The Development of Open Government Data

Connecting Supply and Demand Through Portals

Edited by
Di Wang
Deborah Richards
Ayse Aysin Bilgin
Chuanfu Chen

The Development of Open Government Data

This page intentionally left blank

The Development of Open Government Data: Connecting Supply and Demand Through Portals

BY

DI WANG
Wuhan University, China

DEBORAH RICHARDS
Macquarie University, Australia

AYSE AYSIN BILGIN
Macquarie University, Australia

And

CHUANFU CHEN
Wuhan University, China

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2022

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-80262-316-1 (Print)
ISBN: 978-1-80262-315-4 (Online)
ISBN: 978-1-80262-317-8 (Epub)

ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

Table of Contents

List of Figures and Tables

About the Authors

Di Wang is a Postdoc Researcher from School of Information Management, Wuhan University. She received her PhD in Computing from Macquarie University and PhD in Information Resources Management from Wuhan University. Her research interests include big data analytics, open data policy, open data resources, and data services. She published in journals including *Government Information Quarterly* and *Library Hi Tech* and presented at conferences including JCDL, ICADL, iConference, etc. She has been awarded several research grants from the National Science Foundation of China, the National Social Science Foundation of China, and China Postdoctoral Science Foundation.

Deborah Richards is a Professor in the Department of Computing at Macquarie University. Following 20 years in the IT industry during which she completed a BBus (Comp and MIS) and MAppSc (InfoStudies), she completed a PhD in Artificial Intelligence on the Reuse of Knowledge at the University of New South Wales and joined academia in 1999. While she continues to work on solutions to assist decision-making and knowledge acquisition, for the past decade, her focus has been on intelligent systems, agent technologies, and virtual worlds to support human learning, health, and well-being. She has over 300 refereed publications and been awarded numerous research and teaching grants from the Australian Research Council and other research funders.

Ayse Aysin Bilgin is an Associate Professor, active in the statistical analysis and design of studies in health and statistics education. Her research in health developed to include adolescent health and recently studies related to breast cancer, MS, and Alzheimer's disease. She is one of the pioneers on applying machine learning to health data sets. She was awarded "Excellence in research: Five Future-shaping Priorities (Healthy People)" by Macquarie University. Ayse has won several learning and teaching awards for her outstanding contributions to student learning such as ALTC citation for "Outstanding Contributions to Student Learning." Her pedagogical research interests are focused on statistics education, such as learning approaches in statistics, impact of learning spaces to students' learning. She is the President of International Association for Statistical Education (IASE) (July 2021–July 2023).

Chuanfu Chen is a Distinguished Research Professor and Dean of the Graduate School, Wuhan University. From 2005 to 2012, he was Dean of the School of

Information Management of Wuhan University and was the Deputy President of Library Society of China from 2004 to 2012. He earned his MLS in 1986, and his PhD in Law in 2001, both from Wuhan University. His research focuses on library development, public access to information resources, big data governance, and copyright. He has published in many journals including *ASLIB Proceedings*, *College & Research Libraries*, *Electronic Library*, J*ournal of Library Science of China*, and *Scientometrics*, among others. He has also been awarded numerous research grants from the National Social Science Foundation and the Natural Science Foundation of China.

Abbreviations

AHP	Analytic Hierarchy Process
API	Application programming interface
CR	Consistency ratio
DOI	Diffusion of Innovation theory
FAQ	Frequently asked question
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 behavioral 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

This page intentionally left blank

Preface

Open government data (OGD) have developed rapidly in these years due to various benefits that can be derived through transparency and public access. However, researchers emphasize lack of use instead of lack of disclosure as a key problem in OGD's present development. Previous studies look into this issue either from the supply side, focusing on data quantity and quality, or from the demand side, focusing on factors that affect users' acceptance of OGD, but seldom consider both sides at the same time. This book makes a comparison of the supply and demand sides of OGD and explores possible directions for the future development of OGD portals based on the discovered mismatches lying between the two sides.

The core purpose of this book is to improve OGD utilization by balancing the supply side and demand side of OGD according to the demands of citizens through the development of OGD portals. To achieve this objective, four connected studies were designed and carried out. 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. Sequentially, with citizens as 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 based on Diffusion of Innovation (DOI) theory, using the data collected in the previous two studies. 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 case studies in these four parts were carried out in China.

This research finds that Chinese OGD portals are in an early stage of development. Citizens have limited awareness of OGD and OGD portals. Significant correlations are recognized among citizens, and their demands and utilization of OGD. Mismatches lie between the supply and demand sides of OGD. Following the proposed usability criteria for building an OGD portal could improve citizens' proper utilization of OGD. Future directions for developing OGD are identified.

The key contribution of this book to the present literature is the theoretical and practical understanding of OGD and its user, as well as proposing directions for OGD portals' future development in order to encourage citizens' OGD utilization.

This book originates from the PhD research study of Dr Di Wang in Macquarie University, Australia, under the supervision of Prof. Deborah Richards, Assoc. Prof. Ayse Aysin Bilgin, and Prof. Chuanfu Chen.

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. This book would not be completed without your support and advice.

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

I would like to acknowledge the contributions of my parents, Mr Lin Wang and Mrs Lifen Ma for your continuous loving support. I could not have completed this work without your encouragement.

Besides my supervisors and parents, 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 my research. I would also like to express my appreciation for the kind support of the School of Information Management, Wuhan University, that has allowed me to complete this book.

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

Di Wang
Wuhan, China, December 2021

This page intentionally left blank

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 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. At the same time, supported by digital media and Web 2.0 technologies (Ribeiro, 2017), great volumes of data are generated, collected, and stored because many 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 (Janssen, Charalabidis, & Zuiderwijk, 2012). 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), 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 (Zuiderwijk & Janssen, 2014). In this context, open data policies have been established in different countries, e.g., the United States' Open Government Initiative, the EU Commission's decision on the reuse of Commission documents, and P.R. China's policy for the promotion of the development of big data. Consequently, many open data initiatives have

The Development of Open Government Data, 1–15

doi:10.1108/978-1-80262-315-420221001

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 book, *open government data* (OGD).

1.1 Research Context

The definition by Open Knowledge International provides precise 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. OGD is commonly treated as a subset of open data and is simply government-related data that are made open to the public (Attard, Orlandi, Scerri, & Auer, 2015) with a possibility of redistribution in any form without any copyright restrictions (Chatfield & Reddick, 2017). 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, it can be concluded that OGD is the intersection of open data and PSI (Fig. 1). OGD not only contains datasets directly produced by the government, including those relating to budgets and spending, population and census, but also data which are collected and indirectly owned by public institutions, such as data relating to climate, public transportation, education.

Over the whole life cycle of OGD, the opened data 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). 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 (Bertot, Jaeger,

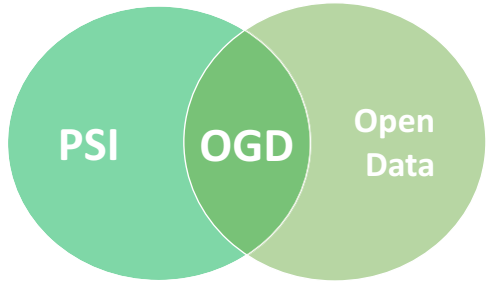

Fig. 1. Relationship between Open Data, OGD, and PSI.

& Grimes, 2010; Dawes, 2010) 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 (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. 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). 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" (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, and data.gov.sg of Singapore.

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 (Corrêa, Paula, Corrêa, & Silva, 2017; Yavuz & Welch, 2014). On the other hand, although more and more data are provided through OGD portals (Meijer, de Hoog, van Twist, van der Steen, & Scherpenisse, 2014), 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 (Ruijer, Grimmelikhuijsen, & Meijer, 2017), which leads to inadequate evidence about the real benefits from OGD initiatives (Open-DataBarometer, 2018). Citizens' lack of knowledge concerning the existence of OGD and OGD portals (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 (Janssen et al., 2012) and creating economic benefits (Willinsky, 2005). Therefore, at present, a disconnection between the supply side and the demand side of OGD is evident. 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 does not show obvious social

impact (OpenDataBarometer, 2018). 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).Many issues, challenges, and barriers can still be found in this field (Safarov, Meijer, & Grimmelikhuijsen, 2017; Wang et al., 2019).

All these issues exist in China, the largest developing country in the world. Encouraged by *Regulations of the People's Republic of China on Open Government Information*, OGD faced a fast development in China since 2007. Till the end of 2020, 20 province level and 156 city level OGD portals have been established and opened to the public in China. Tons of datasets related to education, public health, public safety, transportation, governmental budget, etc. have been uploaded to these portals. Despite the fast development of OGD in China, scholars have revealed insignificant impacts of the opened data on the whole society. Many Chinese citizens are not yet aware of OGD. More details of OGD development in China will be provided in Chapter 2.

1.2 Research Objectives

The main aims of disclosing government data are for its use, reuse, and distribution (Attard et al., 2015). With the development of OGD in the past few years, open data are 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, 2018). As a matter of fact, in contrast to the continual development on OGD's supply side, plenty of challenges can still be recognized when implementing OGD initiatives (Ribeiro, 2017), which resulted in the limited utilization of OGD (Safarov et al., 2017). Therefore, inadequate evidence of real benefits from OGD initiatives can be found, particularly in relation to social impact (OpenDataBarometer, 2018).

Since collaboration between the supply side and demand side of OGD is the premise for making OGD work for people, the ultimate goal of this study is 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. To reach this goal, several objectives were proposed 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 mismatches between the demand side and supply side of OGD.

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

The 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 and demand sides 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; Martin, 2014; 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). Therefore, this book focuses specifically on citizens of the demand side due to their role in the life cycle of OGD.

The objectives of this book for balancing the supply side and demand side are not only due to the inadequacy of studies connecting the two sides of OGD (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).

1.3 Research Questions

Open data are 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 both treated as flagship initiatives of OGD programs (Lourenço, 2015) and act as the bridge between the supply side and demand side of OGD.

Since the most critical challenge in present 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 book:

> 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 subquestions:

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 book (*O1*). Although many evaluations regarding country-level OGD portals have been carried out by different scholars and organizations (Attard et al., 2015; Thorsby, Stowers, Wolslegel, & Tumbuan, 2017), assessments of the potential and capabilities of OGD portals at local levels are limited (Chatfield & Reddick, 2017). Thus, for the first research question, the focus was set on analyzing the local development of OGD portals. *RQ1* is then divided into four subquestions:

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 stakeholders of the demand side, citizens, in relation to their demands and utilization habits of OGD, which reflected the second objective of this book (*O2*). *RQ2* is further divided into three subquestions:

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 findings of the studies to answer *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 (Meijer et al., 2014). To figure out the possible relationships between these two sides of OGD, especially the existing mismatches as stated in the third objective of this book (*O3*), *RQ3* is specifically raised for the study. In order to find possible methods for analyzing the relationship between two sides, *RQ3* is divided into two subquestions:

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 determining future development directions for OGD portals to encourage the utilization of OGD, which reflect the fourth objective of this book (*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 useable 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, *RQ4* is further divided into four subquestions:

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 answers to these four research questions, reported in Chapter 5, one further subquestion was added to focus on alternative help functions:

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

The last two research questions (*RQ5* and *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 are shown in Table 1.

1.4 Research Design and Roadmap

The whole research design of this book is based on the theory of the open data ecosystem and the conflict paradigm. The "Open data ecosystem" was firstly proposed by T. M. Harrison, Pardo, and Cook (2012), 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" (Harrison et al., 2012, p. 900). Open data are the core resource in the ecosystem (Heimstädt, Saunderson, & Heath, 2014).

Table 1. Structure of Objectives and Research Questions.

	Research Questions	
Objectives	**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 mismatches 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?