, P. A. (2007). *Coordination in conversation and rapport.* Paper presented at the Proceedings of the workshop on Embodied Language Processing.

Chakraborty, A., Wilson, B., Sarraf, S., & Jana, A. (2015). Open data for informal settlements: Toward a user′ s guide for urban managers and planners. *Journal of Urban Management, 4*(2), 74-91.

Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly, 34*(2), 231-243. doi:10.1016/j.giq.2017.02.004

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & In Shim, S. (2011). Virtual agents in e-commerce: representational characteristics for seniors. *Journal of Research in Interactive Marketing, 5*(4), 276-297.

Chattaraman, V., Kwon, W., & Gilbert, J. (2009). *Social presence in online stores: building social support and trust among older consumers.* Paper presented at the Retailing Strategic Challenges and Opportunities in Uncertain Times: Special Conference Series.

Chou, J.-R., & Hsiao, S.-W. (2007). A usability study on human–computer interface for middle-aged learners. *Computers in human behavior, 23*(4), 2040-2063.

Chu, H., He, S., & Thelwall, M. (2002). Library and information science schools in Canada and USA: A Webometric perspective. *Journal of education for library and information science*, 110-125.

Clemmensen, T., & Katre, D. (2012). Adapting e-gov Usability: Evaluation to Cultural Contexts. In *Usability in government systems* (pp. 331-344): Elsevier.

CNNIC. (2019). *The 44th China Statistical Report on Internet Development*. Retrieved from http://www.cac.gov.cn/pdf/20190829/44.pdf

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

Cornford, J., Wilson, R., Baines, S., & Richardson, R. (2013). Local governance in the new information ecology: the challenge of building interpretative communities. *Public Money*

& Management, 33(3), 201-208.

Corrêa, A. S., Paula, E. C. d., Corrêa, P. L. P., & Silva, F. S. C. d. (2017). Transparency and open government data. *Transforming Government: People, Process and Policy, 11*(1), 58-78. doi:10.1108/tg-12-2015-0052

Cranefield, J., Robertson, O., & Oliver, G. (2014). Value in the mash: Exploring the benefits, barriers and enablers of open data apps.

Craveiro, G. d. S., Santana, M. T. d., & Pereira, J. P. d. A. (2013). *Assessing open government budgetary data in Brazil.* Paper presented at the International Conference on Digital Society-ICDS 2013.

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

David, P. A., & Foray, D. (2003). Economic fundamentals of the knowledge society. *Policy futures in education, 1*(1), 20-49.

Davies, T. (2011). Open Data: infrastructures and ecosystems. *Open Data Research*, 1-6.

Davies, T. (2014). Open data policies and practice: An international comparison. *Available at SSRN 2492520*.

Davies, T. G. (2013). Transparency and open data.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results.* Massachusetts Institute of Technology,

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of applied social psychology, 22*(14), 1111-1132.

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

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International journal of human-computer studies, 52*(1), 1-22.

Delopoulos, H. N. (2015). A usability evaluation of e-government services: the case of e-deliberation service of Greece. *International Journal of Electronic Governance, 7*(2), 93-112.

Desouza, K. C., & Bhagwatwar, A. (2012). Citizen apps to solve complex urban problems. *Journal of Urban Technology, 19*(3), 107-136.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26): Sage publications.

Dietrich, D. (2015). The Role of Civic Tech Communities in PSI Reuse and Open Data Policies. *European PSI Platform*.

Dix, A. (2009). Human-computer interaction. In *Encyclopedia of database systems* (pp. 1327-1331): Springer.

DMGLabFudanUniversity. (2018). China Open Data Index. Retrieved from http://ifopendata.fudan.edu.cn/method

Dragone, M., Holz, T., & O'Hare, G. M. (2007). *Using mixed reality agents as social interfaces for robots.* Paper presented at the RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication.

Dwivedi, Y. K., & Williams, M. D. (2008). Demographic influence on UK citizens'e-government adoption. *Electronic government, an international journal, 5*(3), 261-274.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological bulletin, 81*(6), 358.

Ebrahim, Z., & Irani, Z. (2005). E-government adoption: architecture and barriers. *Business process management journal, 11*(5), 589-611.

Edelmann, N., Höchtl, J., & Sachs, M. (2012). Collaboration for open innovation processes in public administrations. In *Empowering open and collaborative governance* (pp. 21-37): Springer.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review, 14*(4), 532-550.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management journal, 50*(1), 25-32.

Ellonen, H.-K., Wikström, P., & Jantunen, A. (2009). Linking dynamic-capability portfolios and innovation outcomes. *Technovation, 29*(11), 753-762.

Eschenbrenner, B., Nah, F. F.-H., & Siau, K. L. (2009). 3-D Virtual Worlds in Education. In.

Etzioni, A. (1960). Two approaches to organizational analysis: A critique and a suggestion. *Administrative science quarterly*, 257-278.

Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management, 32*(1), 172-185.

Fajar Marta, R. (2016). Open Government Data Portal Design Principles: Implementing Transparency, Privacy, and Information Quality Assurance by Design.

Feeney, M. K., & Brown, A. (2017). Are small cities online? Content, ranking, and variation of U.S. municipal websites. *Government Information Quarterly, 34*(1), 62-74. doi:10.1016/j.giq.2016.10.005

Fernandez, A., Insfran, E., & Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study. *Information and software Technology, 53*(8), 789-817.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin, 76*(5), 378.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*: John Wiley & Sons.

Florini, A. (2008). Making transparency work. *Global Environmental Politics, 8*(2), 14-16.

Foulonneau, M., Martin, S., & Turki, S. (2014). *How open data are turned into services?* Paper presented at the International Conference on Exploring services science.

Fuentes-Enriquez, R., & Rojas-Romero, Y. (2013). *Developing accountability, transparency and government efficiency through mobile apps: the case of Mexico.* Paper presented at the proceedings of the 7th international conference on theory and practice of electronic governance.

Gao, Y., & Vaughan, L. (2005). *Web hyperlink profiles of news sites: A comparison of newspapers of USA, Canada, and China.* Paper presented at the Aslib proceedings.

Garcia, A. C. B., Maciel, C., & Pinto, F. B. (2005). *A quality inspection method to evaluate e-government sites.* Paper presented at the International Conference on Electronic Government.

Gauld, R., Goldfinch, S., & Horsburgh, S. (2010). Do they want it? Do they use it? The 'Demand-Side'of e-Government in Australia and New Zealand. *Government Information Quarterly, 27*(2), 177-186.

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorological Organization (WMO) Bulletin, 58*(3), 175.

Gitelman, L. (2013). *Raw data is an oxymoron*: MIT press.

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:10.1016/j.giq.2015.09.001

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality, 37*(6), 504-528.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., & Morency, L.-P. (2006). *Virtual rapport.* Paper presented at the International Workshop on Intelligent Virtual Agents.

Graves-Fuenzalida, A. (2013). *Improving the use of open government data using visualizations.* Rensselaer Polytechnic Institute,

Graves, A., & Hendler, J. (2013). *Visualization tools for open government data.* Paper presented at the Proceedings of the 14th Annual International Conference on Digital Government Research.

Grayling, T. (1998). Fear and loathing of the help menu: A usability test of online help. *Technical communication, 45*(2), 168-179.

Grayling, T. (2002). If we build it, will they come? A usability test of two browser-based embedded help systems. *Technical communication, 49*(2), 193-209.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological bulletin, 83*(2), 314.

Grice, G. R. (1966). Dependence of empirical laws upon the source of experimental variation. *Psychological bulletin, 66*(6), 488.

Grudin, J. (1992). Utility and usability: research issues and development contexts. *Interacting with computers, 4*(2), 209-217.

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday, 16*(2).

Halonen, A. (2012). Being open about data: Analysis of the UK open data policies and applicability of open data. *London: The Finnish Institute in London*.

Harrison, D. A., Mykytyn Jr, P. P., & Riemenschneider, C. K. (1997). Executive decisions about adoption of information technology in small business: Theory and empirical tests. *Information systems research, 8*(2), 171-195.

Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet, 4*(4), 900-928.

Harvey, M., McMeekin, A., Randles, S., Southerton, D., Tether, B., & Warde, A. (2001). Between demand and consumption: A framework for research. *Centre for Research on Innovation and Competition Discussion Paper, 40*.

Hasan, L., Morris, A., & Probets, S. (2013). E-commerce websites for developing countries–a usability evaluation framework. *Online Information Review, 37*(2), 231-251.

Hawker, S., & Hawkins, J. M. (2001). *The Oxford popular dictionary & thesaurus: 2 in 1*: Parragon.

Heimstädt, M., Saunderson, F., & Heath, T. (2014). From toddler to teen: growth of an open data ecosystem. *JeDEM-eJournal of eDemocracy and Open Government, 6*(2), 123-135.

Heise, A., & Naumann, F. (2012). Integrating open government data with stratosphere for more transparency. *Web Semantics: Science, Services and Agents on the World Wide Web, 14*, 45-56.

Ho, W. (2008). Integrated analytic hierarchy process and its applications–A literature review. *European Journal of operational research, 186*(1), 211-228.

Hoorn, J., Eliëns, A., Huang, Z., Van Vugt, H. C., Konijn, E. A., & Visser, C. T. (2004). Agents with character: Evaluation of empathic agents in digital dossiers. *Emphatic Agents, AAMAS*.

Hootsuite. (2019). *Global Digital 2019 reports*. Retrieved from https://wearesocial.com/global-digital-report-2019

Horrigan, J. B., Rainie, L., & Page, D. (2015). Americans' views on open government data. *Pew Research Center*.

Horsley, J. P. (2007). China adopts first nationwide open government information regulations. *Freedominfo. org (Online), 9*.

Huang, R., & Wang, C. (2016). An Investigation into the Open Government Data Platform in China. *Information Studies: Theory & Application, 39*(7), 50-55.

Huang, Z., & Benyoucef, M. (2014). Usability and credibility of e-government websites. *Government Information Quarterly, 31*(4), 584-595.

Hui, K.-L., & Png, I. P. (2015). Research note—Migration of Service to the Internet: Evidence from a Federal Natural Experiment. *Information systems research, 26*(3), 606-618.

Iemma, R. (2012). Open Government Data: A Focus on Key Economic and Organizational Drivers. *Available at SSRN 2262943*.

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of documentation, 54*(2), 236-243.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International journal of human-computer studies, 53*(2), 251-267.

Iwarsson, S., & Ståhl, A. (2003). Accessibility, usability and universal design—positioning and definition of concepts describing person-environment relationships. *Disability and rehabilitation, 25*(2), 57-66.

Jaeger, P. (2005). Deliberative democracy and the conceptual foundations of electronic government. *Government Information Quarterly, 22*(4), 702-719.

Jaeger, P., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly, 27*(4), 371-376. doi:10.1016/j.giq.2010.05.003

Jaeger, P., & Burnett, G. (2005). Information access and exchange among small worlds in a democratic society: the role of policy in shaping information behavior in the post-9/11 United States. *The Library Quarterly, 75*(4), 464-495.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly, 28*(4), 446-456.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management, 29*(4), 258-268.

Jetzek, T., & Avital, M. (2013). The generative mechanisms of open government data.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2012). *The value of open government data: A strategic analysis framework.* Paper presented at the SIG eGovernment pre-ICIS

Workshop, Orlando.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of theoretical and applied electronic commerce research, 9*(2), 100-120.

Kaiser, H. F. (1970). A second generation little jiffy. *psychometrika, 35*(4), 401-415.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). *Big data: Issues and challenges moving forward.* Paper presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Linked Open Government Data Analytics. In M. A. Wimmer, M. Janssen, & H. J. Scholl (Eds.), *Electronic Government: 12th IFIP WG 8.5 International Conference, EGOV 2013, Koblenz, Germany, September 16-19, 2013. Proceedings* (pp. 99-110). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kapoor, K., Dwivedi, Y. K., & Williams, M. D. (2013). *Role of innovation attributes in explaining the adoption intention for the interbank mobile payment service in an Indian context.* Paper presented at the International Working Conference on Transfer and Diffusion of IT.

Karthikeyan, R., Venkatesan, K., & Chandrasekar, A. (2016). A comparison of strengths and weaknesses for analytical hierarchy process. *Journal of Chemical and Pharmaceutical Sciences, 9*(3), 12-15.

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly, 30*(4), 508-513. doi:10.1016/j.giq.2013.05.012

Keren, G. B., & Raaijmakers, J. G. (1988). On between-subjects versus within-subjects comparisons in testing utility theory. *Organizational Behavior and Human Decision Processes, 41*(2), 233-247.

Khajouei, R., Hasman, A., & Jaspers, M. W. (2011). Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system. *international journal of medical informatics, 80*(5), 341-350.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & management, 43*(6), 740-755.

Kitchin, R. (2013). Four critiques of open data initiatives. *LSE: The impact blog, 27*.

Kostovski, M., Jovanovik, M., & Trajanov, D. (2012). *Open data portal based on semantic web technologies.* Paper presented at the Proceedings of the 7th South East European Doctoral Student Conference, University of Sheffield, Greece.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association, 47*(260), 583-621.

Kube, M., Hilgers, D., Fueller, J., & Koch, G. (2013). Re-Thinking Citizen Participation in the Public Sector: Adopting Open Innovation for Collaborative Idea Generation. *Available at SSRN 2336673*.

Kučera, J., Chlapek, D., & Nečaský, M. (2013). *Open government data catalogs: Current approaches and quality perspective.* Paper presented at the International Conference on Electronic Government and the Information Systems Perspective.

Kuflik, T., & Shoval, P. (2000). *Generation of user profiles for information filtering—research agenda.* Paper presented at the Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval.

Kuhn, K. (2011). Open government data and public transportation. *Journal of Public*

Transportation, 14(1), 5.

Kuo, B. C., Roldan-Bau, A., & Lowinger, R. (2015). Psychological help-seeking among Latin American immigrants in Canada: Testing a culturally-expanded model of the Theory of Reasoned Action using path analysis. *International Journal for the Advancement of Counselling, 37*(2), 179-197.

Lee, C., & Kwak, N. (1999). Information resource planning for a health-care system using an AHP-based goal programming method. *Journal of the operational research society, 50*(12), 1191-1198.

Lee, J., & Rao, H. R. (2009). Task complexity and different decision criteria for online service acceptance: A comparison of two e-government compliance service domains. *Decision Support Systems, 47*(4), 424-435.

Lee, Y., & Kozar, K. A. (2012). Understanding of website usability: Specifying and measuring constructs and their relationships. *Decision Support Systems, 52*(2), 450-463.

Leontieva, L. S., Khalilova, T. V., Gaynullina, L. F., & Khalilov, A. I. (2015). Social-communicative innovations in anti-corruption activities (Regional aspect). *Asian Social Science, 11*(7), 387.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). *The persona effect: affective impact of animated pedagogical agents.* Paper presented at the CHI.

Li, J., Kizilcec, R., Bailenson, J., & Ju, W. (2016). Social robots and virtual agents as lecturers for video instruction. *Computers in human behavior, 55*, 1222-1230.

Li, R. (2018). *Research on the construction content of Chinese government data opening platform.* (Master's Degree), Jilin University,

Lim, E. T., Tan, C.-W., Cyr, D., Pan, S. L., & Xiao, B. (2012). Advancing public trust relationships in electronic government: The Singapore e-filing journey. *Information systems research, 23*(4), 1110-1130.

Lindman, J., Rossi, M., & Tuunainen, V. K. (2013). *Open data services: Research agenda.* Paper presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on.

Lonsdale, C., Hodge, K., & Rose, E. A. (2006). Pixels vs. paper: Comparing online and traditional survey methods in sport psychology. *Journal of Sport and Exercise Psychology, 28*(1), 100-108.

Lourenço, R. P. (2013). *Open government portals assessment: a transparency for accountability perspective.* Paper presented at the International Conference on Electronic Government.

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

Lourenço, R. P., Sá, P., Jorge, S., & Pattaro, A. F. (2013). Online transparency for accountability: one assessing model and two applications. *Electronic Journal of e-Government, 11*(2), 280-292.

Lundvall, B.-ä., & Johnson, B. (1994). The learning economy. *Journal of industry studies, 1*(2), 23-42.

Magalhaes, G., Roseira, C., & Manley, L. (2014). *Business models for open government data.* Paper presented at the Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance.

Martin, C. (2014). Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective. *Policy & Internet, 6*(3), 217-240.

Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Risk analysis to overcome barriers to open data. *Electronic Journal of e-Government, 11*(1), 348.

Marx, K. (1844). Economic and philosophical manuscripts. *Early writings, 333*.

Mathieson, K. (1991). Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. *Information systems research, 2*(3), 173-191.

McGoldrick, P. J., Keeling, K. A., & Beatty, S. F. (2008). A typology of roles for avatars in online retailing. *Journal of Marketing Management, 24*(3-4), 433-461. doi:10.1362/026725708x306176

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research, 13*(3), 334-359.

Meijer, A., de Hoog, J., van Twist, M., van der Steen, M., & Scherpenisse, J. (2014). Understanding the dynamics of open data: From sweeping statements to complex contextual interactions. In *Open government* (pp. 101-114): Springer.

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

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33*(3), 338-348.

Moon, M. J., & Norris, D. F. (2005). Does managerial orientation matter? The adoption of reinventing government and e-government at the municipal level. *Information systems journal, 15*(1), 43-60.

Murillo, M. J. (2015). Evaluating the role of online data availability: The case of economic and institutional transparency in sixteen Latin American nations. *International Political Science Review, 36*(1), 42-59.

Neil, T. (2009). 6 Tips for a Great Flex UX: Part 5. Retrieved from http://designingwebinterfaces.com/6-tips-for-a-great-flex-ux-part-5

Nielsen, J. (1992). *Finding usability problems through heuristic evaluation.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Nielsen, J. (1994a). *Enhancing the explanatory power of usability heuristics.* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Nielsen, J. (1994b). *Usability engineering*: Elsevier.

Nielsen, J. (1994c). *Usability inspection methods.* Paper presented at the Conference companion on Human factors in computing systems.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Obama, B. (2009). Open Government Initiative. Retrieved from https://obamawhitehouse.archives.gov/open

OECD. Open Government Data - OECD. Retrieved from http://www.oecd.org/gov/digital-government/open-government-data.htm

OECD. (2008). *Recommendation of the Council for enhanced access and more effective use of Public Sector Information*. ([C(2008)36]).

OECD. (2015). Government at a Glance 2015.

Ohemeng, F. L., & Ofosu-Adarkwa, K. (2015). One way traffic: The open data initiative project and the need for an effective demand side initiative in Ghana. *Government Information*

Quarterly, 32(4), 419-428.
OpenDataBarometer. (2016). ODB Methodology - v1.0. Retrieved from http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-Methodology.pdf
OpenDataBarometer. (2017a). *Open Data Barometer 4th Edition — Global Report*. Retrieved from http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf
OpenDataBarometer. (2017b). Open Data Barometer Global Report Fourth Edition. Retrieved from http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf
OpenDataBarometer. (2018a). *Leaders Edition: From promise to progress*. Retrieved from https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf
OpenDataBarometer. (2018b). *Open Data Barometer Leaders Edition: From Promise to Progress*. Retrieved from https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf
OpenDataCensus. (2016). Local sensus. Retrieved from http://cn-city.census.okfn.org/
OpenDataIndex. (2016). Place Overview - Global Open Data Index. Retrieved from https://index.okfn.org/place/
OpenDataIndex. (2017). Methodology - Global Open Data Index.
OpenDataMonitor. (2017). OpenDataMonitor | Methodology. Retrieved from https://www.opendatamonitor.eu/frontend/web/index.php?r=site%2Fmethodology
OpenGovernmentWorkingGroup. (2007). Open Government Data Principles. Retrieved from https://public.resource.org/8_principles.html
OpenKnowledgeFoundation. The Open Definition. Retrieved from http://opendefinition.org
Ossowska, K., Szewc, L., Weichbroth, P., Garnik, I., & Sikorski, M. (2016). *Exploring an ontological approach for user requirements elicitation in the design of online virtual agents.* Paper presented at the EuroSymposium on Systems Analysis and Design.
Pannell, D. J., Marshall, G. R., Barr, N., Curtis, A., Vanclay, F., & Wilkinson, R. (2006). Understanding and promoting adoption of conservation practices by rural landholders. *Australian journal of experimental agriculture, 46*(11), 1407-1424.
Parycek, P., Hochtl, J., & Ginner, M. (2014). Open government data implementation evaluation. *Journal of theoretical and applied electronic commerce research, 9*(2), 80-99.
Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology, 62*(11), 2085-2094.
Petrie, H., & Bevan, N. (2009). The Evaluation of Accessibility, Usability, and User Experience. *The universal access handbook, 1*, 1-16.
Petychakis, M., Vasileiou, O., Georgis, C., Mouzakitis, S., & Psarras, J. (2014). A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *Journal of theoretical and applied electronic commerce research, 9*(2), 34-47.
Piotrowski, S. J., Zhang, Y., Lin, W., & Yu, W. (2009). Key issues for implementation of Chinese open government information regulations. *Public administration review, 69*(s1).
Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211-218.
Power, R., Robinson, B., Rudd, L., & Reeson, A. (2015). *Scenario Planning Case Studies Using Open Government Data.* Paper presented at the ISESS.
PublicSectorTransparencyBoard. (2012). *Public Data Principles*. Retrieved from

https://data.gov.uk/library/public-data-principles.

Putzer, G. J., & Park, Y. (2010). The effects of innovation factors on smartphone adoption among nurses in community hospitals. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association, 7*(Winter).

Qiu, J. (2010). Webometrics. In: Beijing: China Science Publishing.

Rajshree, N., & Srivastava, B. (2012). *Open government data for tackling corruption-a perspective.* Paper presented at the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Rana, N. P., Dwivedi, Y. K., & Williams, M. D. (2015). A meta-analysis of existing research on citizen adoption of e-government. *Information Systems Frontiers, 17*(3), 547-563.

Ranjbartabar, H. (2016). A virtual emotional freedom practitioner to deliver physical and emotional therapy.

Rappaport, A. S., & Halevi, S. (1991). The computerless computer company. *Harvard Business Review, 69*(4), 69-80.

Recker, J. (2012). *Scientific research in information systems: a beginner's guide*: Springer Science & Business Media.

Regulations of the People's Republic of China on Open Government Information. (2016). *Chinese Law & Government, 48*(1), 51-59. doi:10.1080/00094609.2015.1048135

Ribeiro, F. G. d. M. M. (2017). Value creation and barriers in the commercial use of open government data.

Rijsdijk, S. A., & Hultink, E. J. (2003). “Honey, have you seen our hamster?” Consumer evaluations of autonomous domestic products. *Journal of Product Innovation Management, 20*(3), 204-216.

Rogers, E. M. (2010). *Diffusion of innovations*: Simon and Schuster.

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*. doi:10.1016/j.giq.2017.06.003

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly, 34*(1), 45-52. doi:10.1016/j.giq.2017.01.001

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences, 1*(1), 83-98.

Saeed, S., Malik, I. A., & Wahab, F. (2013). Usability evaluation of Pakistani security agencies websites. *International Journal of E-Politics (IJEP), 4*(3), 57-69.

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of Open Government Data: A systematic literature review of types, conditions, effects and users. *Information Polity*. doi:DOI: 10.3233/IP-160014

Sandoval-Almazan, R., & Gil-Garcia, J. R. (2012). Are government internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities. *Government Information Quarterly, 29*, S72-S81. doi:10.1016/j.giq.2011.09.004

Sang, S., Lee, J.-D., & Lee, J. (2009). E-government adoption in ASEAN: the case of Cambodia. *Internet Research, 19*(5), 517-534.

Schellong, A. (2009). EU eGovernment Benchmarking 2010+ General remarks on the future of benchmarking Digital Government in the EU. *Retrieved July, 1*, 2011.

Scott, J. K. (2005). Assessing the quality of municipal government Web sites. *State and Local Government Review, 37*(2), 151-165.

Sekine, T. T. (2004). Marxian Theory of Value, What We Might Learn from It. *Korean journal of Political Economy, 2*, 1-35.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems, 27*(3), 16-24.

Shapira, B., Shoval, P., & Hanani, U. (1997). Stereotypes in information filtering systems. *Information Processing & Management, 33*(3), 273-287.

Shen, J., Han, L., & Hu, G. (2018). Open Government Data Development Velocity Index: a Research Based on Assessment of Provincial Open Government Data Portals in China. *Journal of Intelligence, 37*(11), 156-163. doi:10.3969/j.issn.1002-1965.2018.11.024

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

Smith, A., & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics, 54*(3), 363-380.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education (IJAIED), 12*, 40-62.

Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied ergonomics, 41*(3), 403-410.

Spool, J. (1997). Online help: Sometimes doesn't help. *Eye for design, 4*(3), 9-10.

StateCouncil. (2016). *Notice of the State Council on Issuing the Interim Measures for the Administration of Sharing of Government Information Resources*. Beijing Retrieved from http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

Steuer, R. E., & Na, P. (2003). Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of operational research, 150*(3), 496-515.

Stokes, K., Barker, R., & Pigott, R. (2014). *WHICH DOCTORS TAKE UP PROMISING IDEAS? New insights from open data*. Retrieved from https://www.nesta.org.uk/sites/default/files/which_doctors_take_up_promising.pdf

Strasser, S., Eveland, J., Cummins, G., Deniston, O. L., & Romani, J. H. (1981). Conceptualizing the goal and system models of organizational effectiveness—implications for comparative evaluation research. *Journal of Management Studies, 18*(3), 321-340.

Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). *Adaptive web search based on user profile constructed without any effort from users.* Paper presented at the Proceedings of the 13th international conference on World Wide Web.

SunlightFoundation. (2010). Ten Principles for Opening Up Government Information. Retrieved from https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

Susha, I., Grönlund, Å., & Janssen, M. (2015). Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries. *Information Polity, 20*(1), 19-34.

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.

Symons, V. J. (1991). A review of information systems evaluation: content, context and process.

European Journal of Information Systems, 1(3), 205-212.

Taherdoost, H. (2018). A review of technology acceptance and adoption models and theories. *Procedia manufacturing, 22*, 960-967.

Tan, M., & Teo, T. S. (2000). Factors influencing the adoption of Internet banking. *Journal of the AIS, 1*(1es), 5.

Tan, W.-s., Liu, D., & Bishu, R. (2009). Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics, 39*(4), 621-627.

Tang, R., Gregg, W., Hirsh, S., & Hall, E. (2019). US state and state capital open government data (OGD): A content examination and heuristic evaluation of data processing capabilities of OGD sites. *Proceedings of the Association for Information Science and Technology, 56*(1), 255-264.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53.

Teo, T. S., Srivastava, S. C., & Jiang, L. (2008). Trust and electronic government success: An empirical study. *Journal of management information systems, 25*(3), 99-132.

Thelwall, M. (2002). *An initial exploration of the link relationship between UK university Web sites.* Paper presented at the ASLIB proceedings.

Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly, 34*(1), 53-61. doi:10.1016/j.giq.2016.07.001

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry, 1*(4), 285-293.

Tognazzini, B. (2014). First Principles of Interaction Design (Revised & Expanded). Retrieved from https://asktog.com/atc/principles-of-interaction-design/

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*(22), 0_1.

Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of operational research, 169*(1), 1-29.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of comparative research*: Allyn & Bacon.

Van Veenstra, A. F., & Van Den Broek, T. A. (2013). *Opening moves–drivers, enablers and barriers of open data in a semi-public organization.* Paper presented at the International Conference on Electronic Government.

Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2018). Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Information systems journal*.

Veeckman, C., & van der Graaf, S. (2015). The city as living laboratory: Empowering citizens with the citadel toolkit. *Technology Innovation Management Review, 5*(3).

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly, 31*(2), 278-290. doi:10.1016/j.giq.2013.10.011

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science, 46*(2), 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Venkatesh, V., Sykes, T. A., & Venkatraman, S. (2014). Understanding e-Government portal use in rural India: role of demographic and personality characteristics. *Information systems journal, 24*(3), 249-269.

Venkatesh, V., Thong, J. Y., Chan, F. K., & Hu, P. J. (2016). Managing citizens' uncertainty in e-government services: The mediating and moderating roles of transparency and trust. *Information systems research, 27*(1), 87-111.

Verdegem, P., & Verleye, G. (2009). User-centered E-Government in practice: A comprehensive model for measuring user satisfaction. *Government Information Quarterly, 26*(3), 487-497.

Vreeland, R. C. (2000). Law libraries in hyperspace: A citation analysis of World Wide Web sites. *Law Libr. J., 92*, 9.

Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly, 35*(4), 644-656.

Wang, D., Richards, D., & Chen, C. (2018). *An Analysis of Interaction Between Users and Open Government Data Portals in Data Acquisition Process.* Paper presented at the Pacific Rim Knowledge Acquisition Workshop.

Wang, D., Richards, D., & Chen, C. (2019). *Connecting Users, Data and Utilization: A Demand-Side Analysis of Open Government Data.* Paper presented at the International Conference on Information.

Wang, H.-J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly, 33*(1), 80-88.

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.

Warde, A. (2005). Consumption and theories of practice. *Journal of consumer culture, 5*(2), 131-153.

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.

WEI-ERN, J. W. (2012). *Establishing Rapport with Conversational Agents: Comparing the Effect of Envelope and Emotional Feedback.*

Wen, F. (2019). *Research on Policy System Construction of Open Government Data in China.* (PhD), Wuhan University,

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.

Wieczorkowski, J., & Pawełoszek, I. (2018). Open Government Data, the Case of Polish Public Sector,". *The Online Journal of Applied Knowledge Management, 6*(2).

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

Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science. *First Monday, 10*(8).

Wormell, I. (2001). Informetrics and webometrics for measuring impact, visibility, and connectivity in science, politics, and business. *Competitive Intelligence Review: Published in Cooperation with the Society of Competitive Intelligence Professionals,* *12*(1), 12-23.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication, 10*(3), JCMC1034.

Wu, P. F. (2009). *User acceptance of emergency alert technology: A case study.* Paper presented at the Proceedings of the 6th International ISCRAM Conference.

Yavuz, N., & Welch, E. W. (2014). Factors affecting openness of local government websites: Examining the differences across planning, finance and police departments. *Government Information Quarterly, 31*(4), 574-583. doi:10.1016/j.giq.2014.07.004

Yi, K., & Jin, T. (2008). Hyperlink analysis of the visibility of Canadian library and information science school web sites. *Online Information Review, 32*(3), 325-347.

Yi, M. (2019). Exploring the quality of government open data: Comparison study of the UK, the USA and Korea. *The Electronic Library, 37*(1), 35-48.

Youngblood, N. E., & Mackiewicz, J. (2012). A usability analysis of municipal government website home pages in Alabama. *Government Information Quarterly, 29*(4), 582-588.

Yu, H. (2010). *On the determinants of internet-based disclosure of government financial information.* Paper presented at the Management and Service Science (MASS), 2010 International Conference on.

Yu, H., & Robinson, D. G. (2011). The new ambiguity of open government. *UCLA L. Rev. Discourse, 59*, 178.

Zainal, Z. (2007). Case study as a research method. *Jurnal Kemanusiaan, 5*(1).

Zhou, W., Xia, J., & Dai, L. (2019). Research on Construction Progress and Optimizing Strategy of Local Government Open Data Platforms in China. *Documentation, Information & Knowledge*(3), 62-71. doi:10.13366/j.dik.2019.03.062

Zuiderwijk, A., Helbig, N., Gil-Garcia, 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), 1-2.

Zuiderwijk, A., & Janssen, M. (2013). *A coordination theory perspective to improve the use of open data in policy-making.* Paper presented at the International Conference on Electronic Government.

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly, 31*(1), 17-29. doi:10.1016/j.giq.2013.04.003

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., & Parnia, A. (2013). *The complementarity of open data infrastructures: An analysis of functionalities.* Paper presented at the Proceedings of the 14th Annual International Conference on Digital Government Research.

Appendix A: Ethics Approval Letter of this thesis has been removed as it may contain sensitive/confidential content

APPENDIX B: COMPLETE SURVEY FOR STAGE 2

Survey on Public Needs of Open Government Data and Portal Utilization Habits in China

Government data is any data and information produced or commissioned by public bodies. Open Government Data (OGD) refers to the government data that can be freely used, re-used and distributed by anyone. Recent years have seen the fast development of OGD around the world, which was noticed by the Chinese government. A lot of provinces in China have built OGD portals for the public as an important access to data owned by the government.

This survey focuses on the information need of the public in China for the data owned by the government, and their habits visiting the OGD portals. The results of this survey may be used for the future development of OGD, as well as the portals in China.

School of Information Management, Wuhan University

Please choose your proper answer for the following questions:

1. Your gender:

○ Male
○ Female
○ Other

2. Your age:

○ Under 18
○ 18~25
○ 26~30
○ 31~40
○ 41~50
○ 51~60
○ Over 60

3. Your educational qualification:

○ Below junior high school

○ Junior high school
○ Senior high school
○ Under graduate
○ Post graduate
○ Beyond post graduate

4. Your occupation:
○ Student
○ Production worker
○ Marketing/Salesman
○ Customer service
○ Logistics
○ Human resources
○ Financial/auditor
○ Civilian post
○ Technician
○ Manager
○ Teacher
○ Consultant
○ Specialist (e.g., accountant, lawyer, architect, journalist, etc.)
○ Other: ____________________

5. Your industry field:
○ Agriculture, forestry, stockbreeding, fishery
○ Mining
○ Manufacturing
○ Electric power, heating power, gas, and water supply
○ Construction
○ Wholesale and retail
○ Transportation, warehousing, and postal service
○ Accommodation and catering industry
○ Information transmission, software, and information technology
○ Banking

○ Realty industry
○ Tenancy and business service
○ Scientific research and technology service
○ Water conservancy, environment and public facilities conservancy
○ Residents service, repaire and other services
○ Education
○ Health
○ Culture, sports and entertainment
○ Public administration, social security and social organization
○ International organization
○ Other: ______________

6. Which province/area do you live in right now?

○ Anhui	○ Beijing	○ Chongqing	○ Fujian	○ Gansu
○ Guangdong	○ Guangxi	○ Guizhou	○ Hainan	○ Hebei
○ Heilongjiang	○ Henan	○ Hongkong	○ Hubei	○ Hunan
○ Jiangsu	○ Jiangxi	○ Jilin	○ Liaoning	○ Macao
○ Neimenggu	○ Ningxia	○ Qinghai	○ Shandong	○ Shanghai
○ Shanxi	○ Shanxi	○ Sichuan	○ Taiwan	○ Tianjin
○ Xinjiang	○ Xizang	○ Yunnan	○ Zhejiang	○ Abroad

7. Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

I see myself as:
____ Extraverted, enthusiastic.
____ Critical, quarrelsome.

____ Dependable, self-disciplined.
____ Anxious, easily upset.
____ Open to new experiences, complex.
____ Reserved, quiet.
____ Sympathetic, warm.
____ Disorganized, careless.
____ Calm, emotionally stable.
____ Conventional, uncreative.

8. Have you ever heard of Open Government Data?

○ Yes
○ No

9. Have you ever directly used Open Government Data?

○ Yes
○ No

10. Do you know that there are portals for Open Government Data in China?

○ Yes
○ No

11. Have you ever used Open Government Data portals in China?

○ Yes
○ No

12. To what extent, do you think your need for Open Government Data?

Strongly Don' t Need ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly Need**

13. You will use Open Government Data for: (Multiple choice)

□ Daily life
□ Software development
□ Scientific research
□ News report

□ Business decision
□ Right to know and superintendence
□ Other: ______________

14. To what extent, do you need these kinds of Open Government Data?

	Strongly don't need	Don't need	A little don't need	Can't decide	A little need	Need	Strongly need
Health	○	○	○	○	○	○	○
Education	○	○	○	○	○	○	○
Entertainment	○	○	○	○	○	○	○
Public Service	○	○	○	○	○	○	○
Transportation	○	○	○	○	○	○	○
Maps	○	○	○	○	○	○	○
Public Security	○	○	○	○	○	○	○
Legislation	○	○	○	○	○	○	○
Weather	○	○	○	○	○	○	○
Air Quality	○	○	○	○	○	○	○
Registration	○	○	○	○	○	○	○
Credit Records	○	○	○	○	○	○	○
International Trade	○	○	○	○	○	○	○
Budget & Spending	○	○	○	○	○	○	○
Public Bidding	○	○	○	○	○	○	○

15. Do you need any other kind of Open Government Data?

○ Yes
○ No

If yes, what kind of Open Government Data do you need?

____

To what extent, do you need this kind of Open Government Data?

Strongly Don't Need ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly Need**

The picture above is an example of the portal for Open Government Data in China. Suppose you are now using a portal like this, and answer the following questions:

16. Will you use keyword search for the data you need?

○ Yes

○ No

17. How frequently will you use keyword search?

Strongly not frequent ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Stronly frequent**

18. To what extent do you prefer keyword search?

Strongly don' t like ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly like**

19. Will you browse for the data you need?

○ Yes

○ No

20. How frequently will you browse for the data you need?

Stronly not frequent ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Stronly frequent**

21. What kind of browse do you hope the portal could offer? (Multiple Choice)

□ Browse by subjects

□ Browse by the update time

□ Browse by the government department

□ Other:______________ *

22. How do you like browse for the data you need?

Strongly don' t like ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Stronly like**

23. Will you look at rankings on the portal?
(e.g. click ranking, download ranking, rating ranking, etc.)

○ Yes
○ No

24. Would you like the portal to recommend relating data/datasets when you are browsing?

○ Yes
○ No

25. Would you like the portal to regularly recommend data/datasets according to your visiting history?

○ Yes
○ No

26. To what extent do you need these services?

	Strongly don't need	Don't need	A little don't need	Can't decide	A little need	Need	Strongly need
Download in bulk	○	○	○	○	○	○	○
Free of charge	○	○	○	○	○	○	○
Up to date	○	○	○	○	○	○	○
Update regularly	○	○	○	○	○	○	○
Formatted data	○	○	○	○	○	○	○
Machine-readable data	○	○	○	○	○	○	○

27. How do you like *login before use*?

Strongly don' t accept ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly accept**

28. How do you like *apply before download*?

Stronly don' t accept ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly accept**

29. How difficult do you suppose when using the portal for the Open Government Data you need?

Strongly not difficult ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Stronly difficult**

30. Would you like the portal to offer user help?

○ Yes
○ No

31. To what extent, do you need help for using the portal?

Strongly don' t need ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly need**

32. What kind of help would you like the portal to offer?

□ User guide
□ FAQ
□ Online smart agent
□ Customer service phone
□ Other: ______________ *

An avatar is a computer-based humanlike virtual character. The following questions concern your preferences for a virtual character that would help and support you when using the portal.

33. To what extent, do you accept such an avatar for help?

Strongly don' t accept ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 **Strongly accept**

34. Would you prefer the virtual character to be:

○ Male
○ Female
○ Doesn' t Matter

35. Would you prefer the virtual character to be:

○ Younger than you
○ Older than you
○ Same age as you

36. Do you prefer the virtual character to look like you?

○ Yes
○ No

○ Doesn' t matter

37. What kind of role would you prefer the virtual character to be/play?

○ Teacher

○ Doctor

○ Parent

○ Friend

○ Peer

○ Other:__________________

Thank you very much for your cooperation!

APPENDIX C: DIALOGUES OF THE VIRTUAL AGENT FOR STAGE 4

The dialogues for the virtual agent were first designed in English and then translated into the Chinese version and put into the format needed to generate the dialogue. In this appendix, we first show dialogues in the English version and then show the dialogue csv files in the Chinese version.

Dialogue 1 – Introduction
Start
Avatar: Hi, my name is XXX. I'm the assistant for this open government data portal. I'll help you at any time. Are you new to this portal?
User: Yes, this is my first time to visit this portal. (1)
No, I have visited this portal before. (2)

(1)
Avatar: This is a public sector information portal that lets you find various data of this area. You can download the data and use it for commercial app development, personal analysis or academic study freely. Would you like me to introduce some of the great functions offered by our portal?
User: Please tell me more about enhanced search. (4)
Please tell me more about download queue. (5)
Please tell me more about featured datasets. (6)
Please tell me more about applications. (7)
No thanks. I would like to explore the portal by myself. (8)

(2)
Avatar: Good to see you again! Our portal offers lots of functions like enhanced search, download queue, featured datasets, and applications. If you have any questions, I'll always be there to help.
User: Thank you, see you later. (3)
Please tell me more about enhanced search. (4)
Please tell me more about download queue. (5)
Please tell me more about featured datasets. (6)
Please tell me more about applications. (7)

(3)
Avatar: See you!
User: Finish. (End)
I need your help. (Start next part)

(4)

Avatar: You can search our portal for data through any keyword. We will search the keywords both in the title and in the contents of datasets. We also provide filter functions for search results, so that you could filter by data category, data format and data provider.

User: Thank you. I would like to know about other functions. (9)

Thank you. I would like to start exploring the portal by myself. (8)

(5)

Avatar: You can add a dataset and data resource that you are interested to the queue when you browse through and download them together later.

User: Thank you. I would like to know about other functions. (9)

Thank you. I would like to start exploring the portal by myself. (8)

(6)

Avatar: On the homepage, you can find some datasets that are showcased. These are some of the instances on what is available in the dataset.

User: Thank you. I would like to know about other functions. (9)

Thank you. I would like to start exploring the portal by myself. (8)

(7)

Avatar: On the homepage, you can find applications using data from our portal. You could learn and understand how data from our portal can be used from these applications.

User: Thank you. I would like to know about other functions. (9)

Thank you. I would like to start exploring the portal by myself. (8)

(8)

Avatar: What functions would you like me to introduce?

User: Please tell me more about enhanced search. (4)

Please tell me more about download queue. (5)

Please tell me more about featured datasets. (6)

Please tell me more about applications. (7)

No thanks. I would like to explore the portal by myself. (8)

(9)

Avatar: No problem. Please enjoy your time with our portal. If you have any questions, I'll always be there to help. See you later!

User: See you. (End)

I need your help. (Start next part)

Current State	Next State	Utterance	Meaning	Style	Actions
Start	0	点击开始	N/A	N/A	
0	re_0	你好！我叫小政，是这个开放政府数据平台的助手。我会随时为你提供帮助。这是你第一次访问这个数据平台吗？	N/A	N/A	
re_0	1	是的，这是我第一次访问平台。	N/A	N/A	
re_0	2	不是，我之前就曾访问过。	N/A	N/A	
1	re_1	这是一个开放的数据门户网站。你可以在这个平台上找到各式各样的数据资源。你可以随意下载这些数据，也可以将这些数据用于商业软件开发，数据分析或者学术研究。是不是很棒呀？	N/A	N/A	
re_1	10	听上去不错。	N/A	N/A	
10	re_10	需要我给你介绍一下我们平台的各种功能吗？	N/A	N/A	
re_10	4	请介绍一下关键词搜索。	N/A	N/A	enableobject:SearchImage
re_10	5	请介绍一下数据页。	N/A	N/A	enableobject:DatapageImage
re_10	6	请介绍一下筛选和排序功能。	N/A	N/A	enableobject:FilterImage
re_10	7	请介绍一下数据展示功能。	N/A	N/A	enableobject:ShowcaseImage
re_10	8	不，谢谢。我想要自己探索平台。	N/A	N/A	
2	re_2	很高兴再次见到你！请点击下方链接，开始使用数据平台吧！这个平台提供了很多不同的功能，像关键词搜索，数据筛选和排序。如果你遇到任何问题，点击导航栏"帮助"下的"询问助手"，就能与我对话。我会随时为你提供帮助。	N/A	N/A	link:http://sjfwcn3.winzome.com
re_2	3	谢谢，等会再见。	N/A	N/A	
re_2	4	请介绍一下关键词搜索。	N/A	N/A	link disappear; enableobject:SearchImage
re_2	5	请介绍一下数据页。	N/A	N/A	link disappear; enableobject:DatapageImage
re_2	6	请介绍一下筛选和排序功能。	N/A	N/A	link disappear; enableobject:FilterImage
re_2	7	请介绍一下数据展示功能。	N/A	N/A	link disappear; enableobject:ShowcaseImage
3	re_3	好的，再见！	N/A	N/A	
re_3	finish	结束对话	N/A	N/A	
4	re_4	[illegible]行搜索。只需在这个搜索框中输入关键词，并且单击"搜索"。我们会自动在数据集的标题和元数据描述中搜索是否出现了你选择的关键词，并且会返回给你一个结果列表。现在你已经了解关键词搜索了吗？	N/A	N/A	
re_4	9	谢谢。我想了解一下其他功能。	N/A	N/A	disableobject:SearchImage
re_4	8	谢谢。我想开始自己浏览数据平台。	N/A	N/A	disableobject:SearchImage
5	re_5	如果你打开一个数据集的页面，你可以看到下方列出了很多关于这个数据集的具体信息，就像这个示例中表示的这样。现在，你已经了解数据页了吗？	N/A	N/A	
re_5	9	谢谢。我想了解一下其他功能。	N/A	N/A	disableobject:DatapageImage
re_5	8	谢谢。我想开始自己浏览数据平台。	N/A	N/A	disableobject:DatapageImage
6	re_6	我们针对数据集列表提供了筛选和排序功能。请看这个截图。你可以按照左侧的资源类型、数据领域和数据提供单位进行筛选。也可以点击右上角选择自己需要的排序方式。你知道吗？筛选和排序功能，可以极大缩短你查找到所需数据的时间。赶快试试看吧！	N/A	N/A	
re_6	9	谢谢。我想了解一下其他功能。	N/A	N/A	disableobject:FilterImage
re_6	8	谢谢。我想开始自己浏览数据平台。	N/A	N/A	disableobject:FilterImage
7	re_7	在主页上，你可以看到一些展示出的数据。这些数据都是我们平台最新的，下载最多的，或者评价最高的数据。如果你觉得这些数据很有趣，不妨点击试试看哦。	N/A	N/A	
re_7	9	谢谢。我想了解一下其他功能。	N/A	N/A	disableobject:ShowcaseImage
re_7	8	谢谢。我想开始自己浏览数据平台。	N/A	N/A	disableobject:ShowcaseImage
8	re_8	没问题！请点击右下方链接，开始使用数据平台吧。如果你遇到任何问题，点击导航栏"帮助"下的"询问助手"，就能与我对话。我会随时为你提供帮助。待会儿见！	N/A	N/A	link:http://sjfwcn3.winzome.com
re_8	finish	再见	N/A	N/A	
9	re_9	想让我介绍哪种功能呢？	N/A	N/A	
re_9	4	请介绍一下关键词搜索。	N/A	N/A	enableobject:SearchImage
re_9	5	请介绍一下数据页。	N/A	N/A	enableobject:DatapageImage
re_9	6	请介绍一下筛选和排序功能。	N/A	N/A	enableobject:FilterImage
re_9	7	请介绍一下数据展示功能。	N/A	N/A	enableobject:ShowcaseImage
re_9	8	不，谢谢。我想要自己探索平台。	N/A	N/A	

Dialogue 2 – FAQ

Start

Avatar: Good to see you again! Are there anything I could help?

User: I couldn't find the data I need. (1)

Would you please tell me more about this portal? (2)

(1)

Avatar: I'm sorry to hear that. Have you tried keyword search to find the data?

User: What is keyword search? (3)

No, I'll try that. (4)

Yes, I have tried but failed. (5)

(2)

Avatar: Sure, of course. This portal is built by the Government Information Office and other departments to disseminate different types of government data for free re-use. Do you know that the government and other public organizations hold a wealth of information, like the economic data, the geographical and meteorological data?

User: Oh, I don't know that before. (61)

Yes, please tell me more about it. (61)

(3)

Avatar: You could find the search bar on the top of all pages of this portal. By keyword search, you could simply type in the words of your problem in the search bar and click "Search", then you will get a list of related results. Sounds simple, right?

User: Yes, I'll try that. (4)

I have already tried but failed. (5)

(4)

Avatar: Yes, please go on. Hope you could find the data you need. Are there anything else I could help?

User: No, thanks. I'll try the keyword search first. (6)

Can you tell me more about other functions? (7)

(5)

Avatar: In that case, I would recommend you change another keyword and try again. How do you like this idea?

User: I will try some other keywords. (42)

I have tried several different keywords but still couldn't find what I need. (8)

No. My problem is there are too many results. (22)

(6)

Avatar: OK, no problem. Take your time. If you still have any problem when using our portal, I'll always be there to help.

User: Thank you, see you later. (End)

(7)

Avatar: Well, there are lots of different functions of our portal. Do you want to know about how to find the data quickly or do you want to know about how to get to know of our portal?

User: How to find what I need quickly. (48)

How to get to know about this portal? (49)

(8)

Avatar: Do you have an idea of the category that the data you need belongs to?

User: What do you mean by the category? (9)

Yes, kind of. (10)

No, I don't think so. (11)

(9)

Avatar: Actually, we provide 10 different categories of data on this portal. It was shown on the front page and on the left-hand side of the search result list. Please take a look.

User: Yes, I can see them. (12)

I still don't know your meaning. Could you give me an example? (13)

(10)

Avatar: Then I think you could try browsing the category that the data you need belongs to. Maybe you could find what you need.

User: Sounds good. I'll try that. (42)

But there are too many of them. Are there any other ways? (14)

I have tried but still couldn't find what I need. (19)

(11)

Avatar: In that case, I think maybe you could try browsing the first page of datasets of each category by clicking the name on the front page. It will give you a brief idea of the contents in each category.

User: OK. And what shall I do next? (15)

Where can I find these categories? (17)

(12)

Avatar: You could try browsing the first page of datasets of each category by clicking their name. It will give you a brief idea of the contents in each category.

User: OK. What shall I do after that? (15)

Where can I find these categories? (17)

(13)

Avatar: Could you see words like "Education", "Public Safety", and "Health" on the front page? These are the categories.

User: Yes, I can see them now. (16)

Where can I find these categories? (17)

(14)
Avatar: Yes. We provide filter functions and order functions. It will help you find what you need more quickly. Do you know about these functions?
User: Yes, I know them quite well. (20)
No, please explain more. (21)

(15)
Avatar: Could you get an idea of what category the data you want to find belongs to?
User: Yes, I got an idea. (10)
No, I am still not sure what category it belongs to. (18)

(16)
Avatar: Great! Now you can try browsing the first page of datasets of each category by clicking their name. It will give you a brief idea of the contents in each category.
User: OK. What shall I do after that? (15)
Where can I find these categories? (17)

(17)
Avatar: You could find these categories on the front page and on the left-hand side of the search result list. Can you see them?
User: Yes, I can see them now. (16)
No, can you give me some examples? (13)

(18)
Avatar: Then maybe you could try browsing by the data provider. Do you know how to do that?
User: Yes. (43)
No, please explain more. (44)
I have no idea of the possible provider of the datasets I'm looking for. (45)

(19)
Avatar: I think maybe we haven't released yet the datasets you need. Would you like me to help you submit an application for this dataset?
User: Yes, please. (26)
No, thanks. I will try to find the dataset again. (6)

(20)
Avatar: Please try them on the lists of datasets you get. It will reduce the length of the list so that you can much more easily find what you need.
User: OK, I will try that. (42)
Please explain how to use these functions. (21)
I have tried but still couldn't find what I need. (19)

(21)
Avatar: For filter functions, please look the left-hand side of the datasets list. You can filter the list by data type, data category or data provider. Just click on the name of the data type, data category or provider that you need, the system will filter the datasets list automatically.
User: OK, I will try that. (42)
What about the order function? (23)

(22)
Avatar: In that case, I think our filter functions and order functions will help you out. Do you know about these functions?
User: Yes, I know them quite well. (20)
No, please explain more. (21)

(23)
Avatar: Can you see a bar saying "Default Order" on the top right corner of the data list page?
User: Yes, I can see it. (24)
No, do you have any example? (25)

(24)
Avatar: You could change the order of the datasets in the list by clicking the down arrow and choosing the order method you prefer from the list.
User: I will try it now. (42)

(25)
Avatar: Please take a look at this screenshot. The left square shows the bar.
User: I can see it now. (24)

(26)
Avatar: Please describe a little bit about the content you need.
User: (Input box) (27)

(27)
Avatar: Do you have an idea of the category it belongs to?
User: Yes. (28)
No. (29)

(28)
Avatar: What category does the data you want belongs to?
User: (Multiple choice) (29)

(29)
Avatar: Would you please tell me what format do you want?
User: (Multiple choice) (30)

(30)
Avatar: OK. We are almost done. May I know the purpose you would use the data for?
User: (Multiple choice) (31)

(31)
Avatar: Great! We have recorded your requests. May I have your email address so that we could contact you once we got the data you need.
User: (input box) (32)

(32)
Avatar: Thank you for your patience! We will keep you up to date about your requests.
User: Thank you. (33)

(33)
Avatar: Would it be convenient for you to give us a shot feedback? It only takes less than a minute.
User: Yes, I can. (34)
No, maybe later. (35)

(34)
Avatar: Do you have any suggestions for our portal?
User: Yes. (36)
No. (37)

(35)
Avatar: OK, no problem. Are there anything else I could help?
User: Would you please tell me more about this portal? (2)
No, thanks. I'm fine at the moment. (39)

(36)
Avatar: Please type in your suggestions freely.
User: (Input Box) (37)

(37)
Avatar: How do you like our portal?
User: (Multiple choice) (38)

(38)
Avatar: How do you like the look of this portal?
User: (Multiple choice) (40)

(39)
Avatar: Good. If you need any help later, I'll always be there.
User: Thanks. (End)

(40)

Avatar: How do you like the functions of our portal?

User: (Multiple choice) (41)

(41)

Avatar: Thank you very much for your feedback. Are there anything else I could help?

User: Would you please tell me more about this portal? (2)

No, thanks. I’m fine at the moment. (39)

(42)

Avatar: Yes, please go on and take your time. Hope you could find the data you need. Are there anything else I could help?

User: No, thanks. I’ll try out by myself first. (6)

Can you tell me more about other functions? (7)

(43)

Avatar: Great! You could see if the provider has uploaded the data you need or not.

User: OK, I will try that. (42)

The data list of the provider is too long. (22)

I have tried but still couldn’t find the data I need. (19)

(44)

Avatar: On the left-hand side of the dataset list, you could see several different data providers. Just click on the possible one, and our system will give you a list of datasets uploaded by this provider automatically. It’s easy, right?

User: Yes, I will try that. (42)

I couldn’t find the data providers you said. (46)

(45)

Avatar: Well, I think I could help you submit a data application. If we have not yet uploaded the datasets you need, they will inform you the outcome. If we have already provided the data you need, you will get a link to the dataset.

User: Yes, please help me submit an application. (26)

No, thanks. I will try to find the dataset again. (6)

(46)

Avatar: Please look at this screenshot. The third list on the left-hand side is the provider list.

User: I see. (47)

(47)

Avatar: Now you can try browsing the data list of the possible provider of the dataset you want.

User: Yes, I will try that. (42)

The data list of the provider is too long. (22)

(48)

Avatar: I would recommend keyword search as a first step and then combine it with filter function and order function. Do you know about keyword search?

User: Yes. (50)

No, please explain more. (51)

(49)

Avatar: There are several ways to get to know about the portal. You could start by visiting our homepage. We have listed the data categories we provided on our portal and also showed the latest datasets, the most popular and highest ranked datasets.

User: What do you mean by data category? (55)

What will happen if I click at one of those showcased datasets? (56)

(50)

Avatar: Then would you like me to explain filter function and order function?

User: What is filter function? (52)

What is order function? (53)

No, thanks. (54)

(51)

Avatar: You could find a search bar on the top of all pages of this portal. Type in the words of your problem in the search bar and click "Search", then you will get a list of related results. Sounds easy, right? Would you like to know filter function and order function as well?

User: What is filter function? (52)

What is order function? (53)

No, thanks. (54)

(52)

Avatar: For filter functions, please look the left-hand side of the datasets list. You can filter the list by data type, data category or data provider. Just click on the name of the data type, data category or provider that you need, the system will filter the datasets list automatically. (Picture)

User: What about order function? (53)

(53)

Avatar: On the top right corner of the data list page, there is a bar saying "Default Order". You could change the order of the datasets in the list by clicking the down arrow and selecting the order method you prefer from the list. (Picture)

User: Got it. Thank you. (54)

(54)

Avatar: No worries. Would you like to know about how to get to know of our portal?

User: Yes, please explain. (49)

No, maybe later. (6)

(55)
Avatar: You could see words like "Education", "Health", and "Environment" on the homepage. These 10 words are categories of the data we provide on this portal.
User: I see. (57)

(56)
Avatar: It will lead you to the data page. You could find detailed information about this dataset without the need of downloading and opening the dataset. Sounds cool, right?
User: Yes, I like it. (60)
Maybe. (60)

(57)
Avatar: Do you know about the data page?
User: Yes, I think I know it well. (58)
No, please explain more. (59)

(58)
Avatar: Great! Are there anything else I could help?
User: Please tell me more about this portal itself. (2)
No, I'm fine by myself. (6)

(59)
Avatar: You could be led to a data page if you click on the title of a datasets. On this page, you can find detailed information about this dataset without the need of downloading and opening the dataset. How does it sound to you?
User: I like it. (60)
I don't know. (60)

(60)
Avatar: Are there anything else I could help?
User: Please tell me more about this portal itself. (2)
No, I'm fine by myself. (6)

(61)
Avatar: We believe if the data owned by the government are opened to the public and re-used, it could increase the value and bring the benefits to our whole society. Since we encourage the re-use of our data, would you like to know how you could use the data on the portal?
User: Yes, please tell me more about it. (62)
No, maybe later. (63)

(62)
Avatar: You can use the data for both commercial and non-commercial purposes. But you need to follow our Terms and Conditions of Use, which you could find here (link). Do you know what is forbidden in the use of our data?

User: No, please tell me more about it. (64)
No, but I will find out by myself. (63)
I think I know. (69)

(63)
Avatar: OK, sure. Are there anything else I could help?
User: No, thanks. (6)
Yes, I would like to know about the functions of this portal. (7)

(64)
Avatar: You are encouraged to make all kinds of innovative uses of the data on the portal, like building and deploying applications. But you shouldn't re-sale the data. More details can be found on the page for our Terms and Conditions of Use, which you could find here (link).
User: Thank you for the information. (65)

(65)
Avatar: It's my pleasure. Do you plan to use our data for commercial purposes?
User: Maybe. (66)
I don't think so. (67)

(66)
Avatar: I'm happy to tell you that you don't need to pay any fees for both commercial and non-commercial re-use of data on this portal. So feel free to create cool things with our data!
User: Sounds great! (68)

(67)
Avatar: OK. But just let you know that you don't need to pay any fees for both commercial and non-commercial re-use of data on this portal. So feel free to create cool things with our data!
User: Good to know that! (68)

(68)
Avatar: I'm glad you like it. Are there anything else I could help?
User: No, thanks. (6)
Yes, I would like to know about the functions of this portal. (7)

(69)
Avatar: Excellent! But just remind you that you shouldn't re-sale the data from this portal. More details can be found on the page for our Terms and Conditions of Use, which you could find here (link).
User: Thank you for the information. (65)

Current State	Next State	Utterance	Meaning	Style	Actions
Start	0	点击开始	N/A	N/A	
0	re_0	你好！又见面了！请问有什么我可以帮忙的？	N/A	N/A	
re_0	1	我无法找到我要的数据。	N/A	N/A	
re_0	2	能告诉我更多关于这个平台的信息吗？	N/A	N/A	
1	re_1	请不要着急。你尝试过用关键词搜索你要的数据了吗？	N/A	N/A	
re_1	3	什么是关键词搜索？	N/A	N/A	
re_1	4	还没有，我要尝试一下。	N/A	N/A	
re_1	5	是的，我已经尝试过了，但没找到想要的数据	N/A	N/A	
2	re_2	当然可以。这个平台是由政府信息科技办公室与各部门一起建立的，用来免费发放各种公共资料，提供给市民大众免费使用。你知道吗？政府和各个公共部门拥有很多丰富的数据资源，像经济数据，地理数据和气象数据等。	N/A	N/A	
re_2	61	哦，第一次听说。	N/A	N/A	
re_2	61	嗯，我之前听说过。	N/A	N/A	
3	re_3	你可以在这个平台所有页面的最上端找到搜索栏。如果要使用关键词搜索，你只需要在搜索栏中输入你想要的数据的关键词，并单击"搜索"。这样，你就能够获得一个关于该关键词的搜索结果列表。很简单吧？	N/A	N/A	
re_3	4	是的，我要尝试一下。	N/A	N/A	
re_3	5	我已经尝试过了，但没找到想要的数据。	N/A	N/A	
4	re_4	好的，请一定试试看。希望你能找到想要的数据。还有什么我可以帮忙的吗？	N/A	N/A	
re_4	6	不，谢谢。我想先试一试关键词搜索。	N/A	N/A	
re_4	7	你能给我介绍一下其他功能吗？	N/A	N/A	
5	re_5	这样的话，我推荐你换个其他的关键词再试试看。你觉得这个主意怎么样？	N/A	N/A	
re_5	42	嗯，我会尝试更换其他的关键词。	N/A	N/A	
re_5	8	我已经试了好几个不同的关键词，但还是找不到需要的数据。	N/A	N/A	
re_5	22	不，我的问题是，得到的搜索结果太多了。	N/A	N/A	
6	re_6	好的，没问题。慢慢来。如果你在使用平台过程中遇到其他问题，我会随时提供帮助的。	N/A	N/A	
re_6	finish	谢谢你，稍后再见。	N/A	N/A	
7	re_7	我们平台具有多种不同的功能。你想了解如何快速找到你想要的数据吗？还是想要了解一下关于平台本身的情况呢？	N/A	N/A	
re_7	48	我想了解如何快速找到所需的数据。	N/A	N/A	
re_7	49	我想了解关于平台本身的情况。	N/A	N/A	
8	re_8	你是否知道自己想要的数据大致属于哪一个数据主题范围？	N/A	N/A	
re_8	9	什么是数据主题范围？	N/A	N/A	
re_8	10	嗯，大概知道。	N/A	N/A	
re_8	11	不，我并不知道。	N/A	N/A	
9	re_9	这个平台上的数据大致属于十个不同的数据主题。在主页和搜索结果列表的左侧，你可以看到这十个数据主题。找到了吗？	N/A	N/A	
re_9	12	是的，我找到了。	N/A	N/A	
re_9	13	我还是不知道你说的是什么。你能举个例子	N/A	N/A	
10	re_10	如果知道大致主题范围的话，我觉得你可以试试浏览这个主题中所有的数据集，看看有没有你需要的数据。	N/A	N/A	
re_10	42	听上去不错。我会试试看。	N/A	N/A	
re_10	14	但数据太多了。有其他办法吗？	N/A	N/A	
re_10	19	我已经试过了但还是没找到需要的数据。	N/A	N/A	
11	re_11	这样的话，我建议你，先通过主页上点选各个数据主题，快速浏览一下这些主题数据集列表的第一页。这样可以让你对各个数据主题所涵盖的内容有一个大致的概念。	N/A	N/A	
re_11	15	好的。那我接下来应该做什么？	N/A	N/A	
re_11	17	我都可以在哪里找到这些主题？	N/A	N/A	
12	re_12	你可以先点选各个数据主题，快速浏览一下这些主题数据集列表的第一页。这样可以让你对各个数据主题所涵盖的内容有一个大致的概念	N/A	N/A	
re_12	15	好的。那我接下来应该做什么？	N/A	N/A	
re_12	17	我都可以在哪里找到这些主题？	N/A	N/A	
13	re_13	你可以在首页上找到"教育科技"，"公共安全"，医疗健康"这些词语吗？这些就是各个数据主题	N/A	N/A	
re_13	16	我现在找到了。	N/A	N/A	
re_13	17	我还可以在哪里找到这些主题？	N/A	N/A	
14	re_14	我们针对这种情况提供了筛选和排序功能。它们能够帮助你快速找到所需的数据。你想进一步了解这些功能吗？	N/A	N/A	
re_14	20	不用了。我已经很了解它们了。	N/A	N/A	
re_14	21	请详细介绍一下这些功能。	N/A	N/A	
15	re_15	现在你能大致判断出自己需要的数据属于哪一种主题了吗？	N/A	N/A	
re_15	10	是的，我有了大致的想法。	N/A	N/A	
re_15	18	不，我还是无法判断我要的数据属于哪种主题	N/A	N/A	
16	re_16	太好了。现在你可以点选各个数据主题，快速浏览一下这些主题数据集列表的第一页。这样可以让你对各个数据主题所涵盖的内容有一个大致的概念。	N/A	N/A	
re_16	15	好的。那我接下来应该做什么？	N/A	N/A	
re_16	17	我都可以在哪里找到这些主题？	N/A	N/A	
17	re_17	你可以在首页上，以及在检索结果列表的左侧，找到这些主题。现在，你能看到这些主题	N/A	N/A	
re_17	16	是的，我能找到它们。	N/A	N/A	
re_17	13	不能。你能给我举个例子吗？	N/A	N/A	
18	re_18	那么或许你可以尝试按照数据提供者浏览一下数据。你知道怎么按照数据提供者来浏览数据	N/A	N/A	
re_18	43	知道。	N/A	N/A	
re_18	44	不知道，请解释一下如何操作？	N/A	N/A	
re_18	45	我不知道我想要的数据的提供者应该是谁。	N/A	N/A	
19	re_19	我觉得，我们可能还没有提供你所需要的数据。需要我帮你提交一个数据申请吗？	N/A	N/A	
re_19	26	好的，谢谢。	N/A	N/A	

re_19	6	不用了，谢谢。我想再试试看能不能找到我要的数据。	N/A	N/A	
20	re_20	你可以在数据集列表上使用一下筛选和排序功能。它们能够减少查找到你所需数据的时间。	N/A	N/A	
re_20	42	好的，我会试试看。	N/A	N/A	
re_20	21	能讲解一下怎么使用筛选和排序功能吗?	N/A	N/A	
re_20	19	我已经尝试过了，但还是没找到所需的数据。	N/A	N/A	
21	re_21	对于筛选功能，请先看一下数据列表的左侧。你可以按照资源类型、数据领域和数据提供单位进行筛选。只要点击你所需要的资源类型、数据领域和数据提供单位，系统就会自动对数据列表的内容进行筛选了。	N/A	N/A	
re_21	42	好的，我会试试看。	N/A	N/A	
re_21	23	那怎么使用排序功能呢?	N/A	N/A	
22	re_22	如果是这种情况，我推荐你使用我们提供的筛选和排序功能。你知道如何使用这两种功能	N/A	N/A	
re_22	20	是的，我非常清楚如何使用。	N/A	N/A	
re_22	21	不清楚，请详细讲解一下。	N/A	N/A	
23	re_23	你能看到在数据列表页右上角，有一个写着'默认排序'的方框吗?	N/A	N/A	
re_23	24	是的，我能看到。	N/A	N/A	
re_23	25	不能，你能给我举个例子吗?	N/A	N/A	enableobject:FilterImage
24	re_24	你可以通过点击这个框，在下拉列表中选择你喜欢的排序方式，来改变数据列表中数据的排	N/A	N/A	
re_24	42	我现在试试看。	N/A	N/A	
25	re_25	请看这张截图，右侧高亮的方框表示的就是排序方式。	N/A	N/A	
re_25	24	我看到了。	N/A	N/A	disableobject:FilterImage
26	re_26	请简单描述一下你所需要的数据内容。	N/A	N/A	
re_26	27	(input)点击提交	N/A	N/A	
27	re_27	你知道自己所需的数据大致属于哪个数据主题范围吗?	N/A	N/A	
re_27	28	知道。	N/A	N/A	
re_27	29	不知道。	N/A	N/A	
28	re_28	你所需的数据属于哪个数据主题范围?	N/A	N/A	
re_28	29	政府财政	N/A	N/A	
re_28	29	信用记录	N/A	N/A	
re_28	29	文化休闲	N/A	N/A	
re_28	29	教育科技	N/A	N/A	
re_28	29	环境质量	N/A	N/A	
re_28	29	政府招投标	N/A	N/A	
re_28	29	医疗健康	N/A	N/A	
re_28	29	国际贸易	N/A	N/A	
re_28	29	区域统计	N/A	N/A	
re_28	29	地理地图	N/A	N/A	
re_28	29	法律政策	N/A	N/A	
re_28	29	公共安全	N/A	N/A	
re_28	29	注册登记	N/A	N/A	
re_28	29	道路交通	N/A	N/A	
re_28	29	气候气象	N/A	N/A	
29	re_29	你需要哪种格式的数据?	N/A	N/A	
re_29	30	csv	N/A	N/A	
re_29	30	pdf	N/A	N/A	
re_29	30	xml	N/A	N/A	
re_29	30	xls	N/A	N/A	
re_29	30	url	N/A	N/A	
re_29	30	其他	N/A	N/A	
30	re_30	我们马上就要完成申请了。你能告诉我你将用这个数据干什么吗?	N/A	N/A	
re_30	31	日常生活	N/A	N/A	
re_30	31	科学研究	N/A	N/A	
re_30	31	软件开发	N/A	N/A	
re_30	31	新闻报道	N/A	N/A	
re_30	31	经营决策	N/A	N/A	
re_30	31	监督知情	N/A	N/A	
re_30	31	其他	N/A	N/A	
31	re_31	好的。我们已经记录了你的数据申请。现在你能告诉我你的邮箱地址吗？这样我们可以及时通知你数据申请的结果。	N/A	N/A	
re_31	32	(input)点击提交	N/A	N/A	
32	re_32	谢谢你的耐心回复！我们会及时通知你数据申请的最新进展。	N/A	N/A	
re_32	33	谢谢。	N/A	N/A	
33	re_33	请问你是否愿意花一分钟时间，给我们一个简单的使用反馈?	N/A	N/A	
re_33	34	好的，我愿意。	N/A	N/A	
re_33	35	现在不太方便。	N/A	N/A	
34	re_34	你对我们的平台有什么建议吗?	N/A	N/A	
re_34	36	有。	N/A	N/A	
re_34	37	没有。	N/A	N/A	
35	re_35	好的，没关系。那还有什么我可以帮忙的吗?	N/A	N/A	
re_35	2	能告诉我更多关于这个平台的信息吗?	N/A	N/A	
re_35	39	不，谢谢。暂时没有了。	N/A	N/A	
36	re_36	你可以随意在下面的方框内写下你对我们的建议。	N/A	N/A	
re_36	37	(input)点击提交	N/A	N/A	
37	re_37	你觉得我们的数据平台怎么样?	N/A	N/A	
re_37	38	非常好	N/A	N/A	
re_37	38	比较好	N/A	N/A	
re_37	38	一般	N/A	N/A	
re_37	38	比较差	N/A	N/A	
re_37	38	非常差	N/A	N/A	
38	re_38	你觉得我们的数据平台的外观看起来如何?	N/A	N/A	
re_38	40	非常好	N/A	N/A	
re_38	40	比较好	N/A	N/A	
re_38	40	一般	N/A	N/A	
re_38	40	比较差	N/A	N/A	
re_38	40	非常差	N/A	N/A	
39	re_39	好的。如果你还有什么困难，我随时都会提供帮助的。	N/A	N/A	
re_39	finish	谢谢。	N/A	N/A	
40	re_40	你觉得我们的数据平台的各项功能怎么样?	N/A	N/A	

re_40	41	非常好	N/A	N/A	
re_40	41	比较好	N/A	N/A	
re_40	41	一般	N/A	N/A	
re_40	41	比较差	N/A	N/A	
re_40	41	非常差	N/A	N/A	
41	re_41	非常感谢你的反馈。还有什么我可以帮忙的	N/A	N/A	
re_41	2	能告诉我更多关于这个平台的信息吗?	N/A	N/A	
re_41	39	不，谢谢。暂时没有了。	N/A	N/A	
42	re_42	好的，慢慢来。希望你能找到需要的数据。还有什么我可以帮忙的吗?	N/A	N/A	
re_42	6	不用了，谢谢。我想先自己试试看。	N/A	N/A	
re_42	7	你能给我介绍一下其他功能吗?	N/A	N/A	
43	re_43	非常好！这样你可以选择可能的数据提供者，来浏览一下他们有没有提供你所需要的数据。	N/A	N/A	
re_43	42	好的，我会试试看。	N/A	N/A	
re_43	22	这个数据提供者提供的数据太多了。	N/A	N/A	
re_43	19	我已经尝试过了，但还是没有找到我需要的数据。	N/A	N/A	
44	re_44	在数据列表的左侧，你可以看到很多不同的数据提供者。只要点击你所需的数据可能的数据提供者，我们的系统就会自动反馈给你一个这个数据提供者上传的所有数据的列表。听上去很简单吧?	N/A	N/A	
re_44	42	是的，我会试试看。	N/A	N/A	
re_44	46	我不知道你说的数据提供者在哪里。	N/A	N/A	enableobject:FilterImage
45	re_45	我觉得我可以帮你提交一个数据申请。如果我们确实还没有提供你所需要的数据，那我们会通过邮件告知你数据申请的结果。如果我们已经提供了你所需的数据，那你会收到一个关于你所需数据的链接。	N/A	N/A	
re_45	26	好的，请帮我提交一个数据申请。	N/A	N/A	
re_45	6	不用了，谢谢。我想再试试查找我要的数据。	N/A	N/A	
46	re_46	请看这张截图。左侧第三栏就是我所说的数据提供者。	N/A	N/A	
re_46	47	好的，我懂了。	N/A	N/A	disableobject:FilterImage
47	re_47	现在，你可以尝试浏览一下你所需数据的提供者是否已经上传了你需要的数据。	N/A	N/A	
re_47	42	好的，我试试看。	N/A	N/A	
re_47	22	这个数据提供者上传的数据太多了。	N/A	N/A	
48	re_48	我优先推荐关键词检索，并且结合筛选和排序功能。你了解关键词检索吗?	N/A	N/A	
re_48	50	是的，我了解。	N/A	N/A	
re_48	51	不了解，请再详细介绍一下。	N/A	N/A	
49	re_49	你可以通过多种途径了解我们的平台。你可以从浏览主页开始。我们在主页上列出了本平台涵盖的数据主题范围，还展示了一些最新，最受欢迎的数据资源。	N/A	N/A	
re_49	55	什么是数据主题?	N/A	N/A	
re_49	56	如果我点击了一个展示的数据，会发生什么?	N/A	N/A	
50	re_50	那你需要我介绍一下筛选和排序功能吗?	N/A	N/A	
re_50	52	请介绍一下筛选功能。	N/A	N/A	
re_50	53	请介绍一下排序功能。	N/A	N/A	
re_50	54	不用了，谢谢。	N/A	N/A	
51	re_51		N/A	N/A	
re_51	52	请介绍一下筛选功能。	N/A	N/A	enableobject:FilterImage
re_51	53	请介绍一下排序功能。	N/A	N/A	enableobject:FilterImage
re_51	54	不用了，谢谢。	N/A	N/A	
52	re_52	如果要使用筛选功能，请看数据列表的左侧。你可以按照资源类型，数据领域和数据提供单位进行筛选。只需要点击你需要的资源类型，数据领域和数据提供单位，系统将会自动对数据列表进行筛选。	N/A	N/A	
re_52	53	那排序功能呢?	N/A	N/A	disableobject:FilterImage
53	re_53	在数据列表页右上角，有一个写着"默认排序"的方框。点击这个方框，在下拉列表中，你可以选择你喜欢的排序方式，来改变数据列表中数据的排序。	N/A	N/A	
re_53	54	好的，知道了。谢谢。	N/A	N/A	disableobject:FilterImage
54	re_54	不客气。需要我介绍一下如何了解这个平台	N/A	N/A	
re_54	49	好的，请介绍一下。	N/A	N/A	
re_54	6	不用了，暂时不需要。	N/A	N/A	
55	re_55	在主页上，你能够看到像"教育科技"，"公共安全"，"医疗健康"这样的词语。这十个词语就是这个数据平台所涵盖的数据主题范围。	N/A	N/A	
re_55	24	好的，我懂了。	N/A	N/A	
56	re_56	你会来到关于这个数据资源的页面。在这个页面上，你不需要下载和打开数据资源，就可以了解到关于这个数据资源的具体信息。听上去很棒吧?	N/A	N/A	
re_56	60	是的，这很好。	N/A	N/A	
re_56	60	嗯，大概吧。	N/A	N/A	
57	re_57	你了解数据页吗?	N/A	N/A	
re_57	58	我认为我很了解。	N/A	N/A	
re_57	59	我不太了解，你可以具体解释一下吗?	N/A	N/A	
58	re_58	很好！那还有什么我可以帮你的吗?	N/A	N/A	
re_58	2	能告诉我更多关于这个平台的信息吗?	N/A	N/A	
re_58	6	不，暂时没有了。	N/A	N/A	
59	re_59	如果你点击了一个数据资源，你就会来到关于这个数据资源的页面。在这个页面上，你不需要下载和打开数据资源，就可以了解到关于这个数据资源的具体信息。听上去很棒吧?	N/A	N/A	
re_59	60	还不错。	N/A	N/A	
re_59	60	嗯，大概吧。	N/A	N/A	
60	re_60	还有什么我可以帮你的吗?	N/A	N/A	
re_60	2	能告诉我更多关于这个平台的信息吗?	N/A	N/A	
re_60	6	不，暂时没有了。	N/A	N/A	
61	re_61	我们相信，如果政府将他们所拥有的数据向社会大众开放，能够为社会带来更大的益处。我们非常鼓励对平台上的数据进行再利用。需要我讲解一下，你都可以如何使用平台上的数据	N/A	N/A	
re_61	62	请具体讲解一下吧。	N/A	N/A	A link will show up: http://gfwcn2.winzome.com/govWeb/usage.html
re_61	63	以后再说吧。	N/A	N/A	

62	re_62	你可以将这个平台上的数据用于商业和非商业性质的目的。但你必须要遵循我们平台的使用条款及条件，你可以在下面的链接中找到这些规定。你知道在使用这个平台的数据时，哪种行为是被禁止的吗？	N/A	N/A	
re_62	64	不知道，可以告诉我吗？	N/A	N/A	A link will show up: http://sjfwcn2.winzome.com/govWeb/usage.html
re_62	63	不知道，我想自己了解一下。	N/A	N/A	
re_62	69	我觉得我知道。	N/A	N/A	A link will show up: http://sjfwcn2.winzome.com/govWeb/usage.html
63	re_63	当然可以。还有什么我可以帮你的吗?	N/A	N/A	
re_63	6	暂时没有了，谢谢。	N/A	N/A	
re_63	7	我想了解一下这个平台的各项功能。	N/A	N/A	
64	re_64	你可以充分发掘数据资源的各种创新用途，比如开发各种应用软件。但你不可以转售平台上的数据。更多关于平台数据的使用条款，可以在这个链接中找到。	N/A	N/A	
re_64	65	好的，谢谢!	N/A	N/A	
65	re_65	不客气。你打算将我们的数据用于商业性质的用途吗?	N/A	N/A	
re_65	66	大概会。	N/A	N/A	
re_65	67	应该不会。	N/A	N/A	
66	re_66	我可以很高兴的告诉你，不论你将平台数据用于商业还是非商业性质的用途，政府都不会对本平台公开发布的数据资源收取任何费用。所以，请尽情使用我们平台的资源，开发新鲜有趣的应用吧!	N/A	N/A	
re_66	68	听上去真棒!	N/A	N/A	
67	re_67	好的。不过，我还是告诉你一声，不论你将平台数据用于商业还是非商业性质的用途，政府都不会对本平台公开发布的数据资源收取任何费用。所以，请尽情使用我们平台的资源!	N/A	N/A	
re_67	68	好的，听上去不错。	N/A	N/A	
68	re_68	是的，很高兴你喜欢这一点。还有什么我可以帮你的吗?	N/A	N/A	
re_68	6	暂时没有了，谢谢。	N/A	N/A	
re_68	7	我想了解一下这个平台的各项功能。	N/A	N/A	
69	re_69	你真棒! 不过，还是提醒你一下，转售这个平台上的数据是被禁止的。更多关于平台数据的使用条款，可以在这个链接中找到。	N/A	N/A	
re_69	65	好的，感谢提醒。	N/A	N/A	

Dialogue 3 – Data Application

Start

Avatar: Good to see you again! Do you want to submit an online data application?

User: Yes, I couldn't find the data I need. (1)

No, I would like to see if I could find the data I need by myself. (2)

(6)

Avatar: OK, no problem. Take your time. If you still have any problem when using our portal, I'll always be there to help.

User: Thank you, see you later. (End)

(26)

Avatar: Please describe a little bit about the content you need.

User: (Input box) (27)

(27)

Avatar: Do you have an idea of the category it belongs to?

User: Yes. (28)

No. (29)

(28)

Avatar: What category does the data you want belongs to?

User: (Multiple choice) (29)

(29)

Avatar: Would you please tell me what format do you want?

User: (Multiple choice) (30)

(30)

Avatar: OK. We are almost done. May I know the purpose you would use the data for?

User: (Multiple choice) (31)

(31)

Avatar: Great! We have recorded your requests. May I have your email address so that we could contact you once we got the data you need.

User: (input box) (32)

(32)

Avatar: Thank you for your patience! We will keep you up to date about your requests.

User: Thank you. (33)

(70)

Avatar: It's my pleasure. If you meet with any other problems when using the portal, I'm always be there for help.

User: Thank you, see you later. (End)

Current State	Next State	Utterance	Meaning	Style	Actions
Start	0	点击开始	N/A	N/A	
0	re_0	你好！又见面了！请问你想要提交一个数据申请吗？	N/A	N/A	
re_0	26	是的，我无法找到我需要的数据。	N/A	N/A	
re_0	6	不，我想再试试看能不能找到我要的数据。	N/A	N/A	
6	re_6	好的，没问题。慢慢来。如果你在使用平台过程中遇到其他问题，我会随时提供帮助的。	N/A	N/A	
re_6	finish	谢谢你，稍后再见。	N/A	N/A	
26	re_26	请简单描述一下你所需要的数据内容。	N/A	N/A	
re_26	27	[input]点击提交	N/A	N/A	
27	re_27	你知道自己所需的数据大致属于哪个数据主题范围吗？	N/A	N/A	
re_27	28	知道。	N/A	N/A	
re_27	29	不知道。	N/A	N/A	
28	re_28	你所需的数据属于哪个数据主题范围？	N/A	N/A	
re_28	29	政府财政	N/A	N/A	
re_28	29	信用记录	N/A	N/A	
re_28	29	文化休闲	N/A	N/A	
re_28	29	教育科技	N/A	N/A	
re_28	29	环境质量	N/A	N/A	
re_28	29	政府招投标	N/A	N/A	
re_28	29	医疗健康	N/A	N/A	
re_28	29	国际贸易	N/A	N/A	
re_28	29	区域统计	N/A	N/A	
re_28	29	地理地图	N/A	N/A	
re_28	29	法律政策	N/A	N/A	
re_28	29	公共安全	N/A	N/A	
re_28	29	注册登记	N/A	N/A	
re_28	29	道路交通	N/A	N/A	
re_28	29	气候气象	N/A	N/A	
29	re_29	你需要哪种格式的数据？	N/A	N/A	
re_29	30	csv	N/A	N/A	
re_29	30	pdf	N/A	N/A	
re_29	30	xml	N/A	N/A	
re_29	30	xls	N/A	N/A	
re_29	30	url	N/A	N/A	
re_29	30	其他	N/A	N/A	
30	re_30	我们马上就要完成申请了。你能告诉我你将用这个数据干什么吗？	N/A	N/A	
re_30	31	日常生活	N/A	N/A	
re_30	31	科学研究	N/A	N/A	
re_30	31	软件开发	N/A	N/A	
re_30	31	新闻报道	N/A	N/A	
re_30	31	经营决策	N/A	N/A	
re_30	31	监督知情	N/A	N/A	
re_30	31	其他	N/A	N/A	
31	re_31	好的。我们已经记录了你的数据申请。现在你能告诉我你的邮箱地址吗？这样我们可以及时通知你数据申请的结果。	N/A	N/A	
re_31	32	[input]点击提交	N/A	N/A	
32	re_32	谢谢你的耐心回复！我们会及时通知你数据申请的最新进展的。	N/A	N/A	
re_32	70	谢谢。	N/A	N/A	
70	re_70	不客气。如果你在使用平台过程中遇到其他问题，我会随时提供帮助的。	N/A	N/A	
re_70	finish	谢谢你，稍后再见。	N/A	N/A	

Dialogue 4 – Feedback

Start

Avatar: Good to see you again! Would you like to submit a user feedback?

User: Yes, it's my pleasure. (34)

No, I don't want to submit at present. (71)

(34)

Avatar: Do you have any suggestions for our portal?

User: Yes. (36)

No. (37)

(36)

Avatar: Please type in your suggestions freely.

User: (Input Box) (37)

(37)

Avatar: How do you like our portal?

User: (Multiple choice) (38)

(38)

Avatar: How do you like the look of this portal?

User: (Multiple choice) (40)

(40)

Avatar: How do you like the functions of our portal?

User: (Multiple choice) (41)

(71)

Avatar: OK, no problem. If you meet with any other problems when using the portal, I'm always be there for help.

User: Thank you, see you later. (End)

(72)

Avatar: Thank you very much for filling the feedback. If you meet with any other problems when using the portal, I'm always be there for help.

User: Thank you, see you later. (End)

Current State	Next State	Utterance	Meaning	Style	Actions
Start	0	点击开始	N/A	N/A	
0	re_0	你好！又见面了！请问你想要提交一个使用反馈吗？	N/A	N/A	
re_0	34	是的。	N/A	N/A	
re_0	71	不，现在不太方便。	N/A	N/A	
34	re_34	你对我们的平台有什么建议吗？	N/A	N/A	
re_34	36	有。	N/A	N/A	
re_34	37	没有。	N/A	N/A	
36	re_36	你可以随意在下面方框内写下你对我们的建议。	N/A	N/A	
re_36	37	[input]点击提交	N/A	N/A	
37	re_37	你觉得我们的数据平台怎么样？	N/A	N/A	
re_37	38	非常好	N/A	N/A	
re_37	38	比较好	N/A	N/A	
re_37	38	一般	N/A	N/A	
re_37	38	比较差	N/A	N/A	
re_37	38	非常差	N/A	N/A	
38	re_38	你觉得我们的数据平台的外观看起来如何？	N/A	N/A	
re_38	40	非常好	N/A	N/A	
re_38	40	比较好	N/A	N/A	
re_38	40	一般	N/A	N/A	
re_38	40	比较差	N/A	N/A	
re_38	40	非常差	N/A	N/A	
40	re_40	你觉得我们的数据平台的各项功能怎么样？	N/A	N/A	
re_40	72	非常好	N/A	N/A	
re_40	72	比较好	N/A	N/A	
re_40	72	一般	N/A	N/A	
re_40	72	比较差	N/A	N/A	
re_40	72	非常差	N/A	N/A	
71	re_71	好的，没关系。如果你在使用平台过程中遇到其他问题，我会随时提供帮助的。	N/A	N/A	
re_71	finish	谢谢你，稍后再见。	N/A	N/A	
72	re_72	非常感谢你的反馈。如果你在使用平台过程中遇到其他问题，我会随时提供帮助的。	N/A	N/A	
re_72	finish	谢谢你，稍后再见。	N/A	N/A	

APPENDIX D: COMPLETE INSTRUMENTS FOR STAGE 4

Faculty of Science and Engineering

MACQUARIE University

Start

Participant Information and Consent Form

Name of Project: Using OGD portals

You are invited to participate in a study on 'Using OGD portals'. The purpose of this study is to investigate the adoption of open government data (OGD) portals, which are websites for the public to get access to the government data that can be freely used, re-used and distributed.

The study is being conducted by, Ms. Di Wang, Department of Computing, di.wang18@students.mq.edu.au to meet the requirements of PhD under the supervision of Professor Deborah Richards, 61 (0)2 9850 9567, deborah.richards@mq.edu.au, Department of Computing, Faculty of Science.

If you decide to participate, first you will complete a demographic questionnaire and a questionnaire about your opinion of OGD portals' functions. Then you will use a simulated OGD portal to complete 5 tasks about finding certain data on the portal. Finally, you will answer some questions about your experience using the portal. The duration of the study is expected to be around 35 minutes.

No personal details are collected. Your data will be anonymous. No individual will be identified in any publication of the results. Access to unidentified data may be made available to other researchers if they are interested.

Your name and identity will not be linked to the data generated by your responses to our survey items. A generic ID will replace your name on the data collected.

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics & Integrity (Telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

Participation in this study is entirely voluntary: you are not obliged to participate and if you decide to participate, you are free to withdraw at any time without having to give a reason and without consequence.

It is required for this study to be taken using Firefox or Chrome browsers for Windows, or Safari or Chrome for Mac. This study cannot be completed on mobile phones or tablets.

隐私与同意参与调查表

项目名称：开放政府数据平台使用调查

诚挚邀请您参加一项名为"开放政府数据平台使用"的调查研究。该研究的目的在于了解社会公众对开放政府数据

平台的接受状况。所谓开放政府数据平台，是政府设立的面向社会公众提供可以随意使用，再开发或传播的政府数据的网络平台。

该研究是由麦考瑞大学计算机系王迪同学（di.wang18@students.mq.edu.au）为达到博士毕业要求，在Deborah Richards教授（电话61(0)298509567，邮箱deborah.richards@mq.edu.au）的指导下设计并实施。

如果您决定参加这项研究，首先，您将会回答一个关于您概况的问卷和一个关于您看待开放政府数据平台各项功能的问卷。之后，您会使用一个模拟的开放政府数据平台查找相关数据来完成五个任务。最后，您将会回答一些关于您平台使用体验的问题。本项研究预计耗时35分钟。

我们不会收集任何关于您的私人数据。所有收集的数据都将做匿名处理。您的姓名和身份不会与您在研究中所提供的数据进行关联。一个重新命名的编号将被用来替换您的身份。本项研究发布的所有成果中将不会有任何个人信息被识别。但本研究经过匿名处理后的数据可能会被提供给其他对本研究感兴趣的研究者。

本项研究完全自愿参与。同时，在您决定参加之后，仍然可以在中途任何时间选择退出调查且不用提供任何理由，也无需承担任何后果。

本研究涉及的科研伦理由麦考瑞大学人类研究伦理委员会审核通过。如果您对于您参与的研究有任何关于伦理道德方面的疑问和投诉，您可以与该委员会联系。电话：(02)98507854，邮箱：ethics@mq.edu.au。您的所有投诉都将做保密处理并被仔细调查，同时，您也会被告知处理结果。

本研究需要在Windows系统上使用Firefox或Chrome浏览器，或在Mac系统上使用Safari或Chrome浏览器完成。本研究无法在手机和平板上完成。

- I have the correct browser on a computer and wish to participate in this study.
  我在电脑上使用了正确的浏览器，并愿意参与这项研究。
- I have changed my mind and don't want to participate in this study.
  我并不愿意参与这项研究

Survey 1: Demographic Questionnaire

Your gender 您的性别

- Male 男性
- Female 女性
- Other 其他

Your age 您的年龄

- Under 18 18岁以下
- 18-25 18岁-25岁
- 26-30 26岁-30岁
- 31-40 31岁-40岁
- 41-50 41岁-50岁
- 51-60 51岁-60岁
- Over 60 60岁以上

Your educational qualification 您已获得的最高学历

- ○ Below junior high school 初中以下
- ○ Junior high school 初中毕业
- ○ Senior high school 高中毕业
- ○ Under graduate 学士学位
- ○ Post graduate 硕士学位
- ○ Beyond post graduate 博士学位

Your occupation 您目前的职业

- ○ Student 全日制学生
- ○ Production worker 生产人员
- ○ Marketing/Salesman 市场/公关人员
- ○ Customer service 客服人员
- ○ Logistics 行政/后勤人员
- ○ Human resources 人力资源
- ○ Financial/auditor 财务/审计人员
- ○ Civilian post 文职/办公人员
- ○ Technician 技术/研发人员
- ○ Manager 管理人员
- ○ Teacher 教师
- ○ Consultant 顾问/咨询
- ○ Specialist (e.g., accountant, lawyer, architect, journalist, etc.) 专业人士(如会计师、律师、建筑师、医护人员、记者等)
- ○ [ ] Other 其他

Your monthly personal income 您的月收入

○ Below 2000 低于2000元
○ 2001-3000 2001-3000元
○ 3001-5000 3001-5000元
○ 5001-7000 5001-7000元
○ Above 7000 高于7000元

Have you ever used Open Government Data before?
您是否使用过开放政府数据?

○ Yes 是
○ No 否

How did you know Open Government Data?
您是如何了解到开放政府数据的? (多选)

☐ Heard from friend/college/family members 从朋友、同事、家人那里听说
☐ Search online 从网络搜索获知
☐ Heard from the government 从政府获知
☐ Heard from news 从新闻报道获知
☐ [ ] Other 其他

What types of data have you used?
您使用了哪些主题的开放政府数据? (多选)

☐ Budget & Spend 政府财政

☐ Credit Records 信用记录
☐ Cultural activity 文化休闲
☐ Education 教育科技
☐ Environment Quality 环境质量
☐ Government Bid 政府招投标
☐ Health 医疗健康
☐ International Trade 国际贸易
☐ Local Statistics 区域统计
☐ Maps 地理地图
☐ Policies & Legislation 法律政策
☐ Public Safety 公共安全
☐ Registration 注册登记
☐ Transportation 道路交通
☐ Weather 气候气象
☐ ______ Other 其他

On average, how often do you use Open Government Data?
您使用开放政府数据的平均频率是:

○ Used once only 仅使用过一次
○ More than once but rarely use 多于一次但极少使用
○ Daily 平均每天一次
○ Weekly 平均每周一次
○ Monthly 平均每月一次
○ ______ Other 其他

Have you ever used an Open Government Data portal to obtain data before?
您是否使用过开放政府数据平台获取数据?

- ○ Yes 是
- ○ No 否

How did you know Open Government Data portal?
您是如何了解到开放政府数据平台的？（多选）

- ☐ Heard from friend/college/family members 从朋友、同时、家人那里听说
- ☐ Search online 从网络搜索获知
- ☐ Heard from the government 从政府获知
- ☐ Heard from news 从新闻报道获知
- ☐ ______ Other 其他

What types of data have you obtained from the portal?
您使用开放政府数据平台获取了哪些主题的数据？（多选）

- ☐ Budget & Spend 政府财政
- ☐ Credit Records 信用记录
- ☐ Cultural activity 文化休闲
- ☐ Education 教育科技
- ☐ Environment Quality 环境质量
- ☐ Government Bid 政府招投标

☐ Health 医疗健康
☐ International Trade 国际贸易
☐ Local Statistics 区域统计
☐ Maps 地理地图
☐ Policies & Legislation 法律政策
☐ Public Safety 公共安全
☐ Registration 注册登记
☐ Transportation 道路交通
☐ Weather 气候气象
☐ [ ] Other 其他

On average, how often do you use Open Government Data Portal?
您使用开放政府数据平台的平均频率是:

○ Used once only 仅使用过一次
○ More than once but rarely use 多于一次但极少使用
○ Daily 平均每天一次
○ Weekly 平均每周一次
○ Monthly 平均每月一次
○ [ ] Other 其他

You used Open Government Data for:
您使用开放政府数据的目的为：（多选）

☐ Daily life 日常生活

- ☐ Software development 软件开发
- ☐ Scientific research 科学研究
- ☐ News report 新闻报道
- ☐ Business decision 经营决策
- ☐ Right to know and superintendence 监督知情
- ☐ [ ] Other 其他

Do you play computer games?
您是否玩儿电脑游戏?

- ○ Yes 是
- ○ No 否

On average, how many hours of computer games do you play in one week?
您平均每周玩儿多少小时的电脑游戏?

	1	20	38	57	75	94	112	131	149	168
Hours of computer games per week 平均每周电脑游戏时长	○									

Here are a number of personality traits that may or may not apply to you. Please select on a scale of 1-7 the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.
以下是一些性格特质，它们可能适用于您，也可能并不适用。请选择您多大程度上同意或者不同意这些陈述。您应该选择每一对性格描述在何种程度上适用于您，即使那一对词语中的一个描述比另一个更贴切。

I see myself as: 我认为我是:

	Disagree strongly 完全不同意	Disagree moderately 大致不同意	Disagree a little 有点不同意	Neither agree nor disagree 不清楚	Agree a little 有点同意	Agree moderately 大致同意	Agree strong 完全同意
Extroverted, enthusiastic 外向的，有热情的	○	○	○	○	○	○	○
Calm, emotionally stable 平静的，情绪稳定的	○	○	○	○	○	○	○
Reserved, quiet 内向的，安静的	○	○	○	○	○	○	○

	Disagree strongly 完全不同意	Disagree moderately 大致不同意	Disagree a little 有点不同意	Neither agree nor disagree 不清楚	Agree a little 有点同意	Agree moderately 大致同意	Agree strong 完全同意
Open to new experiences, complex 接受新经验的，复杂型的	○	○	○	○	○	○	○
Dependable, self-disciplined 可靠的，自律的	○	○	○	○	○	○	○
Conventional, uncreative 行为典型的，缺乏创造性的	○	○	○	○	○	○	○
Anxious, easily upset 焦虑的，容易烦乱的	○	○	○	○	○	○	○
Disorganized, careless 缺乏组织的，粗心大意的	○	○	○	○	○	○	○
Critical, quarrelsome 善批评的，好争论的	○	○	○	○	○	○	○
Sympathetic, warm 有同情心的，温暖的	○	○	○	○	○	○	○

Please rank your attitudes towards the following statements:
请选择您多大程度上同意以下这些观点:

	Disagree strongly 完全不同意	Disagree moderately 大致不同意	Disagree a little 有点不同意	Agree a little 有点同意	Agree moderately 大致同意	Agree strongly 完全同意
I usually trust people until they give me a reason not to trust them. 我通常在对方给我不相信他/她的理由前，选择相信对方。	○	○	○	○	○	○
My typical approach is to trust new acquaintances until they prove I should not trust them. 我通常在见到一个陌生人时选择信任对方，除非他/她表现出我不应信任他/她。	○	○	○	○	○	○
I generally give people the benefit of the doubt when I first meet them. 我通常在第一次见到对方时，尽量把他/她往好处想。	○	○	○	○	○	○

Survey 2-1a: Technology Acceptance Model Questionnaire

Please select to what degree do you think the importance of each function? 请选择您认为以下功能的重要程度:

	Extremely important 极度重要	Very important 非常重要	Moderately important 中度重要	Slightly important 有点重要	Not at all important 完全不重要
Download without registration 无需注册即可下载	○	○	○	○	○
Users' Feedback 用户反馈	○	○	○	○	○

	Extremely important 极度重要	Very important 非常重要	Moderately important 中度重要	Slightly important 有点重要	Not at all important 完全不重要
Different formats to choose 多种数据格式	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Browse 浏览	○	○	○	○	○
Helping Function 帮助功能	○	○	○	○	○
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○
Meta data 元数据信息	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○
Filtering on search results 检索结果筛选	○	○	○	○	○

Instructions to search-help

You are now going to use a simulated OGD portal of some place in China. Five questions are given to you. You could find answers to these questions by using the open data on this portal. If you cannot find related data on the portal to answer these questions, please submit a data request on the portal. You are expected to spend around 8 minues for these questions.

A minimum timer has been set to explore the portal and answer these questions. After that, you are allowed to proceed to the next part of the survey. If you need to take a rest, please do so before going on for the following questions.

您现在将会使用一个模拟中国某地的开放政府数据平台。在问卷下一页，您将会得到五个问题。您可以通过查找数据平台上的开放数据，找到这些问题的答案。如果您无法在数据平台上检索到相关数据，请在数据平台上提交一个数据申请。您需要花费大约8分钟时间来查找问题的答案。

我们设置了倒计时。**倒计时结束后**，您才可以提交答案并前进到后续问卷内容。如果您需要短暂的休息，请在此时进行。休息过后再进行后续的答题。

Are you ready to use the portal?
您准备好使用开放政府数据平台了吗?

○ Yes 准备好了

Please click the following link before moving forward:
请先点击以下链接，再单击下一步：

访问开放政府数据平台

Questions to search-a

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

0459

Please find through "FAQ" in "Help" to see if the following statement is true or not:

Fees will be claimed for business use of the data on the portal.

请通过"帮助"项下的"常见问题"，查找以下说法是否正确：

将本平台数据用于营业用途会被政府收取一定的费用。

- True正确
- False不正确

How many kindergartens are there in this area?
本区域示范性幼儿园共有多少个?

- 39
- 59
- 79
- 无法找到相关信息并提交了数据申请

Please find through "FAQ" in "Help" to see if the following statement is true or not:
It is legal to resale the data downloaded from this portal.
请通过"帮助"项下的"常见问题"，查找以下说法是否正确:
可以转售通过本平台获取的数据。

- True正确
- False不正确

What's the top first cause for death in 2016?
2016年人口死因第一位的是?

- 呼吸系病
- 消化系病
- 循环系病
- 无法找到相关信息并提交了数据申请

How many swimming pools are there in this area?
本区域游泳池共有多少个?

- ○ 14
- ○ 24
- ○ 34
- ○ 无法找到相关信息并提交了数据申请

Survey2-2a: Technology Acceptance Model Questionnaire

Please select to what degree do you think the ease of use of each function? 请选择您认为以下功能的易用性:

	Extremely easy 非常易用	Somewhat easy 比较易用	Neither easy nor difficult 不易用也不难用	Somewhat difficult 有点难用	Extremely difficult 非常难用
Different formats to choose 多种数据格式	○	○	○	○	○
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Filtering on search results 检索结果筛选	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○

	Extremely easy 非常易用	Somewhat easy 比较易用	Neither easy nor difficult 不易用也不难用	Somewhat difficult 有点难用	Extremely difficult 非常难用
Users' Feedback 用户反馈	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Helping Function 帮助功能	○	○	○	○	○
Browse 浏览	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○
Download without registration 无需注册即可下载	○	○	○	○	○
Metadata 元数据信息	○	○	○	○	○

Please select to what degree do you think the usefulness of each function? 请选择您认为以下功能的有用性:

	Extremely useful 极度有用	Very useful 非常有用	Moderately useful 中度有用	Slightly useful 稍微有用	Not at all useful 完全没用
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Users' Feedback 用户反馈	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○

	Extremely useful 极度有用	Very useful 非常有用	Moderately useful 中度有用	Slightly useful 稍微有用	Not at all useful 完全没用
Browse 浏览	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Filtering on search results 检索结果筛选	○	○	○	○	○
Different formats to choose 多种数据格式	○	○	○	○	○
Metadata 元数据信息	○	○	○	○	○
Helping Function 帮助功能	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Download without registration 无需注册即可下载	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○

Survey 2-1b: Technology Acceptance Model Questionnaire-Avatar

Please select to what degree do you think the importance of each function? 请选择您认为以下功能的重要程度:

	Extremely important 极度重要	Very important 非常重要	Moderately important 中度重要	Slightly important 有点重要	Not at all important 完全不重要

	Extremely important 极度重要	Very important 非常重要	Moderately important 中度重要	Slightly important 有点重要	Not at all important 完全不重要
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Users' Feedback 用户反馈	○	○	○	○	○
Avatar 虚拟智能助理	○	○	○	○	○
Metadata 元数据信息	○	○	○	○	○
Browse 浏览	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Download without registration 无需注册即可下载	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Different formats to choose 多种数据格式	○	○	○	○	○
Filtering on search results 检索结果筛选	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○

Instructions to search-Avatar

You are now going to use a simulated OGD portal of some place in China. Five questions are given to you. You could find answers to these questions by using the open data on this portal. If you cannot find related data on the portal to answer these questions, please submit a data request on the portal. You are expected to spend around 8 minutes for these questions. Because the website has an avatar to help you and she will contact with you by speaking, **please make sure you have weared headphones or turned on your computer sound**. Thank you for your cooperation!

A minimum timer has been set to explore the portal and answer these questions. After that, you are allowed to proceed to the next part of the survey. If you need to take a rest, please do so before going on for the following questions.

您现在将会使用一个模拟中国某地的开放政府数据平台。在问卷下一页，您将会收到五个问题，您可以通过查找数据平台上的开放数据，找到这些问题的答案。如果您无法在数据平台上检索到相关数据，请在数据平台上提交一个数据申请。您需要花费大约8分钟时间来查找问题的答案。由于本网站搭载了一个虚拟智能助理，她将会通过语音与你交谈，**请确保你已佩戴耳机或电脑音量已经开启**，谢谢合作!

我们设置了倒计时。**倒计时结束后**，您才可以提交答案并

前进到后续问卷内容。如果您需要短暂的休息，请在此时进行。休息过后再进行后续的答题。

Are you ready to use the portal?
您准备好使用开放政府数据平台了吗?

◯ Yes 准备好了

Please click the following link before moving forward:
请先点击以下链接，再单击下一步：

访问开放政府数据平台

Have you clicked the link in the previous page?
请问你点击上一页的链接了吗?

If yes, please go on.
如果已经点击了，请继续下一步。

If not, please click the back button to the previous page.
如果没有，请返回上一步并点击链接。

Questions to search-bAvatar

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

05 00

Please find through "Ask the agent" in "Help" to see if the following statement is true or not:
Fees will be claimed for business use of the data on the portal.
请通过“帮助”项下的“询问助手”，查找以下说法是否正确：
将本平台数据用于营业用途会被政府收取一定的费用。

- True正确
- False不正确

How many kindergartens are there in this area?
本区域示范性幼儿园共有多少个？

- 39
- 59
- 79

○ 无法找到相关信息并提交了数据申请

Please find through "Ask the agent" in "Help" to see if the following statement is true or not:
It is legal to resale the data downloaded from this portal.
请通过"帮助"项下的"询问助手"，查找以下说法是否正确：
可以转售通过本平台获取的数据。

○ True正确
○ False不正确

What's the top first cause for death in 2016?
2016年人口死因第一位的是?

○ 呼吸系病
○ 消化系病
○ 循环系病
○ 无法找到相关信息并提交了数据申请

How many swimming pools are there in this area?
本区域游泳池共有多少个?

○ 14
○ 24
○ 34

○ 无法找到相关信息并提交了数据申请

Survey2-2b: Technology Acceptance Model Questionnaire-Avatar

Please select to what degree do you think the ease of use of each function?
请选择您认为以下功能的易用性:

	Extremely easy 非常易用	Somewhat easy 比较易用	Neither easy nor difficult 不易用也不难用	Somewhat difficult 有点难用	Extremely difficult 非常难用
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Browse 浏览	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Users' Feedback 用户反馈	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Different formats to choose 多种数据格式	○	○	○	○	○
Avatar 虚拟智能助理	○	○	○	○	○
Filtering on search results 检索结果筛选	○	○	○	○	○

	Extremely easy 非常易用	Somewhat easy 比较易用	Neither easy nor difficult 不易用也不难用	Somewhat difficult 有点难用	Extremely difficult 非常难用
Download without registration 无需注册即可下载	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○
Metadata 元数据信息	○	○	○	○	○

Please select to what degree do you think the usefulness of each function?
请选择您认为以下功能的有用性:

	Extremely useful 极度有用	Very useful 非常有用	Moderately useful 中度有用	Slightly useful 稍微有用	Not at all useful 完全没用
Avatar 虚拟智能助理	○	○	○	○	○
Visualize for quick viewing 可视化预览	○	○	○	○	○
Ranking 排行榜	○	○	○	○	○
Request new datasets 申请数据	○	○	○	○	○
Download without registration 无需注册即可下载	○	○	○	○	○
Keyword Search 关键词检索	○	○	○	○	○
Browse 浏览	○	○	○	○	○

	Extremely useful 极度有用	Very useful 非常有用	Moderately useful 中度有用	Slightly useful 稍微有用	Not at all useful 完全没用
Filtering on search results 检索结果筛选	○	○	○	○	○
Metadata 元数据信息	○	○	○	○	○
Open, machine-processable format 开放且可机读格式	○	○	○	○	○
Different formats to choose 多种数据格式	○	○	○	○	○
Users' Feedback 用户反馈	○	○	○	○	○

Survey3-1: Diffusion of Innovation Questions

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

Open Government Data Portal is more useful than other portals for data that I have used.
开放政府数据平台优于其他我使用过的数据平台。

○ ○ ○ ○ ○ ○

Open Government Data Portal offered me real advantages over the way I usually get open data online.
相比我平时在线获取开放数据的方式，开放政府数据平台提供给我更多优势。

○ ○ ○ ○ ○ ○

I had more control over open data by using open government data portal.
通过使用开放政府数据平台，我能够更好地掌控开放数据。

○ ○ ○ ○ ○ ○

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

Open Government Data Portal is better than other methods for getting government data (like going to government departments, use search engines).
开放政府数据平台优于其他获取政府数据的方式（如前往政府部门，使用搜索引擎等）。

○ ○ ○ ○ ○ ○

I found the data I need more quickly and easily because of using Open Government Data Portal.
因为使用开放政府数据平台，我可以更快更容易地找到我需要的数据。

○ ○ ○ ○ ○ ○

Using Open Government Data Portal made getting government data a better experience than I would have otherwise.
相比不使用开放政府数据平台，使用该平台让获取政府数据成为一种更好的体验。

○ ○ ○ ○ ○ ○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

The name "Open Government Data Portal" made we want to use the portal.
"开放政府数据平台"这一名称让我愿意使用这一平台。

○ ○ ○ ○ ○ ○

I think other portals should be built to gather open data like open government data portals.
我认为应该仿照开放政府数据平台，建设其他汇集开放数据的平台。

○ ○ ○ ○ ○ ○

Open Government Data Portal fit right into the way I like to find open data online.
开放政府数据平台非常符合我所喜爱的在线获取开放数据的方式。

○ ○ ○ ○ ○ ○

Open Government Data Portal helped me to learn more about technology while also learning about open government data.
开放政府数据平台不仅帮助我更加了解技术，也让我了解了开放政府数据。

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
○	○	○	○	○	○

Using Open Government Data Portal made getting access to the government data easier to me.
对我来说，使用开放政府数据平台让获取政府数据更为容易。

○ ○ ○ ○ ○ ○

Open Government Data Portal helped me find the data I need more easily while also showed me lots of data offered by the government.
开放政府数据平台不仅帮助我更简便地找到我所需的数据，同时向我展示了很多政府提供的数据。

○ ○ ○ ○ ○ ○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I had no difficulty in using the functions of the Open Government Data Portal.
我可以轻松地使用开放政府数据平台的各项功能。

○ ○ ○ ○ ○ ○

I had no difficulty using Open Government Data Portal online.
我可以轻松地在线使用开放政府数据平台。

○ ○ ○ ○ ○ ○

I had no difficulty understanding how to get around in Open Government Data Portal.
我可以轻松了解如何使用开放政府数据平台。

○ ○ ○ ○ ○ ○

I had no difficulty controlling the functions of the Open Government Data Portal.
我可以轻松地控制开放政府数据平台的各项功能。

○ ○ ○ ○ ○ ○

I had no difficulty understanding the data on Open Government Data Portal.
我可以轻松理解开放政府数据平台上的数据。

○ ○ ○ ○ ○ ○

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I had no difficulty understanding how Open Government Data Portal technically worked.
我可以轻松了解开放政府数据平台是如何运作的。

○ ○ ○ ○ ○ ○

I had no difficulty finding the data that I wanted.
我可以轻松找到我所需要的数据。

○ ○ ○ ○ ○ ○

I had no difficulty understanding what Open Government Data is.
我可以轻松理解什么是开放政府数据。

○ ○ ○ ○ ○ ○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法:

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I am more likely to want to use Open Government Data Portal because of being part of this pilot test.
由于参加这项实验，我因而更乐于使用开放政府数据平台。

○ ○ ○ ○ ○ ○

I really won't lose much by trying Open Government Data Portal, even if I don't like it.
尝试使用开放政府数据平台并不会给我带来多少损失，哪怕我不喜欢它。

○ ○ ○ ○ ○ ○

Being able to try out Open Government Data Portal was important in my deciding whether or not to use it.
能够试用开放政府数据平台，对于我决定是否使用它非常重要。

○ ○ ○ ○ ○ ○

I like being able to try out Open Government Data Portal before deciding whether I like it or not.
我愿意在决定是否喜欢开放政府数据平台之前，有机会先尝试使用它。

○ ○ ○ ○ ○ ○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法:

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I would have no difficulty in telling others finding for open government data how open government data portal improved it. 我可以毫无困难地告诉其他正在搜寻开放政府数据的人，开放政府数据平台是如何改善了搜索开放政府数据这一体验的。	○	○	○	○	○	○
I would have no difficulty in telling friends what open government data portal is like. 我可以毫无困难地告诉朋友开放政府数据平台是什么样子的。	○	○	○	○	○	○
People can tell that I know more about open government data since I have used Open Government Data Portal. 人们可以看出我因为使用开放政府数据平台而更了解开放政府数据。	○	○	○	○	○	○

Survey3-2: Diffusion of Innovation Questions - Avatar

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法:

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
An avatar offered me real advantages over traditional helping pages I usually use. 虚拟智能助理相比我平时使用的传统帮助界面，能够给我提供更多优势。	○	○	○	○	○	○
I used the portal more quickly and easily because of interacting with an avatar. 由于与虚拟智能助理交流，我可以更快更容易地使用数据平台。	○	○	○	○	○	○
I had more fun using the portal because of interacting with an avatar. 由于与虚拟智能助理交流，我可以更有意思地使用数据平台。	○	○	○	○	○	○
An avatar is more interesting than helping pages I have used on other portals. 虚拟智能助理比我在其他平台使用过的帮助界面更有趣。	○	○	○	○	○	○
Interacting with an avatar made using the portal a better experience than I would have otherwise. 相比没有虚拟智能助理，与她交流让平台的使用体验更好。	○	○	○	○	○	○
An avatar is better than helping pages for assisting me using the portal. 在帮助我使用数据平台上，虚拟智能助理优于帮助界面。	○	○	○	○	○	○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
The Avatar helped me find data more quickly and easily. 虚拟智能助理帮助我更快更容易地找到数据。	○	○	○	○	○	○

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
Interacting with the avatar make using the portal easier to me. 对我来说，与虚拟智能助理交流让使用数据平台变得更加容易。	○	○	○	○	○	○
An Avatar fits right into the way I like to have assistance. 虚拟智能助理非常符合我所希望的帮助方式。	○	○	○	○	○	○
I think other websites should have an avatar like this. 我认为其他网站也应该有一个类似这样的虚拟智能助理。	○	○	○	○	○	○
The name "Xiao Zheng" made me want to interact with it. 她的名字"小政"让我乐于与她交流。	○	○	○	○	○	○
The Avatar made my experience happier and relax. 虚拟智能助理让我的使用体验更加愉快而轻松。	○	○	○	○	○	○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I had no difficulty understanding how the avatar technically worked. 我可以轻松了解虚拟智能助理是如何工作的。	○	○	○	○	○	○
I had no difficulty getting the information that I wanted with the avatar. 因为有虚拟智能助理，我可以轻松获取我所需的信息。	○	○	○	○	○	○
I had no difficulty understanding how to interact with the avatar. 我可以轻松了解如何与虚拟智能助理交流。						

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
	○	○	○	○	○	○
I had no difficulty interacting with the avatar on a portal. 我可以轻松地与数据平台的虚拟智能助理交流。	○	○	○	○	○	○
I had no difficulty understanding the information given by the avatar. 我可以轻松地理解虚拟智能助理提供的信息。	○	○	○	○	○	○
I had no difficulty in interacting with the avatar. 我可以无困难地与一个虚拟智能助理交流。	○	○	○	○	○	○
I had no difficulty controlling the interaction with the avatar. 我可以轻松地控制与虚拟智能助理的交流。	○	○	○	○	○	○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I like being able to try out interacting with the avatar before deciding whether I like it or not. 我愿意在决定是否喜欢一个虚拟智能助理前，先尝试与他进行交流。	○	○	○	○	○	○
I am more likely to want to interact with the avatar because of being part of this pilot test. 由于参加了这项实验，我因而更乐于与虚拟智能助理交流。	○	○	○	○	○	○
Being able to try out the avatar was important in my decision to interact with it. 能够尝试使用虚拟智能助理对于我决定是否与她交流非常重要。	○	○	○	○	○	○

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I really won't lose much by trying the avatar, even if I don't like it.
尝试使用虚拟智能助理并不会给我带来多少损失，哪怕我并不喜欢她。

○ ○ ○ ○ ○ ○

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I would have no difficulty in telling others who using open government data portal how the avatar improved it.
我可以毫无困难地告诉其他使用开放政府数据平台的人，虚拟智能助理是如何改善了这一使用体验的。

○ ○ ○ ○ ○ ○

I would have no difficulty in telling friends what the avatar is like.
我可以毫无困难地告诉朋友虚拟智能助理是什么样子的。

○ ○ ○ ○ ○ ○

People can tell that I know more about open government data portals since I have interacted with the avatar on it.
人们可以看出我因为与平台上的虚拟智能助理交流而更了解开放政府数据平台。

○ ○ ○ ○ ○ ○

Survey4: Trustworthiness

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
In general, open government data portal is a safe place to get open government data. 总体来说，开放政府数据平台是一个获取开放政府数据的安全途径。	○	○	○	○	○	○
I can always rely on the open government data portal in a situation needing open government data. 在需要开放政府数据时，我可以总是依赖开放政府数据平台。	○	○	○	○	○	○
Faced with a situation that required me to look for open government data, I would use open government data portal. 当我面临需要搜寻开放政府数据的情况时，我会使用开放政府数据平台。	○	○	○	○	○	○
I feel that I could count on open government data portal to help with finding open government data. 我觉得我可以依赖开放政府数据平台来帮我找到开放政府数据。	○	○	○	○	○	○
In my opinion, open government data portal is trustworthy. 在我看来，开放政府数据平台是值得信赖的。	○	○	○	○	○	○
I think I can trust the open government data on the portal. 我认为我可以信赖开放政府数据平台上的数据。	○	○	○	○	○	○
When I need to find open government data, I would feel comfortable depending on the open government data portal. 当我需要寻找开放政府数据时，依赖开放政府数据平台会让我感到舒适。	○	○	○	○	○	○
I think I can trust the government. 我认为我可以信赖政府。	○	○	○	○	○	○

Survey5a: Rapport Questions-Helping pages

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法：

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I would like to look at the helping pages. 我乐于看着帮助界面。	○	○	○	○	○	○
I had difficulty understanding helping pages. 我对于理解帮助界面存在困难。	○	○	○	○	○	○
The helping pages was weird. 帮助界面很奇怪。	○	○	○	○	○	○
I don't like the way helping pages looks. 我不喜欢帮助界面的外观。	○	○	○	○	○	○
I would like to have helping pages help me. 我愿意让帮助界面帮助我。	○	○	○	○	○	○
Helping pages would be poor problem solvers. 帮助界面是解决问题的外行。	○	○	○	○	○	○
I would like to use the helping pages. 我乐于使用帮助界面。	○	○	○	○	○	○
I would be able to engage with the helping pages. 我能够与帮助界面建立友好关系。	○	○	○	○	○	○
Interacting with helping pages was believable. 与帮助界面交流是可信的。	○	○	○	○	○	○

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I think helping pages and I understood each other.
我认为我和帮助界面能够互相理解。

○ ○ ○ ○ ○ ○

I liked the helping pages.
我喜欢帮助界面。

○ ○ ○ ○ ○ ○

I felt uncomfortable during the session.
在实验过程中我感到不舒适。

○ ○ ○ ○ ○ ○

Helping pages were warm and caring.
帮助界面热情而有爱心。

○ ○ ○ ○ ○ ○

The helping pages were distractive.
帮助界面是干扰的。

○ ○ ○ ○ ○ ○

I couldn't get anything accomplished with the helping pages.
我不能在帮助界面的帮助下完成任何任务。

○ ○ ○ ○ ○ ○

I would listen to the helping pages.
我愿意听从帮助界面。

○ ○ ○ ○ ○ ○

My attention was distracted by the helping pages during the session.
在实验过程中，我的注意力被帮助界面干扰了。

○ ○ ○ ○ ○ ○

I felt embarrassed during the session.
在实验过程中我感到尴尬。

○ ○ ○ ○ ○ ○

The helping pages were helpful.
帮助界面是有帮助的。

○ ○ ○ ○ ○ ○

I felt I had a connection with helping pages.
我感到我与帮助界面具有了一定的联系。

○ ○ ○ ○ ○ ○

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I think the helping pages and I established rapport. 我认为帮助界面和我建立了良好关系。	○	○	○	○	○	○
I would be comfortable to use helping pages. 我对于使用帮助界面感到舒适。	○	○	○	○	○	○

Survey5b: Rapport Questions-Avatar

Please select your opinion towards the following statements:
请选择您对于以下陈述的看法:

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I would like to have someone like the avatar help me. 我愿意让一个像虚拟智能助理的人帮助我。	○	○	○	○	○	○
I liked the avatar. 我喜欢虚拟智能助理。	○	○	○	○	○	○
Interacting with the avatar was believable. 与虚拟智能助理交流是可信的。	○	○	○	○	○	○
I felt I had a connection with the avatar. 我感到我与虚拟智能助理具有了一定的联系。	○	○	○	○	○	○
The avatar was distractive. 虚拟智能助理是干扰的。	○	○	○	○	○	○

Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意

I would listen to the avatar.
我愿意听从虚拟智能助理。

○	○	○	○	○	○

I would be able to engage with the avatar.
我能够与虚拟智能助理建立友好关系。

○	○	○	○	○	○

I would be comfortable to talk to an avatar.
我对于与虚拟智能助理对话感到舒适。

○	○	○	○	○	○

I had difficulty understanding the avatar.
我对于理解虚拟智能助理存在困难。

○	○	○	○	○	○

The avatar was weird.
虚拟智能助理很奇怪。

○	○	○	○	○	○

This avatar was warm and caring.
虚拟智能助理热情而有爱心。

○	○	○	○	○	○

The avatar would be a poor problem solver.
虚拟智能助理是解决问题的外行。

○	○	○	○	○	○

My attention was distracted by the avatar during the session.
在实验过程中，我的注意力被虚拟智能助理干扰了。

○	○	○	○	○	○

I would like to talk to the avatar.
我乐于跟虚拟智能助理对话。

○	○	○	○	○	○

I think the avatar and I understood each other.
我认为我和虚拟智能助理能够互相理解。

○	○	○	○	○	○

The avatar was helpful.
虚拟智能助理是有帮助的。

○	○	○	○	○	○

	Strongly agree 非常同意	Agree 同意	Somewhat agree 有点同意	Somewhat disagree 有点不同意	Disagree 不同意	Strongly disagree 非常不同意
I think the avatar and I established rapport. 我认为虚拟智能助理和我建立了良好关系。	○	○	○	○	○	○
I don't like the way the avatar looks. 我不喜欢虚拟智能助理的外观。	○	○	○	○	○	○
I felt embarrassed during the session. 在实验过程中我感到尴尬。	○	○	○	○	○	○
I felt uncomfortable during the session. 在实验过程中我感到不舒适。	○	○	○	○	○	○
I couldn't get anything accomplished with the avatar. 我不能在虚拟智能助理的帮助下完成任何任务。	○	○	○	○	○	○
I would like to virtually chat (typing) with the avatar. 我乐于与虚拟智能助理通过视觉文字（打字）交谈。	○	○	○	○	○	○

Powered by Qualtrics

APPENDIX E: ADVERTISEMENT FOR RECRUITMENT OF STAGE 4

Using open government data portal to find data directly from the government

You are invited to participate in a study on 'Using OGD portals'. The purpose of this study is to investigate the adoption of open government data (OGD) portals. The study is being conducted by, Ms. Di Wang, Department of Computing, di.wang18@students.mq.edu.au to meet the requirements of PhD under the supervision of Professor Deborah Richards, 61 (0)2 9850 9567, deborah.richards@mq.edu.au, Department of Computing, Faculty of Science.

Participation in this study is entirely voluntary. If you decide to participate, first you will complete a demographic questionnaire and a questionnaire about your adoption of OGD and OGD portals. Then you will use a simulated OGD portal to complete 5 tasks about finding certain data on the portal. You will be asked your opinion of OGD and OGD portals after you complete all the tasks. Finally, you will answer some questions about your experience using the portal. The duration of the study is expected to be around 40 minutes.

No personal details are collected. Your data will be anonymous. No individual will be identified in any publication of the results. Access to unidentified data may be made available to other researchers if they are interested.

If you would like to participate, please copy the following URL into your browser – (https://mqedu.qualtrics.com/jfe/form/SV_86A3D41H6MZcfWd).

We could not offer any financial reward for participation. However, it is a good chance to know about such kind of websites as an additional way to find information.