

How do institutional dimensions of open government data affect innovation? Evidence from research institutes in China

Aslib Journal of
Information
Management

Rui Mu and Xiaxia Zhao

*School of Public Administration and Policy, Dalian University of Technology,
Dalian, China*

Received 8 July 2023
Revised 16 January 2024
Accepted 24 March 2024

Abstract

Purpose – This study investigates the individual and binary (i.e. combined) effects of institutional dimensions of open government data (which include instructional, structural and accessible rules) on scientific research innovation, as well as the mediating roles that researchers' perceived data usefulness and data capability play in between.

Design/methodology/approach – Based on a sample of 1,092 respondents, this study uses partial least squares structural equation modeling (PLS-SEM) and polynomial regression with response surface analysis to evaluate the direct and indirect effects of individual and binary institutional dimensions on scientific research innovation.

Findings – The findings demonstrate that instructional, structural and restricted access data have a positive effect on scientific research innovation in the individual effect. While the binary effect of institutional dimensions produces varying degrees of scientific research innovation. Furthermore, this study discovers that the perceived usefulness and data capability of researchers differ in the mediating effect of institutional dimensions on scientific research innovation.

Originality/value – Theoretically, this study contributes new knowledge on the causal links between data publication institutions and innovation. Practically, the research findings offer government data managers timely suggestions on how to build up institutions to foster greater data usage.

Keywords Open government data, Institutional design, Innovation, Scientific research, China, Data interaction

Paper type Research paper

1. Introduction

Government data refers to the collection of information and datasets that are generated and maintained by governmental organizations. It includes various types of data such as demographic statistics, economic indicators, public health records, environmental measurements, etc. Open government data (OGD) refers to the government data that are published and made accessible to the public, in a format that is easily discoverable, downloadable and readable, allowing users to freely manipulate, analyze, reuse and distribute it for various purposes (Kassen, 2020). One of the primary goals of OGD is to promote public engagement and innovation. By making government data open, it enables

© Rui Mu and Xiaxia Zhao. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This work is funded by the National Natural Science Foundation of China (No: 72174036), Liaoning "Xingliao Talent Plan" and "Four Batch Talents" Program (No: XLYC2210021) and Liaoning Provincial Science Public Welfare Research Fund (Soft Science Research Program) in 2023 (No: 2023JH4/10600015).

Conflict of interest disclosure: The authors claim that there is no conflict of interest.



Aslib Journal of Information
Management
Emerald Publishing Limited
2050-3806
DOI 10.1108/AJIM-07-2023-0243

citizens, businesses, researchers and developers to analyze and utilize the data to conduct research and create innovative ideas, as well as improve public services and hold governments accountable (Zuiderwijk *et al.*, 2014). However, substantial studies on OGD indicate that the usage of OGD is lagged; even though a huge number of datasets are available, only a small number are actively used (Zuiderwijk and de Reuver, 2021). And the OGD-driven innovation only exists sporadically in private organizations and government-led hackathons (Mu and Wang, 2022; Sussha *et al.*, 2015).

Scholars attribute this low level of data usage to underdeveloped institutions for data publication (Kassen, 2018; Li and Chen, 2021). The key proponents of institutional theory argue that the adoption of OGD is influenced by existing institutional arrangements and that proper institutional design will contribute positively to the transition from closeness to openness (Altayar, 2018; Safarov, 2019). For instance, based on the system theory, Janssen *et al.* (2012) explain that the conventional system boundaries of governments are dissolving and becoming open when data is made public, which requires different steering institutions to manage data, motivate data usage and stimulate system feedback from external stakeholders. Besides, from the human-data interaction perspective, Victorelli *et al.* (2020) perform a comprehensive evaluation of the literature on human-data interaction and conclude that institutions on data representation, data interaction and data processing should be formulated in advance to facilitate users to understand and interact with the government data and to take the best use of the data.

Although scholars have reached the consensus that institutions are needed for publicizing data, they do not elaborate on what institutional dimensions should be considered and how these institutional dimensions will affect innovation? Do these dimensions work individually or together? Without answering these questions, people merely know institutional design is important but will never know what institutions should be at play and how they take effect. As a response, this article distills three institutional dimensions from the existing literature: *the instructional, structural and accessible dimensions*. And we examine how these institutional dimensions will affect *Scientific Research Innovation (SRI)*. SRI refers to the process of formulating novel research questions, creating and developing new methods and tools and exploring new knowledge and theories. We choose SRI as the outcome variable because researchers, among the various stakeholders, are one of the largest groups whose innovation activities may highly depend on government data and thus be affected by data presentation (Lassinantti *et al.*, 2019; Lnenicka *et al.*, 2022). For example, Li *et al.* (2019) demonstrates that tabular and graphical data presentations may affect how researchers understand an under-researched area and subsequently influence how they formulate research questions. Again, Harron *et al.* (2017) argue that effective data linkage environments can facilitate researchers to answer questions that require large sample sizes or detailed data on hard-to-reach populations and generate findings with a high level of external validity and applicability for policy making.

Therefore, our main research question reads as:

RQ1. How do the instructional, structural and accessible institutional dimensions, affect scientific research innovation in singular and binary forms?

Additionally, the literature suggests that data adoption and OGD-driven innovation may hinge on users' characteristics. For instance, Venkatesh and Davis (1996) point out that only when an individual believes that the data is relevant and will assist him or her in improving job performance will he or she adopt it. In addition, Janssen *et al.* (2012) put forward a myth of OGD and explain that only the individual who has the resources, expertise and capacities to collect and process data, will use the data. In line with these thoughts, we introduce two mediating variables, researchers' perceived data usefulness and data capability, the former measuring researchers' subjective perception of data usefulness and the latter measuring

researchers' objective abilities of data collection, process and interpretation. Thus, this article addresses an additional research question:

RQ2. Do researchers' perceived data usefulness and data capability play mediating effects between OGD institutions and SRI?

The empirical evidence of this study comes from the Chinese context. China is a typical case of a transition economy that performs well in scientific research innovation but institutional construction for OGD has just started. Therefore, to sustain and even boost scientific research innovation and to inject new data-based dynamics to innovation, the establishment of proper institutions for data publication are imperative. For other transition countries as well, a broad range of institutional changes and design work are needed to release data value and specially to promote effective data-driven innovation. However, the OGD institutions in western, developed countries might not fit the socio-economic contexts of the transition economies. Thus, China, as a representative case, need to consider what the institutional dimensions, in singular and binary forms, can set the stage for OGD usage and engender positive societal impact, such as scientific research innovation in our case.

This study has both theoretical contributions and practical implications. Theoretically, the study compensates for the theoretical gap between OGD institutions and SRI. In addition, it examines the role of researchers' perceived data usefulness and data capability in linking OGD institutions and SRI. In practice, the conclusions on the influence of OGD institutions on SRI provide a timely reminder for policymakers, administrators and practitioners to manage OGD.

2. Theoretical background

2.1 *Institutional dimensions of open government data*

Although institutions are defined in many ways, they are generally understood as formal rules (e.g. laws, regulations, policies, standards, or guidelines) and informal norms (e.g. cultures, customs, or traditions) that constrain and encourage individual behaviors and social, political and economic interactions (North, 1991). In this study, we only consider formal institutions because we aim to analyze how formal rules for data publication influence innovation; informal institutions are not considered because we do not attempt to explore the influence of cultures and norms on data usage. Three fundamental institutional dimensions that would affect data usage and influence scientific research innovation were extracted from the extant OGD literature (Altayar, 2018; Janssen *et al.*, 2012; Li and Chen, 2021; Machova *et al.*, 2018; Victorelli *et al.*, 2020): (1) the instructional dimension; (2) the structural dimension; and (3) the accessible dimension.

2.1.1 The instructional dimension: whether governments should provide instructional and security rules for data usage, or just publicizing data without any instructions or monitoring? One of the dominant debates about OGD is whether governments should publish data usage instructions and security rules along with the opened data. The current literature on data's instructional dimension does not offer an explicit answer. Institutionalism theorists assert that the failures of the human-data interaction are the consequence of inadequate rule design that limits the users' understanding of the opened data and the data context (Gonzalez-Zapata and Heeks, 2015; Niebel, 2021). Take this regard, Wang and Lo (2016) argue that data providers should be responsible for giving data instructions (e.g. data provenance, suggestions and requirements on how to use the data and recommended processing software) along with data publication. As a supplement, Bonina and Eaton (2020) recommend that, except for instructional rules, the arm's-length connection between data providers and data users should be governed by security rules to protect data use.

On the contrary, some scholars argue that too many data instructions and security rules will constrain data usage and exploitation. For example, [Martin et al. \(2019\)](#) points out that in some circumstances data instructions and security rules may hinder innovation because the innovators may (1) abandon the innovation ideas to focus on others that face fewer data regulations; (2) feel instruction burdensome and be discouraged from using the data and (3) minimize data usage and thus reduce their attempts to access data. Similarly, [Niebel \(2021\)](#) reports that data protection rules have a negative impact on innovation because innovators must adhere to the security standards, which will increase the costs of innovation.

2.1.2 The structural dimension: whether governments should provide highly structured data or just the data in loosely structured formats?. The second debate centers around the structure of data, that is, whether governments should supply highly structured data or just drop it and do not bother about the format. A specific response is not provided by the literature currently available on data's structural dimension. According to [Kitchin \(2014\)](#), structured data "are those that can be easily organized, stored and transferred in a defined data model, such as numbers/text set out in a table or relational database that have a consistent format." In contrast, unstructured data do not have a predefined data model or common identifiable structure, such as narrative text, audio, photo, or video. Some scholars argue that governments are better to publish structured datasets because such data are "machine-processable", meaning that calculus and algorithms can read, combine, process and analyze them easily, and computers can depict them using graphs and maps ([Attard et al., 2015](#); [Zuiderwijk and Janssen, 2014](#)). In the field of scientific research, [Figlio et al. \(2017\)](#) propose that the integration of structured datasets not only provides researchers with a full-sample data resource that reduces the generation of random errors during empirical analysis, but also offers new opportunities to reveal the full picture of event development under dynamic longitudinal panel data.

However, other scholars point out the problems of highly structured data. [Grossman and Pedahzur \(2020\)](#) argue that in the big data era, only 15–20% of existing data are structured data, while most available data are unstructured, including political speeches, pictures, video recordings, media broadcastings, policy/regulatory documents and massive blog posts generated by the wider public. And these unstructured data are growing much faster than structured data ([Zikopoulos et al., 2012](#)). For these data, governments are not advised to open them in a highly structured format, since data structuring process is a reductive process that inevitably entails the loss of details and context, and the structuring process may not keep the pace of data generation ([Grossman and Pedahzur, 2020](#)). These data are qualitative in nature and human-readable. Therefore, users can convert the unstructured data into structured data, depending on the users' own needs and purposes, through imposing a common structure upon the data by classification and codification.

2.1.3 The accessible dimension: whether governments should set up data access requirements or let public access data directly and effortlessly?. The third debate focuses on the data accessibility issue, that is, whether governments should set up certain data access restrictions (e.g. registration, application, or payment) or anyone is able to obtain the data without any additional efforts. There is no clear solution provided by the literature currently available on data's accessible dimension. Some scholars argue that complete and immediate disclosure of government data is needed, and the insurance of free public access is considered a significant foundation for open government and transparency ([Dawes, 2010](#); [Lourenco, 2015](#)). For instance, [The Open Government Working Group \(2007\)](#) emphasizes several important access principles for government data, including complete (all government data should be made available online), accessible (data can be obtained directly, not through navigating web forms or additional technical tools), non-discriminatory (government data do not require registration and application) and license-free (government data does not have any copyright and thus data access should be free of charge).

However, other scholars challenge the legitimacy of the above-mentioned claims. For example, [Janssen et al. \(2012\)](#) argue that a greater amount of opened data does not necessarily lead to better data usage and exploitation. Data publication without prior screening can result in information overload at the societal level and lead the public to less understanding and more confusion of government data. In line with Janssen, [Wang and Shepherd \(2020\)](#) evaluate the UK's OGD practice and point out that most published data hardly attracted public attention and the British government even halted the commenting and forum functions due to inactive citizens. In addition, [Janssen and van den Hoven \(2015\)](#) state that governments always need to consider the privacy and security issues when releasing data and thus will inevitably require users to register for data access and usage. In some circumstances, governments use the "disclosure upon application" strategy, meaning that, in order to avoid information overload and reduce privacy risks, users need to apply for permits/licenses to access certain needed data ([Lnenicka and Komarkova, 2019](#)).

In summary, we can see that the current literature on OGD has recognized the institutional aspects of data publication. However, disagreements still exist on how to design the rules, i.e. whether instructional rules should be presented along with data publication; whether rules should stipulate data structure; and whether accessibility should be restricted to a certain degree. Our study aims to fill the gap by investigating the effects of both sides of each institutional dimension on scientific research innovation and thus unpacking the logic of how the institutional dimensions of OGD influence data usage.

2.2 Scientific research innovation

Currently, there is not a unified term to describe the "newness" produced from scientific research activities. The most influential term is "scientific creativity" ([Simonton, 2003](#)), which means the capacity of researchers to conduct scientific studies that are novel, original, valuable and unexpected. This definition focuses on creative individuals and emphasizes the individuals' mental processes and cognitive operations that lead to scientific discoveries, but does not pay much attention to creative products. As a response, some scholars depart from the Schumpeterian tradition and use the term "research novelty" to describe creative products generated from scientific research ([Lee et al., 2015](#); [Schumpeter, 1934](#)). Here, creative products do not only generate from scratch but also from the unprecedented and distant combination of existing bits of knowledge ([Wang et al., 2017](#)). Thus, this definition emphasizes recombination of existing knowledge and does not capture the creative products that are made from new and fresh inputs of research data. In this study, we borrow in the innovation theory (the data-driven innovation literature in particular) which argues that innovative products can be designed not only through existing knowledge but also through new analytical and productional materials such as open data ([Jetzek et al., 2014](#); [Rizk et al., 2022](#)). Therefore, we build upon the concepts of "scientific creativity" and "research novelty" and introduce the term "scientific research innovation" (SRI) as our outcome variable of OGD's institutional construction. It generally refers to the scientific research products that are "new" or contain "newness" in relative to the existing knowledge. The new research products can be novel research questions, inventions of research methods, instruments and tools and discoveries of new relationships between variables to reveal what we otherwise had not known or conceived ([Corley and Gioia, 2011](#); [Heinze et al., 2007](#)).

2.3 Perceived data usefulness and data capability

In this study, we propose and examine two mediating variables linking the institutional dimensions of OGD and scientific research innovation. One mediating variable is researchers' perceived data usefulness, which reflects the researchers' subjective perception of OGD. It measures the extent to which the researchers think that OGD would improve their

opportunities to generate SRI (Yoon and Kim, 2017). If the researchers believe OGD is useful and valuable, then they will put efforts into consciously collecting and analyzing the data, which lays the foundation for SRI (Weerakkody *et al.*, 2017). By contrast, the researchers are likely to avoid finding and using OGD if they think OGD is useless. As such, we propose that perceived data usefulness plays a mediating role between the institutional dimensions of OGD and SRI.

Another mediating variable is researchers' data capability, which reflects the researchers' objective aspect of OGD. It refers to the abilities or skills of the researchers to collect, translate, convert, analyze and exhibit data (Li *et al.*, 2019). When the researchers have sufficient data capability, they know what they need, why they need the data and how to create the interfaces and systems of data (Rizk *et al.*, 2022). That means, data capability can help the researchers understand and digest a large amount of data. Meanwhile, strong data capability facilitates researchers to explore and maximize the growth of data value by using proficiently appropriate methods and uncovering valuable information and patterns. Hence, if the researchers do not have the capability to use OGD, the new ideas, insights and designs with the data might be forgone and SRI becomes bleak (Jetzek *et al.*, 2014).

3. Methodology and data

3.1 Questionnaire development and variable operationalization

This study relied on the following procedures to develop the survey questionnaire. First, we created a set of initial items from the literature. Second, we invited 15 experts from different disciplines to check the wording of the item and received 11 suggestions for revision. The comments from these experts are listed in Table A1. Based on these suggestions, we redefined the scope of OGD, revised the description of item IRA3 and added self-report items in the questionnaire. Next, we employed the pilot study that re-distributed the questionnaires to 144 respondents to further examine the quality of the questionnaire. Then, we evaluated the discriminant validity and the convergent validity of the questionnaire, removed the problematic items and finally obtained the formal 24-item questionnaire. The detailed items and their sources for the questionnaire are presented in Table A2. In the questionnaire, we added researchers' gender, age and dependency of discipline on OGD as control variables. Furthermore, we regarded the level of research institute with which researchers are affiliated as an additional control variable (Lee *et al.*, 2015), because in China research institutes are ranked, for instance, into the "985" and "211" series. In our models, the first-level research institutes include the Chinese Academy of Science (CAS) and "Project 985" universities [1], the second level includes "Project 211" universities [2] and the third level includes all other universities. We measured the dependent and independent variables using a five-point Likert scale ranging from 1 (i.e. strongly disagree) to 5 (i.e. strongly agree). The control variables were measured by category.

3.2 Sample and data collection

From August to September 2021, this study collected data through Wenjuanxing (www.wjx.cn), a popular and most prominent online survey platform in China. The respondents were researchers who used OGD in their research, including Ph.D. students, postdoctoral fellows and research fellows. All participants were recruited by posting the survey recruitment information on the crowdsourcing platform, such as muchong.com, which contains a large number of active researchers from multiple disciplines. Finally, a total of 1,611 questionnaires were received, of which 1,092 valid surveys were employed for this study. Table 1 shows the respondents' demographic information. Among the 1,092 respondents, 56.96% are male and 43.04% are female. Most respondents are 21–30 years old, up to 82.23%. A majority of

Characteristics	Categories	Frequency	Percentage
Gender	Male	622	56.96
	Female	470	43.04
Age	21–30 years old	898	82.23
	31–40 years old	151	13.83
	41–50 years old	25	2.29
	≥51 years old	18	1.65
Research institute level	First-level	819	75.00
	Second level	87	7.97
	Third level	186	17.03
The dependency of discipline on OGD	1-strongly disagree	34	3.11
	2-disagree	109	9.98
	3-uncertain	373	34.16
	4-agree	413	37.82
	5-strongly agree	163	14.93

Source(s): Table by authors

Table 1.
Sample's demographic
information

respondents come from the first-level research institutes (75.00%), while 7.97% are from the second level and 17.03% are from the third level. Regarding the dependency of discipline on OGD, 52.75% of the respondents agree that their research disciplines rely on OGD, while 13.09% disagreed and 34.16% of respondents are uncertain.

3.3 Strategies for data analysis

First, partial least squares structural equation modeling (PLS-SEM) was used to examine the direct and indirect effects of individual institutional dimensions on SRI. We chose PLS-SEM to test the relationship because the PLS-SEM method has several advantages over other statistical methods used for structural equation modeling (Hair *et al.*, 2019). Most importantly, the PLS-SEM method is highly predictive and appropriate for research where the goal is theory development rather than theory testing. It is appropriate for our study because previous research has not extensively investigated the effects of the institutional aspects of OGD on innovation, and factor loadings and cross-loading external models allow the PLS-SEM method to predict and explain such underdeveloped causal relations. In addition, research models with complex latent variables that are measured by multi-layer constructs can be predicted using the PLS-SEM approach. We find that the three institutional dimensions are measured by complicated structures of items, making this method an appropriate analytical approach for our study. Furthermore, our study involves a formative variable, the unrestricted accessibility rule, which requires the PLS-SEM method because this method permits formative variables.

Second, polynomial regression with response surface analysis was applied to test the direct and indirect effects of binary institutional dimensions on SRI. Polynomial regression can provide more insights by evaluating the relationship between the interaction of two predictors on the dependent variable (Shanock *et al.*, 2010). This allows for capturing more nuanced and intricate patterns in the data. It thus fits our research purpose of examining the binary effects of the institutional dimensions on SRI. Furthermore, polynomial regression is often used in response surface analysis, which offers us with visualized, easy-to-interpret non-linear relationships between the binary institutional predictors and SRI. Moreover, the coefficients of the polynomial terms provide insights into the direction and magnitude of the relationships between the binary predictors and SRI. This enables us to identify the optimal conditions or settings for the institutional dimensions that maximize SRI.

3.4 Testing the quality of the research model

To assess the internal consistency and reliability of the questionnaire, Cronbach's alpha and composite reliability (CR) were used as indicators. Meanwhile, the convergent validity was assessed by the average variation extraction (AVE). Table 2 shows that all constructs' Cronbach's alpha coefficients are greater than 0.6, the CR of latent variables is greater than 0.7, and the AVE is greater than 0.5, indicating that all constructs in the research model are highly reliable and convergent (Chin *et al.*, 2003).

The discriminant validity of the square root of AVE was then tested using cross-loading analysis and the association of the square root of AVE with other components (Hair *et al.*, 2019). Cross-loading revealed that all factor loadings are greater than the suggested value of 0.70. Table 3 further shows that the squared root of all variables' AVE is greater than their correlation with other factors. As a result, there is good discriminant validity between variables. Then we calculated the variance inflation factor (VIF) to see if there was multicollinearity. If the value of VIF exceeds the 3.3 threshold, multicollinearity is a worry (Diamantopoulos and Winklhofer, 2001). The maximum value of VIF in our analysis is 2.44, indicating that multicollinearity is unlikely to be a problem.

We also investigated the possibility of common method bias. First, SPSS 23 was used to run Harman's single-factor test (Podsakoff *et al.*, 2003). The findings revealed that the first component only explains 26.46% of the variation, falling short of the 50% requirement. Second, SmartPLS 3.3.3 was used to run the unmeasured common latent technique, which involves adding a major construct of all variables in the research model, as described by Liang *et al.* (2007). We found that the average factor loading value in the substantively principal constructs is substantially greater than the common approach when we compared the average variance of each item in the substantively principal constructs and the common method. As a result, the typical procedure bias would not be a threat.

Constructs	Items	Factor loading	AVE	CR	Cronbach's α
Instructional dimension	ID1	0.846	0.708	0.879	0.795
	ID2	0.812			
	ID3	0.866			
Structural dimension-unstructured	UST1	0.812	0.567	0.796	0.650
	UST2	0.729			
	UST3	0.714			
Structural dimension-structured	ST1	0.713	0.566	0.796	0.622
	ST2	0.827			
	ST3	0.711			
Accessible dimension-restricted	RA1	0.811	0.654	0.849	0.746
	RA2	0.892			
	RA3	0.712			
Perceived data usefulness	PDU1	0.844	0.758	0.904	0.840
	PDU2	0.895			
	PDU3	0.871			
Data capability	DC1	0.787	0.630	0.872	0.803
	DC2	0.824			
	DC3	0.824			
	DC4	0.735			
Scientific research innovation	SRI1	0.825	0.711	0.908	0.865
	SRI2	0.851			
	SRI3	0.867			
	SRI4	0.830			

Table 2.
The results of
reliability and
convergent validity
analysis

Note(s): Accessible dimension-unrestricted were measured with one item

Source(s): Table by authors

Variables	1	2	3	4	5	6	7	8
1. Instructional dimension	<i>0.841</i>							
2. Structural dimension-unstructured	0.295	<i>0.753</i>						
3. Structural dimension-structured	0.271	0.426	<i>0.752</i>					
4. Accessible dimension-unrestricted	0.060	0.127	0.204	<i>1.000</i>				
5. Accessible dimension-restricted	0.146	0.257	0.112	-0.334	<i>0.809</i>			
6. Perceived data usefulness	0.325	0.149	0.150	0.029	0.146	<i>0.870</i>		
7. Data capability	0.394	0.148	0.224	0.078	0.158	0.545	<i>0.793</i>	
8. Scientific research innovation	0.451	0.254	0.246	0.040	0.198	0.490	0.480	<i>0.843</i>

Note(s): The italic data are the square root of the average variation extraction. And the left and the bottom values are the variable correlation coefficients

Source(s): Table by authors

Table 3.
The results of correlations and convergent validity analysis

4. Results and analyses

The results are presented in two parts. The first part shows how institutional dimensions affect SRI individually, as well as the roles of mediating variables in linking the institutional dimensions and SRI. The second part displays the binary effects of the institutional dimensions on SRI and the roles of mediating variables.

4.1 Testing the direct and indirect effects of individual institutional dimensions on SRI

Table 4 presents the results of the direct effects of individual institutional dimensions on SRI. As can be seen, significant effects of instructional rules on SRI emerge ($\beta = 0.337, p = 0.000$). That means, when governments provide instructions on data usage along with data publication, SRI is more likely to be produced. This result is consistent with the opinion of institutionalism theorists who argue that proper instructions on data usage will ease the interface between data and users and thus is advantageous to data exploitation (Mutambik et al., 2023).

At the structural dimension, the results show that both unstructured data ($\beta = 0.078, p = 0.023$) and structured data ($\beta = 0.064, p = 0.035$) have significant effects on SRI. This

Dimensions	Path coefficient	t-value	p-value	Conclusion
<i>Instructional dimension</i>				
Instructional → SRI	0.337	10.259	0.000	Supported
<i>Structural dimension</i>				
Unstructured → SRI	0.078	2.278	0.023	Supported
Structured → SRI	0.064	2.106	0.035	Supported
<i>Accessible dimension</i>				
Unrestricted → SRI	0.008	0.270	0.787	Not supported
Restricted → SRI	0.083	2.501	0.012	Supported
<i>Control variables</i>				
Gender → SRI	0.002	0.084	0.933	Not supported
Age → SRI	-0.002	0.093	0.926	Not supported
Research institute level → SRI	0.060	2.366	0.018	Supported
The dependency of discipline on OGD → SRI	0.282	9.292	0.000	Supported

Note(s): SRI = scientific research innovation

Source(s): Table by authors

Table 4.
The results of direct effects of individual institutional dimensions on SRI

finding is in line with what we argued in the theoretical framework: governments can publish both quantitative, machine-readable datasets in rows and columns and qualitative, human-readable texts, audios and videos. The reason is that structured and unstructured data have their respective advantages: structured data that is organized into a predefined consistent format is easier to store, search, retrieve and analyze and thus allows for efficient data processing and simplified data management; unstructured data can provide richer and diverse information, especially valuable contextual information and can offer researchers a more comprehensive and holistic view of the data and allowing for deeper analysis and understanding (Grossman and Pedahzur, 2020).

However, at the accessible dimension, the result only supports the positive relationship between restricted accessibility and SRI ($\beta = 0.083$, $p = 0.012$); the effect of unrestricted accessibility on SRI is not supported. This indicates that the use of restricted data is more beneficial for SRI. This logic may be interpreted by the fact that restricted data ensures that sensitive or confidential information is already processed and protected by governments, reducing researchers' risks of data misuse and unintended ethical consequences and increasing researchers' trust on the data. Moreover, restricted data might be of higher quality because it often undergoes more rigorous quality control measures within governments, ensuring its accuracy and reliability for researchers' usage (Meijer *et al.*, 2014).

In addition, among the control variables, the results show that research institute level and dependency of discipline on OGD are related to the dependent variable, while researchers' gender and age are not.

Apart from the direct effects, we also test the indirect effects of institutional dimensions on SRI through researchers' perceived data usefulness and data capability. As Table 5 shows, there is a "complementary" mediating effect of perceived data usefulness between instructional rules and SRI (95% confidence interval (CI) = 0.046–0.094). Meaning that instructional rules transfer their effects to SRI partly through perceived data usefulness. It suggests that data instructions can make researchers think the data is useful because they provide clear guidance on how to use and interpret the data effectively; when researchers perceive the data is useful or of high quality, they will be satisfied with the data and generate greater reuse intention (Wang *et al.*, 2023).

Besides, there is an "indirect-only" mediating effect of perceived data usefulness between restricted data accessibility and SRI (95% CI = 0.009–0.045). It means that perceived data usefulness functions as a necessary condition for restricted accessibility to influence SRI. The underlying logic would be that when restrictions exist in data accessibility, researchers would perceive higher data security, have greater trust on the data and will attach more value to the data (Bargh *et al.*, 2016). In that case, researchers will feel more confident to generate valuable new insights or knowledge.

Regarding researchers' data capability, it plays a partial intermediary role in linking instructional rules and SRI (95% CI = 0.040–0.088). It suggests that clear data instructions and guidance (e.g. how to use the data and suggested analytical tools) will enhance researchers' data capacity and further lead to innovative discoveries (Wilson and Cong, 2021; Li *et al.*, 2019).

However, it plays complete mediating effects between structured data and SRI (95% CI = 0.009–0.036), unrestricted data and SRI (95% CI = 0.003–0.029) and restricted data and SRI (95% CI = 0.010–0.042). These results indicate that the generation of SRI depends highly on researchers' data capability when the data used are in structured formats. And no matter whether there are access restrictions or not, it requires data capability for researchers to produce SRI. Figure 1 summarizes the direct and indirect effects of individual institutional dimensions on SRI.

Dimensions	Direct effect with mediators β (95% CI)	Indirect effect Perceived data usefulness β (95% CI)	Mediated?	Data capability β (95% CI)	Mediated?
<i>Instructional dimension</i> Instructional \rightarrow SRI	0.227*** [0.163–0.289]	0.069*** [0.046–0.094]	Complementary	0.064*** [0.040–0.088]	Complementary
<i>Structural dimension</i> Unstructured \rightarrow SRI	0.076* [0.016–0.138]	0.003 [–0.015 to 0.022]	No	–0.009 [–0.023 to 0.004]	No
Structured \rightarrow SRI	0.045 [–0.012 to 0.103]	0.011 [–0.006 to 0.028]	No	0.021** [0.009–0.036]	Indirect-only
<i>Accessible dimension</i> Unrestricted \rightarrow SRI	–0.006 [–0.057 to 0.047]	0.009 [–0.007 to 0.025]	No	0.015* [0.003–0.029]	Indirect-only
Restricted \rightarrow SRI	0.052 [–0.006 to 0.117]	0.025*** [0.009–0.045]	Indirect-only	0.024** [0.010–0.042]	Indirect-only

Note(s): SRI = scientific research innovation, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source(s): Table by authors

Table 5.
The results of indirect
effects of individual
institutional
dimensions on SRI

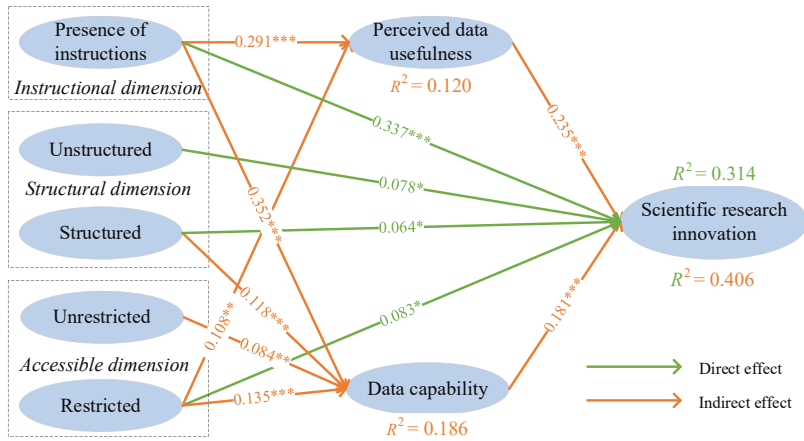


Figure 1. A summary of the direct and indirect effects of individual institutional dimensions on SRI

Note(s): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source(s): Figure by authors

4.2 Testing the direct and indirect effects of binary institutional dimensions on SRI

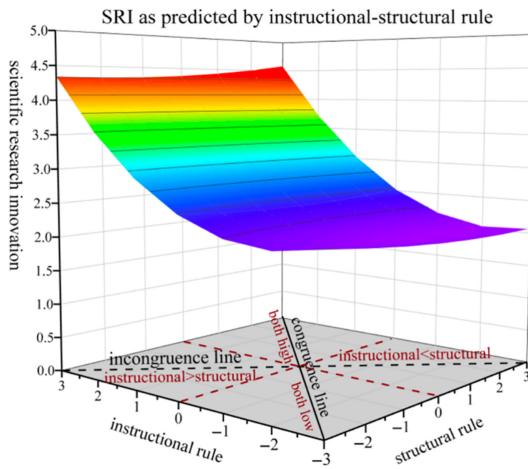
In our study, we construct three models of binary institutional dimensions. Table 6 reports the polynomial regressions results as well as the slopes and curvatures along the congruence line and incongruence line for the three models respectively. Here, congruence means two combined institutional dimensions exhibit concurrently high-high or low-low statuses, while incongruence measures the opposite statuses, high-low or low-high, of the combined institutional dimensions. In the regression models, we controlled researchers' gender, age, institute level and dependency of discipline on OGD. Additionally, when we combined two dimensions, we treated the other dimension as the control variable. Figures 2–4 illustrate the three-dimensional response surface based on the coefficients.

First, in Model 1 (instructional and structural rules), Figure 2b shows a positive significance (slope = 0.380, $p = 0.000$) along the congruence line. It also indicates that the performance of SRI is higher at the rear corner (high instructional rules and high structured data) than at the front corner (low instructional rules and low structured data) (also see Figure 2a). The incongruence

Dimensions	Effect	β	S.E.	t -value	p -value
Instructional – structural	Slope along instructional = structural	0.380	0.038	10.000	0.000
	Curvature on instructional = structural	0.111	0.049	2.265	0.025
	Slope along instructional = –structural	0.378	0.038	9.947	0.000
	Curvature on instructional = –structural	0.084	0.042	2.000	0.047
Instructional – accessible	Slope along instructional = accessible	0.468	0.037	12.649	0.000
	Curvature on instructional = accessible	0.018	0.039	0.462	0.645
	Slope along instructional = –accessible	0.282	0.042	6.714	0.000
	Curvature on instructional = –accessible	0.235	0.058	4.052	0.000
Structural – accessible	Slope along structural = accessible	0.089	0.039	2.282	0.023
	Curvature on structural = accessible	0.034	0.048	0.708	0.480
	Slope along structural = –accessible	–0.089	0.033	–2.697	0.008
	Curvature on structural = –accessible	0.035	0.035	1.000	0.318

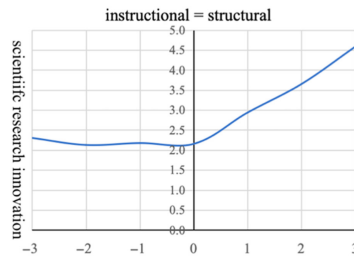
Table 6. The results of three polynomial regressions of binary institutional dimensions

Source(s): Table by authors

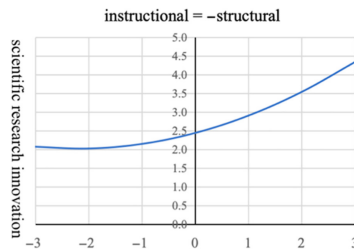


(a)

Source(s): Figure by authors

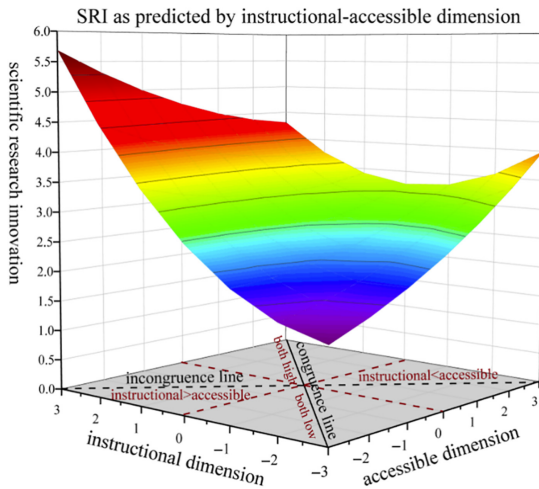


(b)



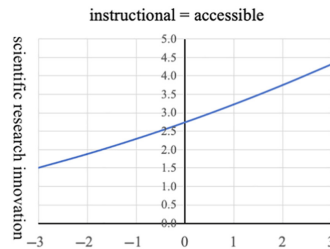
(c)

Figure 2.
Plotted results of the
combination of
instructional and
structural dimensions

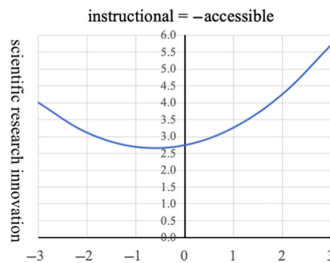


(a)

Source(s): Figure by authors



(b)



(c)

Figure 3.
Plotted results of the
combination of
instructional and
accessible dimensions

line also reports a positive significance (slope = 0.378, $p = 0.000$). Figure 2c shows that the performance of SRI is higher in the left corner (high instructional rules and low structured data) than in the right corner (low instructional rules and high structured data) (also see Figure 2a). When we compare the left corner along the incongruence line with the rear corner along the

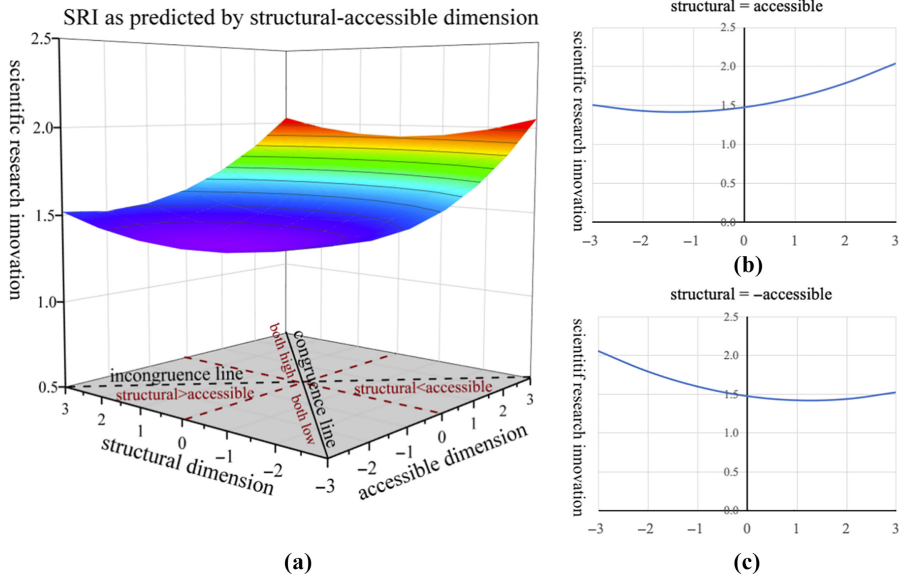


Figure 4. Plotted results of the combination of structural and accessible dimensions

Source(s): Figure by authors

congruence line, we find that the rear corner brings about a slightly higher performance of SRI, meaning that the most powerful condition in triggering SRI in model 1 is the combination of instructional and structured rules.

Second, in Model 2 (instructional and accessible rules), Figure 3b shows a positive significance (slope = 0.468, $p = 0.000$) along the congruence line. It also indicates that the performance of SRI is higher in the rear corner (high instructional rules and high accessibility) than in the front corner (low instructional rules and low accessibility) (also see Figure 3a). Along the incongruence line, a positive significance also appears (slope = 0.282, $p = 0.000$). Figure 3c indicates that the performance of SRI is higher at the left corner (high instructional rules and low accessibility) than at the right corner (low instructional rules and high accessibility) (also see Figure 3a). If we compare the left corner along the incongruence line with the rear corner along the congruence line, we find that the left corner brings about a higher level of SRI, meaning that the condition of high instructional rules and low accessibility is more beneficial for the generation of SRI.

Third, in Model 3 (structural and accessible rules), Figure 4b shows a positive significance (slope = 0.089, $p = 0.023$) along the congruence line. It also shows that the performance of SRI is higher in the rear corner (high structured data and high accessibility) than in the front corner (low structured data and low accessibility) (also see Figure 3a). Along the incongruence line, however, the result reports a negative significance (slope = -0.089 , $p = 0.008$). As Figure 4c shows, the performance of SRI is higher at the right corner (high accessibility and low structured data) than at the left corner (low accessibility and high structured data) (also see Figure 4a). However, this response surface does not show whether the rear corner condition or the right corner condition produces a higher level of SRI.

Apart from the direct effects of the binary institutional dimensions, we also tested the indirect effects of the combined variables through the two mediating variables. As Table 7 shows, both perceived data usefulness and data capability play partial mediating roles between three pairs of binary institutional dimensions and SRI.

5. Conclusions and discussions

5.1 Summary of findings

With a sample of 1,092 respondents in China, our empirical findings indicate that both the presence of data instructions and the presence of data access restrictions have positive impacts on SRI; however, it does not matter for SRI whether governments publish data in structured or unstructured formats. In addition, our findings also reveal that researchers' perceived data usefulness plays a partial mediating role between the instructional rules and SRI; however, it plays as a necessary mediating condition between restricted accessibility and SRI. This implies that only when researchers perceive government data as extremely important, data accessibility restrictions would not limit data usage for SRI. Regarding researchers' data capability, it plays a partial mediating role between the instructional rules and SRI; however, it plays a complete mediating role between structured data format and SRI. This implies that when data are presented in structured format, researchers need to depend on their data capability to interpret and process data and subsequently generate SRI.

Furthermore, we investigated the effects of binary institutional dimensions on SRI to gain a better understanding of the intricate interplay between the rules and their interactive impact on SRI. As a result of Section 4.2, we can conclude that when governments provide instructional rules in data publication, restricted accessible rules or unstructured rules are preferred to go together to facilitate SRI. A possible explanation might be that, when governments put forward instructional and regulatory rules on data usage (e.g. for privacy and abuse-avoidance concerns), data acquisition is usually non-free and set with registration requirements. And to avoid losing data details and context information, governments usually publish unstructured data and in doing so, data usage instructions are preferably accompanied with the unstructured data to direct or guide data classification and codification. By contrast, unrestricted accessible rules must exist when governments do not provide instructional rules for data publication, and there are no requirements on data structure rules. This is possible because, when governments do not intend to impose any data usage instructions to further promote equal participation, any acquisition restriction rules such as registration and payment should be removed. And any data, no matter in structured or unstructured formats, should be encouraged to open to stimulate equal and easy participation and data usage.

5.2 Theoretical contributions

This work adds to the body of knowledge on OGD in the following ways. First, this study proposes an analytical framework of institutional dimensions of OGD, which argues that instructional rules, structural rules and accessible rules are three institutional pillars supporting OGD-driven innovation. This framework advances the current understanding of what kind of institutions will affect data usage and exploitation. As we explained in the

Independent variable in three groups	Direct effect with mediators	Perceived data usefulness	Indirect effect (95% CI)		Conclusions
			Data capability		
Instructional – structural	0.682***	0.158 [0.124–0.191]	0.160 [0.131–0.191]		Both complementary
Instructional – accessible	0.698***	0.151 [0.120–0.182]	0.151 [0.121–0.180]		Both complementary
Structural – accessible	0.539*	0.229 [0.032–0.375]	0.232 [0.018–0.402]		Both complementary

Note(s): * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Source(s): Table by authors

Table 7.
The results of
mediation testing with
binary institutional
dimensions on SRI

introduction, although scholars recognize the importance of institutional design for OGD, they do not clarify what rules should be considered. Our framework just bridges this gap.

Second, this study adds to the OGD literature by empirically testing the causal relationships between the institutional dimensions and scientific research innovation (SRI). The findings from the individual and binary effect of direct and indirect of OGD and SRI tests improve our understanding of the mechanism of OGD-driven innovation. Comparing our findings with the opinions in the OGD literature, the similarity is the discovery of importance of instructional rules and access restrictions to promote data usage and exploitation, but the difference lies in the fact that our findings do not support the significance of data format in influencing data-driven innovation. Additionally, and most importantly, we add the OGD literature by analyzing the interplay between different institutional dimensions and their joint effects on innovation. This is the new knowledge added to the OGD literature.

5.3 Practical implications

This research also has practical implications. These practical implications can be transferred to other transition countries that strive for advancing their scientific research innovation and building up suitable institutional frameworks for OGD.

First, this study confirms the importance of instructional rules and accessible rules in promoting innovation. Therefore, it is essential for governments to develop and continuously construct instructions regarding data usage and to establish necessary data acquisition restrictions. In the field of scientific research in particular, governments also need to be aware that users' perceived data usefulness and data capability will influence the performance of OGD-driven innovation. As such, governments are encouraged to carry out events that can enhance potential users' perception on data usefulness (such as data marketing and educational activities) and that can increase users' capability in using data (such as technical training activities).

Secondly, our findings reveal the effects of binary institutional dimensions on innovation. A practical implication is that when governments provide instructional rules for data publication, restricted accessible rules or unstructured rules are preferable to go together to facilitate SRI. However, when governments fail to provide instructional guidelines for data usage, freely accessible rules must be available. And in this situation, there are no mandates for data structure rules, indicating that governments may or may not issue requirements on data structure.

5.4 Limitations and future research

There are three main limitations to this study that can be addressed in future research. First, this study relies on questionnaire data. This inevitably raises the subjectivity problems. In future research, objective data are encouraged to use to measure the variables. Second, our respondents are researchers in the Chinese research institutes, the scope condition for generalizing our findings is limited to the scientific research field. How the institutional dimensions should be designed to promote innovation in other fields, such as business models or public services, are subject to future research. Third, this study only considers researchers' perceived data usefulness and data capability as two mediating variables, but more mediating variables, such as researchers' social capital or organizational environment, should be tested in future research.

Notes

1. Project 985 is a national strategy proposed by the Chinese government in 1998 to cultivate world-class universities in the 21st century, including 39 universities. It is always located on the first ladder in the ranking of Chinese higher education institutions.

-
2. Project 211 is the Chinese government's endeavor initiated in 1995 to construct 100 universities and a series of critical disciplines in the 21st century. And the ranking of these universities is lower than the project 985 universities.

References

- Altayar, M.S. (2018), "Motivations for open data adoption: an institutional theory perspective", *Government Information Quarterly*, Vol. 35 No. 4, pp. 633-643, doi: [10.1016/j.giq.2018.09.006](https://doi.org/10.1016/j.giq.2018.09.006).
- Attard, J., Orlandi, F., Scerri, S. and Auer, S. (2015), "A systematic review of open government data initiatives", *Government Information Quarterly*, Vol. 32 No. 4, pp. 399-418, doi: [10.1016/j.giq.2015.07.006](https://doi.org/10.1016/j.giq.2015.07.006).
- Bargh, M.S., Choenni, S. and Meijer, R. (2016), "On design and deployment of two privacy-preserving procedures for judicial-data dissemination", *Government Information Quarterly*, Vol. 33 No. 3, pp. 481-493, doi: [10.1016/j.giq.2016.06.002](https://doi.org/10.1016/j.giq.2016.06.002).
- Bertot, J., McDermott, P. and Smith, T. (2012), "Measurement of open government: metrics and process", *Proceeding of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA*, IEEE Computer Society, pp. 2491-2499.
- Bonina, C. and Eaton, B. (2020), "Cultivating open government data platform ecosystems through governance: lessons from Buenos Aires, Mexico City and Montevideo", *Government Information Quarterly*, Vol. 37 No. 3, 101479, doi: [10.1016/j.giq.2020.101479](https://doi.org/10.1016/j.giq.2020.101479).
- Chin, W.W., Marcolin, B.L. and Newsted, P.R. (2003), "A partial least squares latent variable modeling approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study", *Information Systems Research*, Vol. 14 No. 2, pp. 189-217, doi: [10.1287/isre.14.2.189.16018](https://doi.org/10.1287/isre.14.2.189.16018).
- Corley, K.G. and Gioia, D.A. (2011), "Building theory about theory building: what constitutes a theoretical contribution?", *Academy of Management Review*, Vol. 36 No. 1, pp. 12-32, doi: [10.5465/amr.2009.0486](https://doi.org/10.5465/amr.2009.0486).
- Curry, R. (2015), "Beyond 'data thrifting': an investigation of factors influencing research data reuse in the social science", available at: <https://surface.syr.edu/etd/266> (accessed 5 September 2022).
- Dawes, S.S. (2010), "Stewardship and usefulness: policy principles for information-based transparency", *Government Information Quarterly*, Vol. 27 No. 4, pp. 377-383, doi: [10.1016/j.giq.2010.07.001](https://doi.org/10.1016/j.giq.2010.07.001).
- Diamantopoulos, A. and Winklhofer, H.M. (2001), "Index construction with formative indicators: an alternative to scale development", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 269-277, doi: [10.1509/jmkr.38.2.269.18845](https://doi.org/10.1509/jmkr.38.2.269.18845).
- Faniel, I.M., Kriesberg, A. and Yakel, E. (2016), "Social scientists' satisfaction with data reuse", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 6, pp. 1404-1416, doi: [10.1002/asi.23480](https://doi.org/10.1002/asi.23480).
- Figlio, D., Karbownik, K. and Salvanes, K. (2017), "The promise of administrative data in education research", *Education Finance and Policy*, Vol. 12 No. 2, pp. 129-136, doi: [10.1162/edfp_a_00229](https://doi.org/10.1162/edfp_a_00229).
- Gonzalez-Zapata, F. and Heeks, R. (2015), "The multiple meanings of open government data: understanding different stakeholders and their perspectives", *Government Information Quarterly*, Vol. 32 No. 4, pp. 441-452, doi: [10.1016/j.giq.2015.09.001](https://doi.org/10.1016/j.giq.2015.09.001).
- Grossman, J. and Pedahzur, A. (2020), "Political Science and big data: structured data, unstructured data, and how to use them", *Political Science Quarterly*, Vol. 135 No. 2, pp. 225-257, doi: [10.1002/polq.13032](https://doi.org/10.1002/polq.13032).
- Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24, doi: [10.1108/ebrev-11-2018-0203](https://doi.org/10.1108/ebrev-11-2018-0203).
- Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimae, M., Barreto, M.L. and Goldstein, H. (2017), "Challenges in administrative data linkage for research", *Big Data and Society*, Vol. 4 No. 2, pp. 1-12, doi: [10.1177/20539517177456](https://doi.org/10.1177/20539517177456).

-
- Heinze, T., Shapira, P., Senker, J. and Kuhlmann, S. (2007), "Identifying creative research accomplishments: methodology and results for nanotechnology and human genetics", *Scientometrics*, Vol. 70 No. 1, pp. 125-152, doi: [10.1007/s11192-007-0108-6](https://doi.org/10.1007/s11192-007-0108-6).
- Janssen, M. and van den Hoven, J. (2015), "Big and open linked data (BOLD) in government: a challenge to transparency and privacy?", *Government Information Quarterly*, Vol. 32 No. 4, pp. 363-368, doi: [10.1016/j.giq.2015.11.007](https://doi.org/10.1016/j.giq.2015.11.007).
- Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012), "Benefits, adoption barriers and myths of open data and open government", *Information Systems Management*, Vol. 29 No. 4, pp. 258-268, doi: [10.1080/10580530.2012.716740](https://doi.org/10.1080/10580530.2012.716740).
- Jetzek, T., Avital, M. and Bjorn-Andersen, N. (2014), "Data-driven innovation through open government data", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9 No. 2, pp. 100-120, doi: [10.4067/S0718-18762014000200008](https://doi.org/10.4067/S0718-18762014000200008).
- Kassen, M. (2018), "Adopting and managing open data: stakeholder perspective, challenges and policy recommendations", *Aslib Journal of Information Management*, Vol. 70 No. 5, pp. 518-537, doi: [10.1108/ajim-11-2017-0250](https://doi.org/10.1108/ajim-11-2017-0250).
- Kassen, M. (2020), "Open data and its peers: understanding promising harbingers from Nordic Europe", *Aslib Journal of Information Management*, Vol. 72 No. 5, pp. 765-785, doi: [10.1108/ajim-12-2019-0364](https://doi.org/10.1108/ajim-12-2019-0364).
- Kim, Y. (2021), "A study of the roles of metadata standard and data repository in science, technology, engineering and mathematics researchers' data reuse", *Online Information Review*, Vol. 45 No. 7, pp. 1306-1321, doi: [10.1108/oir-09-2020-0431](https://doi.org/10.1108/oir-09-2020-0431).
- Kitchin, R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage Publication, Thousand Oaks, CA.
- Lassinantti, J., Stahlbrost, A. and Runardotter, M. (2019), "Relevant social groups for open data use and engagement", *Government Information Quarterly*, Vol. 36 No. 1, pp. 98-111, doi: [10.1016/j.giq.2018.11.001](https://doi.org/10.1016/j.giq.2018.11.001).
- Lee, Y.N., Walsh, J.P. and Wang, J. (2015), "Creativity in scientific teams: unpacking novelty and impact", *Research Policy*, Vol. 44 No. 3, pp. 684-697, doi: [10.1016/j.respol.2014.10.007](https://doi.org/10.1016/j.respol.2014.10.007).
- Li, S. and Chen, Y. (2021), "Explaining the resistance of data providers to open government data", *Aslib Journal of Information Management*, Vol. 73 No. 4, pp. 560-577, doi: [10.1108/ajim-09-2020-0270](https://doi.org/10.1108/ajim-09-2020-0270).
- Li, Q., Wang, P., Sun, Y., Zhang, Y. and Chen, C. (2019), "Data-driven decision making in graduate students' research topic selection: cognitive processes and challenging factors", *Aslib Journal of Information Management*, Vol. 71 No. 5, pp. 657-676, doi: [10.1108/ajim-01-2019-0019](https://doi.org/10.1108/ajim-01-2019-0019).
- Liang, H., Saraf, N., Hu, Q. and Xue, Y. (2007), "Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management", *MIS Quarterly*, Vol. 31 No. 1, pp. 59-87, doi: [10.2307/25148781](https://doi.org/10.2307/25148781).
- Lnenicka, M. and Komarkova, J. (2019), "Big and open linked data analytics ecosystem: theoretical background and essential elements", *Government Information Quarterly*, Vol. 36 No. 1, pp. 129-144, doi: [10.1016/j.giq.2018.11.004](https://doi.org/10.1016/j.giq.2018.11.004).
- Lnenicka, M., Nikiforova, A., Saxena, S. and Singh, P. (2022), "Investigation into the adoption of open government data among students: the behavioural intention-based comparative analysis of three countries", *Aslib Journal of Information Management*, Vol. 74 No. 3, pp. 549-567, doi: [10.1108/ajim-08-2021-0249](https://doi.org/10.1108/ajim-08-2021-0249).
- Lourenco, R.P. (2015), "An analysis of open government portals: a perspective of transparency for accountability", *Government Information Quarterly*, Vol. 32 No. 3, pp. 323-332, doi: [10.1016/j.giq.2015.05.006](https://doi.org/10.1016/j.giq.2015.05.006).
- Machova, R., Hub, M. and Lnenicka, M. (2018), "Usability evaluation of open data portals: evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective", *Aslib Journal of Information Management*, Vol. 70 No. 3, pp. 252-268, doi: [10.1108/ajim-02-2018-0026](https://doi.org/10.1108/ajim-02-2018-0026).

- Martin, N., Matt, C., Niebel, C. and Blind, K. (2019), "How data protection regulation affects startup innovation", *Information Systems Frontiers*, Vol. 21 No. 6, pp. 1307-1324, doi: [10.1007/s10796-019-09974-2](https://doi.org/10.1007/s10796-019-09974-2).
- Meijer, R., Conradie, P.W. and Choenni, S. (2014), "Reconciling contradictions of open data regarding transparency, privacy, security and trust", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9 No. 3, pp. 32-44, doi: [10.4067/S0718-18762014000300004](https://doi.org/10.4067/S0718-18762014000300004).
- Mu, R. and Wang, H. (2022), "A systematic literature review of open innovation in the public sector: comparing barriers and governance strategies of digital and non-digital open innovation", *Public Management Review*, Vol. 24 No. 4, pp. 489-511, doi: [10.1080/14719037.2020.1838787](https://doi.org/10.1080/14719037.2020.1838787).
- Mutambik, I., Almuqrin, A., Liu, Y.D., Halboob, W., Alakeel, A. and Derhab, A. (2023), "Increasing continuous engagement with open government data: learning from the Saudi experience", *Journal of Global Management*, Vol. 31 No. 1, pp. 1-21, doi: [10.4018/jgim.322437](https://doi.org/10.4018/jgim.322437).
- Niebel, C. (2021), "The impact of the general data protection regulation on innovation and the global political economy", *Computer Law and Security Review*, Vol. 40, 105523, doi: [10.1016/j.clsr.2020.105523](https://doi.org/10.1016/j.clsr.2020.105523).
- North, D.C. (1991), "Institutions", *Journal of Economic Perspectives*, Vol. 5 No. 1, pp. 97-112, doi: [10.1257/jep.5.1.97](https://doi.org/10.1257/jep.5.1.97).
- Podsakoff, P.M., Mackenzie, S.B., Lee, J.Y. and Podsakoff, N.P. (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 88 No. 5, pp. 879-903, doi: [10.1037/0021-9010.88.5.879](https://doi.org/10.1037/0021-9010.88.5.879).
- Rizk, A., Stahlbrost, A. and Elragal, A. (2022), "Data-driven innovation processes within federated networks", *European Journal of Innovation Management*, Vol. 25 No. 6, pp. 498-526, doi: [10.1108/ejim-05-2020-0190](https://doi.org/10.1108/ejim-05-2020-0190).
- Safarov, I. (2019), "Institutional dimensions of open government data implementation: evidence from The Netherlands, Sweden, and the UK", *Public Performance and Management Review*, Vol. 42 No. 2, pp. 305-328, doi: [10.1080/15309576.2018.1438296](https://doi.org/10.1080/15309576.2018.1438296).
- Schumpeter, J. (1934), *The Theory of Economic Development*, Harvard University Press, Cambridge, MA.
- Shanock, L.R., Baran, B.E., Gentry, W.A., Pattison, S.C. and Heggstad, E.D. (2010), "Polynomial regression with response surface analysis: a powerful approach for examining moderation and overcoming limitations of differences scores", *Journal of Business and Psychology*, Vol. 25 No. 4, pp. 543-554, doi: [10.1007/s10869-010-9183-4](https://doi.org/10.1007/s10869-010-9183-4).
- Simonton, D.K. (2003), "Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives", *Psychological Bulletin*, Vol. 129 No. 4, pp. 475-494, doi: [10.1037/0033-2909.129.4.475](https://doi.org/10.1037/0033-2909.129.4.475).
- Susha, I., Gronlund, A. and Janssen, M. (2015), "Driving factors of service innovation using open government data: an exploratory study of entrepreneurs in two countries", *Information Polity*, Vol. 20 No. 1, pp. 19-34, doi: [10.3233/ip-150353](https://doi.org/10.3233/ip-150353).
- The Open Government Working Group (2007), "8 principles of open government data", available at: https://public.resource.org/8_principles.html (accessed 5 September 2022).
- Venkatesh, V. and Davis, F.D. (1996), "A model of the antecedents of perceived ease of use: development and test", *Decision Sciences*, Vol. 27 No. 3, pp. 451-481, doi: [10.1111/j.1540-5915.1996.tb00860.x](https://doi.org/10.1111/j.1540-5915.1996.tb00860.x).
- Victorelli, E.Z., Dos Reis, J.C., Hornung, H. and Prado, A.B. (2020), "Understanding human-data interaction: literature review and recommendations for design", *International Journal of Human-Computer Studies*, Vol. 134, pp. 13-32, doi: [10.1016/j.ijhcs.2019.09.004](https://doi.org/10.1016/j.ijhcs.2019.09.004).
- Wang, H.J. and Lo, J. (2016), "Adoption of open government data among agencies", *Government Information Quarterly*, Vol. 33 No. 1, pp. 80-88, doi: [10.1016/j.giq.2015.11.004](https://doi.org/10.1016/j.giq.2015.11.004).
- Wang, V. and Shepherd, D. (2020), "Exploring the extent of openness of open government data – a critique of open government datasets in the UK", *Government Information Quarterly*, Vol. 37 No. 1, 101405, doi: [10.1016/j.giq.2019.101405](https://doi.org/10.1016/j.giq.2019.101405).

-
- Wang, J., Veugelers, R. and Stephan, P. (2017), "Bias against novelty in science: a cautionary tale for users of bibliometric indicators", *Research Policy*, Vol. 46 No. 8, pp. 1416-1436, doi: [10.1016/j.respol.2017.06.006](https://doi.org/10.1016/j.respol.2017.06.006).
- Wang, F., Zhang, Z., Ma, X., Zhang, Y., Li, X. and Zhang, X. (2023), "Path to open government data reuse: a three-dimensional framework of information need, data and government preparation", *Information and Management*, Vol. 60 No. 8, 103879, doi: [10.1016/j.im.2023.103879](https://doi.org/10.1016/j.im.2023.103879).
- Weerakkody, V., Kapoor, K., Balta, M.E., Irani, Z. and Dwivedi, Y.K. (2017), "Factors influencing user acceptance of public sector big open data", *Production Planning and Control*, Vol. 28 Nos 11-12, pp. 891-905, doi: [10.1080/09537287.2017.1336802](https://doi.org/10.1080/09537287.2017.1336802).
- Wilson, B. and Cong, C. (2021), "Beyond the supply side: use and impact of municipal open data in the U.S", *Telematics and Informatics*, Vol. 58, 101526, doi: [10.1016/j.tele.2020.101526](https://doi.org/10.1016/j.tele.2020.101526).
- Yoon, A. and Kim, Y. (2017), "Social scientists' data reuse behaviors: exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories", *Library and Information Science Research*, Vol. 39 No. 3, pp. 224-233, doi: [10.1016/j.lisr.2017.07.008](https://doi.org/10.1016/j.lisr.2017.07.008).
- Zhang, X. and Bartol, K.M. (2010), "Linking empowering leadership and employee creativity: the influence of psychological empowerment, intrinsic motivation, and creative process engagement", *Academy of Management Journal*, Vol. 53 No. 1, pp. 107-128, doi: [10.5465/amj.2010.48037118](https://doi.org/10.5465/amj.2010.48037118).
- Zikopoulos, P., Eaton, C., de Roos, D., Deutsch, T. and Lapis, G. (2012), *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill, New York.
- Zuiderwijk, A. and de Reuver, M. (2021), "Why open government data initiatives fail to achieve their objectives: categorizing and prioritizing barriers through a global survey", *Transforming Government: People Process and Policy*, Vol. 15 No. 4, pp. 377-395, doi: [10.1108/TG-09-2020-0271](https://doi.org/10.1108/TG-09-2020-0271).
- Zuiderwijk, A. and Janssen, M. (2014), "Open data policies, their implementation and impact: a framework for comparison", *Government Information Quarterly*, Vol. 31 No. 1, pp. 17-29, doi: [10.1016/j.giq.2013.04.003](https://doi.org/10.1016/j.giq.2013.04.003).
- Zuiderwijk, A., Gasco, M., Parycek, P. and Janssen, M. (2014), "Special issue on transparency and open data policies: guest editor's introduction", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9 No. 3, pp. I-IX, doi: [10.4067/S0718-18762014000300001](https://doi.org/10.4067/S0718-18762014000300001).

Comment id	Feedback information	Experts id (discipline)
Comment 1	I think you could explain the meaning of OGD in the section on informed content, especially if your research goes beyond the realm of employing experimental data in engineering, science, and other fields	Expert 1 (Engineering) Expert 3 (Science) Expert 4 (Law)
Comment 2	The three UST1–UST3 elements seem to be somewhat similar. To cut down on the time needed to complete the survey, you can think about combining them	Expert 2 (Education) Expert 5 (Engineering)
Comment 3	According to the disciplinary background and paradigmatic approaches, I believe you should look for academics who have experience with utilizing OGD to condense the scope of the survey	Expert 6 (Education) Expert 7 (Science)
Comment 4	My doubt is whether the research database is a kind of OGD	Expert 8 (Agriculture)
Comment 5	It is advised to include self-report and reverse elements in the questionnaire	Expert 9 (Economics) Expert 10 (Administration)
Comment 6	For item RA3, since some institutions would purchase data, it is necessary to distinguish whether it is the research institutions or the researchers that pay for the data	Expert 11 (Administration)

Note(s): We invited researchers from different disciplines (e.g. education, administration) drawn from the “Degree Granting and Talent Training Discipline Catalog that set by the Ministry of Education in China”

Source(s): Table by authors

Table A1.
Expert feedback on the
questionnaire

Dimensions and items	Sources
<p><i>Instructional dimension (ID)</i> The data usage instructions published by governments, such as statistical standards, download interfaces, etc., have provided me with convenience in using the data The data correction and inquiry instructions published by governments, such as correcting information and providing feedback to user inquiries, has increased my confidence in using the data Governments' supervision of data usage, such as anti-crawling and information verification, etc., enables me to use data reasonably and effectively</p>	Bertot <i>et al.</i> (2012), Wang and Lo (2016)
<p><i>Structural dimension-unstructured (UST)</i> I always obtain qualitative government data in the human-readable format, e.g. descriptive government reports I always obtain qualitative government data in image or video format, e.g. political speeches, pictures I always obtain qualitative government data in audio format, e.g. voice datasets</p>	Curty (2015), Wang and Shepherd (2020)
<p><i>Structural dimension-structured (ST)</i> I always obtain quantitative government data in the machine-readable format I always obtain quantitative government data fitted into rows and columns, such as EXCEL, CVS, XLS, JSON, XML datasets I always obtain quantitative government data from API or SPARQL search interfaces</p>	
<p><i>Accessible dimension-restricted (RA)</i> RA1: When I access government data, I need to register RA2: When I access government data, I need to initiate a request and fill in certain forms RA3: When I access government data, I need to pay for it</p>	Curty (2015), Faniel <i>et al.</i> (2016)
<p><i>Accessible dimension-unrestricted (URA)</i> URA: When I access government data, I don't need to register, apply, or pay</p>	
<p><i>Scientific research innovation (SRI)</i> OGD enable me to define research questions from new or different perspectives OGD enable me to reveal the new relationship between variables OGD enable me to establish or apply a new method for solving the research problem OGD enable me to develop the new research tool and software</p>	Zhang and Bartol (2010), Kim (2021)
<p><i>Perceived data usefulness (PDU)</i> I think OGD is very useful for me I think using OGD would save my research time and reduce monetary costs in my research I think using OGD would optimize my research design and verify my research findings</p>	Yoon and Kim (2017)
<p><i>Data capability (DC)</i> I can find the data that are useful for my research I can collect and organize the data that are useful for my research I can choose suitable tools and methods for data processing according to my research purpose I can create charts or figures to exhibit the meaningful information hidden in the datasets</p>	Li <i>et al.</i> (2019)
<p>Source(s): Table by authors</p>	

Table A2.
 The measurement items and literature sources

About the authors

Rui Mu is Professor at School of Public Administration and Policy, Dalian University of Technology in China. She has specialized in open government data and collaborative governance, as well as smart cities and urban sustainability. She published a number of articles around these areas in *Public Management Review*, *Policy and Politics*, *International Review of Administrative Sciences*, *Urban Policy and Research*, *Journal of Cleaner Production*, *Transport Policy*, *Journal of Transport Geography* and *Policy and Society*.

Xiaxia Zhao is Ph.D. candidate at the School of Public Administration and Policy, Dalian University of Technology in China. She does research in the field of science and technology management, specifically on the topic of the relationship between open government data and scientific research innovation. She published articles on "Process Analysis of Public Sector Open Innovation from the Perspective of Knowledge Management" in the Chinese Journal of Administration and Law. Xiaxia Zhao is the corresponding author and can be contacted at: xiaxia682860@mail.dlut.edu.cn

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com