ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202211.00467

Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis (Postprint)

Authors: Wang, Fang, Zhao, An, Zhao, Hong, Chu, Jun, Wang, Fang

Date: 2022-11-29T00:00:00+00:00

Abstract

For many government departments, uncertainty aversion is a source of barriers in the advancement of data openness. A more active response to potential risks is needed and necessitates an in-depth examination of risks related to open government data (OGD). With a cross-case study in which three cases from the United Kingdom, the United States and China are examined, this study identifies potential risks that might emerge at different stages of the lifecycle of OGD programs and constructs a taxonomy model for them. The taxonomy model distinguishes the "risks from OGD" from the "risks to OGD" , which can help government departments make better responses. Finally, risk response strategies are suggested based on the research results.

Full Text

Preamble

Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis

Fang Wang[1],[2]†, An Zhao[1], Hong Zhao[1],[3] & Jun Chu[1]
[1]Business School of Nankai University, Tianjin 300071, China
[2]Center for Network Society Governance of Nankai University, Tianjin 300071, China
[3]CETC Big Data Research Institute Co. Ltd., Guiyang 550081, China

Keywords: Open government data; Risk; Taxonomy; Lifecycle; Risk management

Citation: F. Wang, A. Zhao, H. Zhao & J. Chu. Building a holistic taxonomy model for OGD-related risks: Based on a lifecycle analysis. *Data Intelligence* 1(2019), 309-332. doi: 10.1162/dint_a_{00018}

Received: January 25, 2019; **Revised:** May 10, 2019; **Accepted:** May 18, 2019

ABSTRACT

For many government departments, uncertainty aversion poses a significant barrier to advancing data openness. A more proactive response to potential risks requires an in-depth examination of risks related to open government data (OGD). Through a cross-case study examining three cases from the United Kingdom, the United States, and China, this study identifies potential risks that may emerge at different stages of the OGD lifecycle and constructs a taxonomy model for these risks. The taxonomy model distinguishes "risks from OGD" from "risks to OGD," which can help government departments formulate better responses. Finally, risk response strategies are suggested based on the research findings.

1. INTRODUCTION

Government information disclosure has played an important role in the democratic development of human society since Sweden passed the Freedom of the Press Act in 1766. In particular, since the United States enacted the Freedom of Information Act (FOIA) in 1966, many countries have implemented laws or regulations to protect the public' s "right to know." With the advent of the big data era, significant attention has focused on the value of open government data (OGD) in promoting government transparency and accountability, public participation, and social innovation [1, 2, 3, 4]. By 2016, more than 70 countries had joined the "Open Government Partnership" initiative [5]. In China, 46 local governments had launched OGD websites as of May 2018 [6]. In January 2019, the OPEN Government Data Act of the United States was signed into law [7].

Nevertheless, alongside the worldwide advancement of OGD, various barriers have been reported by leading OGD countries such as the UK [8], the Netherlands [9], and the USA [10]. Numerous studies have explored these barriers, which we categorize into six classes according to their perspectives on OGD: (1) the data user perspective, e.g., lack of knowledge about accessing OGD datasets [11, 12]; (2) the data provider perspective, e.g., institutional barriers [13], basic resources, organizational arrangements, and technical capacity [14], fear of false conclusions [9], and economic issues [15]; (3) the data perspective, e.g., fragmented datasets [16], poorly documented metadata [17], poor data quality [13, 18], poor information usability [11], poor machine-processability [19, 20], and complex data formats [21]; (4) the legislation perspective [13], e.g., difficulties in evaluating dataset eligibility [11], challenges in assessing privacy sensitivity of datasets [10, 22], and complexity of data copyright [23]; (5) the technology perspective [10, 13], e.g., the manner in which data are stored, obtained, and used by departments [24]; and (6) the environment perspective, e.g., external pressures [25].

Some of these barriers exist in reality, while others derive from insufficient understanding of potential risks. For instance, a database of registered usernames and email addresses from data.gov.uk was reportedly leaked [26]. In 2018, the BBC reported that sensitive military information had been disclosed in a data visualization map from a fitness tracking company [27]. Additionally, concerns were raised that geospatial data published on data.gov by the U.S. Department of Agriculture might be used to locate crops targeted for eradication via infestation or even to commit acts of biological warfare [28]. As these cases demonstrate, fear of potential risks may hinder OGD advancement. The culture of risk aversion is one source of existing institutional barriers to OGD [13], and uncertainty avoidance plays a negative moderating role in the relationship between organizational resources and OGD capacity in government departments [14].

The best strategy for addressing risks is understanding and managing them more effectively rather than ignoring or avoiding them. Therefore, we aim to address three questions: "What kinds of risks are involved when government departments implement OGD initiatives?", "How are these risks distributed across the OGD lifecycle?", and "What strategies could be adopted?" To answer these questions, we conducted a cross-case study of three OGD-related cases from three countries: the cessation of the care.data program in the UK, the IRS data breach in the USA, and the slow progress of the OGD program in China.

We identified 14 risks and categorized them into a taxonomy model that distinguishes "risks to OGD" from "risks from OGD." This study deepens current understanding of OGD-related risks and provides new insights into mechanism design for advancing OGD.

2.1 Risks Associated with OGD

Risk refers to the uncertainty of future results under given conditions and within a particular period [29]. Frequently discussed risks related to OGD include data leakage [30, 31, 32], privacy invasion [33, 34, 35], and other information security issues. One study summarized 11 government risks in data release: copyright, trade secret protection, privacy, infrastructure security, publication of improper data or information that might lead to negative attitudes toward public institutions, inaccurate data, misinterpretation of data, absence of data consumers, reduced willingness to cooperate, data overlap, and increased numbers of data requests [36]. The UK Office for National Statistics also proposed better balancing openness with privacy protection [37]. Another scholar argues that risks posed to official statistics departments by big data relate to mission drift, damage to reputation and loss of public trust, inconsistent access and continuity, fragmentation of approaches across jurisdictions, resource constraints and cutbacks, privatization and competition, etc. [38]. Additionally, OGD is vulnerable to risks regarding effectiveness, relevance, and trust [39]. One study categorized risks that may hinder OGD as those related to governance, economic issues, licenses and legal frameworks, data characteristics, metadata, access, and skills

[15]. Based on a single case study of Shanghai, another work summarized potential OGD risks at the legislative, management, and data levels [40].

2.2 Risk Management Related to OGD

In response to risks faced by government departments, the UK National Audit Office (NAO) proposed embedding risk management into core decision-making and planning management processes. The NAO suggested that risk management—including identification, evaluation, processing, and reporting—can help governments make credible decisions and support innovation [41]. In 2013, a subordinate body of the US Treadway Commission released the Internal Control-Integrated Framework to address enterprise risk management across five aspects: controlled environment, risk assessment, control activities, information and communication, and regulatory activities [42]. One study proposed mitigation strategies for identified OGD risks, including monitoring and assessing data demand, proper specification of datasets to avoid duplication, compliance assessment, data anonymization and aggregation, quality control of data publication, establishment of internal and external data catalogs, linking to already-published datasets, properly formulated terms and prompts for third-party data, clearly explained duties, and continuous monitoring of OGD initiative impacts [36]. Another scholar contends that applying tax-related data quality control and data technology is a key element in controlling tax risks [43]. In 2017, the British government updated the Data Protection Act originally promulgated in 1998 to reinforce citizen rights, such as the right to be forgotten, personal data protection, privacy information, and data migration, stipulating that identifying personal information from anonymous data and tampering with such data would result in criminal charges [44]. In 2018, the EU General Data Protection Regulation came into force [45], focusing on protecting and empowering all EU citizens regarding data privacy and reshaping how all organizations, including governments, approach data privacy.

The aforementioned regulations and studies have thoroughly discussed risk management strategies associated with OGD, but they did not address the context in which specific risks occur—for example, the stage in the data lifecycle, or their sources and consequences. Moreover, a single-country context may be insufficient to reveal the variety of risks across more comprehensive institutional backgrounds. These limitations may weaken the effectiveness of suggested strategies. In view of this, we conducted a cross-case study in an international context from the perspective of the OGD lifecycle.

2.3 Lifecycle Model of OGD

Lifecycle models can guide the process of opening up data [46]. Lifecycle analysis draws on biological analysis methods, dividing the development process of objects (records, data, products, projects, organizations, etc.) from generation to extinction into several stages. OGD is also subject to changes throughout

its lifecycle. An analysis based on the OGD lifecycle can help identify risks that exist at different stages. Meanwhile, risk management itself can also be divided into three phases—latency, occurrence, and crisis response [47]—which necessitate different measures.

According to one study, the OGD lifecycle comprises three sections: a pre-processing section (data creation, selection, harmonization, and publishing), an exploitation section (data interlinking, discovery, exploration, and exploitation), and a maintenance section (data curation) [2]. Based on an investigation in the Netherlands, another work developed a community-driven open data lifecycle model comprising data identification, data preparation, data issuance, and data reuse and evaluation [46]. Yet another study discussed OGD barriers in China based on a data-centered lifecycle model, including data organization and processing, storage and distribution, discovery and acquisition, and appreciation and evaluation [48].

With different research purposes, the aforementioned OGD lifecycle models are either data management-centered or value realization-centered. Since OGD risks are mainly borne by government departments, including both data providers and users, a government-centered lifecycle model is needed. Such a model should consist of at least five stages: data creation and collection, data organization, data release, data utilization, and data maintenance.

2.4 Analytical Framework

Based on the literature review above and the OGD lifecycle, we developed an analytical framework for the subsequent case study, as shown in Figure 1 [Figure 1: see original paper].

3.1 Research Procedure

The present study adopts a cross-case study method. Although previous studies have discussed some OGD-related risks, they have neither integrated them into a comprehensive whole nor examined them in detail from the perspective of the OGD lifecycle. To bridge these theoretical gaps, this study employs the case study method. Compared with a single case study, a cross-case study is used to make generalizations and examine themes, similarities, and differences across cases in quantitative or qualitative analysis. Research designs involving multiple cases are generally regarded as more robust than single-case designs, as they provide observation and analysis of a phenomenon across several settings [49].

In this study, we conducted an analysis across three OGD cases from different countries using content analysis to examine the lifecycle distributions of OGD-related risks. The research steps are as follows:

1. Select three cases of OGD in the USA, the UK, and China based on their theoretical potential and representativeness across different stages of the

OGD lifecycle.
2. Collect case data through Internet searches and semi-structured interviews.
3. Identify risks associated with OGD using within-case and cross-case analysis.
4. Categorize all identified risks into a taxonomy model.
5. Analyze the distributions of all risks across five stages of the OGD lifecycle.
6. Suggest countermeasures for OGD risk management.

3.2 Case Selection

Three cases were selected for their theoretical richness and representativeness in answering the research questions and addressing different stages of OGD programs: the UK healthcare data program "care.data," the American Internal Revenue Service (IRS) data breach event, and the Open Data initiative of Shanghai, China. The "care.data" case is typical in revealing risks in data collection and sharing. The IRS data breach is not a typical OGD case, but it reveals mismanagement and malicious use of government data, making it representative for identifying risks characteristic of the OGD maintenance phase. The Shanghai OGD case is significant for revealing risk concerns of government departments at the early stage of an OGD initiative. These three cases, which are both country- and industry-specific and focus on different phases of the OGD lifecycle, can jointly support the identification of risks across the entire lifecycle.

3.3 Data Collection

Case data were collected via Internet searches and semi-structured interviews. The former applies to cases from the UK and the USA and covers news reports and online commentaries. The latter applies to the China case and includes three one-to-one in-depth interviews and one focus group interview with heads of seven government departments. All interviews lasted nine hours total, and 55,600 Chinese words were transcribed within one week of the interviews.

3.4 Data Analysis

A bottom-up coding approach combined with cross-case comparison was adopted to analyze the data. The data analysis was a continuous process that began with data collection. First, all concepts or entities related to risks or actual harms were identified and then compared and categorized within each case. Second, the results from the three cases were listed, analyzed, compared, and categorized. Third, the sources and consequences of all risks were analyzed. Fourth, a taxonomy model of OGD-related risks was constructed. Finally, the distribution of risks across the five stages of the OGD lifecycle was analyzed, and strategies to address them were suggested.

4.1.1 Case Introduction

In 2013, the UK' s long-established National Health Service (NHS) and the Health and Social Care Information Centre (HSCIC) initiated a program called care.data, aiming to improve patient safety and care through information use. The program also sought to create an extensive health records database whose target users would include pharmacies, mental health services, opticians, dentists, and education and training institutions, eventually supporting all healthcare facilities. The data were anonymized and only covered patients' age ranges, gender, and area of residence, though in exceptional circumstances—such as during a pandemic—a researcher could apply to the Minister of Health for removal of these privacy protections. Researchers believed the data would help them develop new treatments and evaluate NHS services. However, the care.data project did not run smoothly. In February 2014, the NHS acknowledged a serious crisis of confidence regarding care.data and instructed family doctors to postpone uploading patient data for up to six months. In fall 2014, the NHS decided to launch four new pilots to collect medical health data for two million patients, but the first pilot was not officially launched until June 2015. Due to poor public communication—such as the absence of press conferences and failed delivery of brochures to families—the NHS' s data collection was collectively denounced by patients, doctors, the British Medical Association, the privacy campaign group Big Brother Watch, and the Association of Medical Research Charities. Consequently, one million people withdrew from the program. Under enormous pressure, care.data was terminated by the NHS on July 6, 2016.

4.1.2 Risk Identification of Case 1

The NHS is an important source of population data, and care.data represents a typical example of government data collection and utilization. Through analysis of data collected from online news reports, blogs, and commentaries, the following risks were identified:

1. **Privacy leakage risk.** Health data are personal and sensitive in nature. Mudie, an opponent of care.data, stated: "The human cost to the patient whose identity and medical history are made public is potentially disastrous. Careers could be ended, jobs lost, insurance refused, and relationships destroyed if sensitive medical facts are made public or used by private firms, other people, or indeed the media" [50]. Although patient data collected by care.data had been anonymized, data users typically combined open data with holders' closed and accessible data, and the public worried that malicious analysts could use specific techniques to identify patients' private information [51].
2. **Risks from implicit operational specifications.** A lack of explicit operational specifications created various difficulties in data opening and collection. Public accusations included: "the ambiguous standards for obtaining health data pose a risk of trust between doctors and patients

[52]" and "regulations over data access, data inspection, and balance are not yet established or implemented [53]." The NHS did not systematically organize and process collected datasets, leading to perplexing usage and potential safety hazards when cooperating with other agencies.

3. **Risks of improper data use, especially in cooperation with commercial organizations.** Public concern focused on the risk of sharing sensitive medical information with commercial companies without explicit patient consent. Care.data handed over coded patient data to the insurance industry to help actuaries calculate average premiums, but the public believed this could identify individuals' tendencies to become ill, causing bias against insured individuals when purchasing insurance. In cooperation between NHS and Google' s DeepMind, the public also worried that commercial organizations would use patients' private data for profit. As one critic noted, "Any effort to ignore the use of data and only discuss open data policy will fail as they are faced with disorder in reality and compromise in practice [53]."

4. **Data quality risks.** These risks arise from data collection methods, such as incompleteness, distortion, and inaccuracies. For example, care.data required general practitioners to collect patient data, yet data pertaining to young, healthy males might be missing. Some people obtained prescription drugs but flushed them down the toilet when unmonitored, causing distortions in drug performance data.

5. **Risks from immature techniques.** One reason for public opposition to care.data was the lack of technical sophistication. Sheila Bird, a professor of statistics at Strathclyde University, stated: "Data-sharing as proposed by care.data was disastrously incompetent—both ethically and technically. Professionals rebelled and prevailed in out-casting care.data, thereby ensuring that future proposals will not succeed unless both technically proficient and in the public interest" [54].

6. **The risk of public trust crisis prompted by immature government regulation.** Many objections against care.data could be regarded as a function of the public trust crisis caused by a flawed government supervision system. As one critic observed, "When you propose to share our most confidential medical records, ambiguous promises and fictitious regulatory frameworks are disturbing to the public" [50].

7. **The risk of poor communication.** Huge external pressures on care.data stemmed from lack of publicity and ineffective communication. While the project was valuable, it required understanding, support, and cooperation from citizens and other organizations. In November 2014, the All Party Parliamentary Group in the British Parliament investigated the care.data project, accusing it of lacking transparency and poor publicity, which eventually led to its failure.

8. **The risk of unsustainable funding for public communication.** Be-

tween October 2013 and 2014, the NHS announced £2 million to publicize the care.data project to the general public, but the actual cost of advocacy was merely £1 million. A survey found that less than one-third of the public received publicity brochures due to insufficient investment in public communication, which eventually forced the project to cease under tremendous external pressure.

9. **The risk of oversized external pressures.** After the care.data program was announced in early 2013, some privacy organizations launched the Medical Data Confidentiality Initiative, calling attention to security risks in medical data use. Since then, these organizations have applied significant pressure to initiatives collecting medical data.

10. **The risk of unprofessional information governance.** In January 2015, the NHS' s Independent Information Governance Oversight Panel released a report indicating that initial commitments to the care.data project were not completed, partly due to the lack of experts in information governance.

4.2.1 Case Introduction

In 2015, a data breach occurred at the Internal Revenue Service (IRS) of the United States [55]. With assistance from companies such as IBM, the IRS had established a complex data platform networked with other government agencies. The site featured an application called "Get Transcript" that allowed citizens to easily access previous tax records. In 2015, hackers illegally accessed approximately 724,000 taxpayers' tax returns via the Get Transcript app. This malicious behavior was not detected until three months later. After investigation, J. Russell George, the Treasury Inspector General for tax administration, accused the IRS of not deploying web systems according to requirements and recommendations, which caused serious data breaches amid reduced staffing and increased capital input [56].

4.2.2 Risk Identification of Case 2

1. **The risk of hacking.** Hacking is a constant threat to government applications, with immediate consequences including leakage of privacy, trade secrets, or even national infrastructure security information.

2. **The risk of poorly implemented government regulations.** After implementation of the My Data project, the US Open Data Action Plan established clear requirements regarding network and data security systems for government departments, but the IRS did not seriously follow these requirements, resulting in a colossal security breach.

3. **The risk of poor communication and cooperation between departments.** After the data breach, the IRS, the Treasury Department,

and other departments were at odds with each other. Poor communication and cooperation resulted in an inadequate response to the risk.

4. **The risk of delayed system updates and maintenance.** Maintenance and updates of an open system platform must be carried out periodically; otherwise, security risks are likely to occur. In this case, some IRS applications were outdated with many security vulnerabilities, coupled with poor maintenance of systems and platforms. Together, these factors created vulnerabilities to hacking that remained undetected for three months.

5. **The risk of unsustainability in capital investment.** According to the IRS, an increasing number of government information systems security services are being outsourced; with the delay of at least $ 400 million in investments due to IT budget cuts, many of the operations, including the maintenance and replacement of old IT systems, are affected, thereby resulting in system failure and security loopholes.

4.3 Case 3: The Tardy Progress of the OGD Program in China

During the short history of OGD in China, no major crisis has taken place yet. Therefore it is impossible to have an event as a case study that exhibits the lifecycle. Instead, we only investigate the relevant risk factors by examining a representative city government. Between December 2016 and May 2017, we interviewed seven government departments in Shanghai, including the Human Resources and Social Security Bureau, the Agriculture Commission, the Food and Drug Administration, the Audit Bureau, the Health and Family Planning Commission, the Trade and Industry Bureau, and the Planning and Land Resources Administration. According to the analysis of interview data, the risk concerns of the departments related to OGD can be summarized into the following six aspects:

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 1). Risk of data distortion. Government statistical data, though eye-catching and frequently used by the public, are prone to distortion in the layer-by-layer hierarchical reporting system; this has become a systemic problem. One interviewee said: “It is not that our attitude to OGD is inactive, but that the authenticity of data reported by the subordinate departments cannot be guaranteed” . 2). Risk of low data quality. Given that some government departments’ data are collected from enterprises or reported by subordinate departments and other diversified channels, some data are thus flawed with inconsistent format, unstandardized metadata, incompleteness and other quality problems, which forestalls some departments from engaging in OGD pilots. One respondent noted: “The data collected directly from medical institutions and health administration bureaus at county or district level were found countless quality problems. We’ ve been doing data

cleaning for two-three consecutive years, so a lot of data have not yet been released."3). Risk of data aggregation. Although open data have been anonymized and desensitized, some departments are still worried about the existence of security risks for collective data release. One interviewee said: "The bulk of our data are disease information or disease prevention information, if released, it implicates personal privacy." 4). Risk of undefined operational norms. Some departments expressed that due to the lack of technical standards and codes of practice, they are quite at sea for some problems in open data. One interviewee suggested that "the scope of disclosure also needs to be stipulated. We are willing to participate in the construction and also want to provide useful data, but the specification of the application needs to be well defined." 5). Risk of imperfect mechanisms. The lack of supervision, feedback and incentive mechanism led to the slow progress of the OGD project. On account of the scanty feedback on the use of open data, the enthusiasm of some departments for OGD has been dampened. One interviewee said: "We are in want of a mechanism for feedback and supervision, we would like to know where and to what extent the data we provide have realized their value. All these require feedback, which is also a positive incentive for us." 6). Risk of data value-sparseness. Some departments want to know if their data are really useful to the public. "Every year we provide a lot of open data to the community (for social services), and a great deal of work has been done, but I am still a stranger to the data they (neighborhoods) are using and how they use them." "We' d love to know who is using the data."

5. RISK INTEGRATION AND CLASSIFICATION

Through iterative comparison and merging, the risks identified from the three cases are integrated into 14 types of risks: content legitimacy, data quality, data value, data management, platform support, information security, organizational support, resource input, institutional support, business process, use, imperfect regulations and standards, scarce external resources and external pressures. The 14 types of risks are then further categorized based on their sources and consequences, respectively.

5.1 Classification Based on the Risk Source

According to the source, the 14 types of risks are categorized into five classes: data (related) risks, technology (related) risks, management (related) risks, utilization (related) risks and external environment (related) risks.

5.1.1 Data Risks

Most of the data published on OGD websites are submitted by various government departments, while a small part of them are created by the OGD depart-

ment itself. These data vary in their attributes accordingly with the clear-cut department functions. Some attributes may lead to uncertain consequences if the data are open. For instance, some data concern citizens' privacy, and their inappropriate disclosure may bring harm to the individuals involved. These risks arising from data attributes are categorized as data risk, which is mainly related to the legitimacy of the open content, the data quality, and the data value.

The risk in the legitimacy of open content refers to the possibility that the open content does not comply with laws, regulations, or other social norms. The risk of low data quality includes the possible distortion, error, obsolescence, or incompleteness of data content, format chaos, absent links or bad metadata. This kind of risks may bring inestimable losses to data users or damage the government' s credibility. The main problems lie in the phase of data collection and pre-processing. Some desensitized data also suffer from quality degradation after de-identification. The risk of low data value refers to the possible investment loss of resources due to opening data that have trivial value, have no clear users, or are rarely used.

5.1.2 Technology Risks

Technology risks result from inadequate technical capacity and tardy understanding of new technologies, such as data management, platform support and information security. Technology problems in data management may lead to unsatisfactory effects of OGD, such as the use of outdated techniques for data collection, improper methods for data cataloguing, unstandardized metadata and improper data formats that encumber analysis and utilization. The low capacity and delayed updates of the OGD platform bring disastrous consequences, as evinced in case 2. Because of the loopholes in the information security technologies of OGD platforms, worrisome consequences, such as privacy leakage, data tampering and corruption, platform damage or falsely authorized certification, may occur.

5.1.3 Management Risks

Management risks are caused by problems that exist in OGD-related organizational structures, business processes, management styles, or mechanism designs; they comprise the following four types: organizational risks, resource input risks, institutional risks and business process risks. The organizational problems may exist in the lack of special position setting, the unbalanced allocation of power and responsibility between the data provider and receiver, poor interdepartmental communication and cooperation, weak OGD promotion strategies, poor management of business outsourcing, etc. Insufficient investment in talent or Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 funds in OGD-related businesses may directly hinder the promotion of OGD. Incomplete management

rules, business guidelines, incentive mechanisms or ill-conceived business processes upon OGD may also bring unwanted consequences.

5.1.4 Utilization Risks

Utilization risks arise from misuse, malicious or improper use, or insufficient use of the open data, which may result from insufficient literacy or capacity of OGD users. This may result in erroneous decision making, invasion of privacy, or insufficient exploitation of data value.

5.1.5 Environment Risks

Environment risk refers to a harmful impact upon the process, mode, or outcome of an OGD program due to the limitations of current institutions, resources, or public expectations, including: (1) Risk of imperfect regulations and standards. Some obstacles to OGD may be attributable to imperfect laws or regulations, inoperable technical standards, or rigid government administration systems. (2) Risk of scarce external resources, e.g., the lack of data governance experts, insufficient talent supply, or immature knowledge of OGD. (3) Risk of excessive external pressure. As shown in case 1, online negative opinions of pressure groups exerted significant pressure on the initiative of OGD.

5.2 Classification Based on the Risk Consequence

Judging from the consequences, the afore-mentioned 14 subtypes of risks can be divided into two classes: risks to OGD and risks from OGD. The risks to OGD may hinder the smooth running of an OGD program but will not undermine the legitimacy of OGD itself. These risks comprise10 sub-categories: risks in data management, information security, platform support, organizational adjustment, resource investment, institutional provision, business process, imperfect regulations and standards, scare external resources, and excessive external pressures. The risks from OGD refers to the possible negative effects caused by OGD, which may lead to challenges to the legitimacy of the OGD, including illegal contents, low data value, poor data quality and improper data utilization. The risks from OGD may lead to a sceptical attitude to OGD initiative, while the risks to OGD may influence the success of OGD program. The fear of both may become the actual barriers to OGD.

5.3 A Holistic Taxonomy Model of Risks Associated with OGD

Based on the analysis above, a holistic taxonomy model of OGD-related risks is constructed, as shown in Figure 2 [Figure 2: see original paper]. First, all risks identified from the three cases are listed together and compared. Those that are similar are clustered and categorized into an upper class according to their sources and consequences, respectively. Next, the correspondence between

the source-based and consequence-based classifications is analyzed. Finally, a holistic taxonomy model of OGD-related risks is constructed. This model can help government departments and the public better understand OGD-related risks and thus devise more efficient response strategies.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 Figure 2. A taxonomy model of open government data (OGD)-related risks.

6.1 Distribution of Risks from Five Sources

Using a content analysis for the qualitative data of the three cases, the lifecycle distributions of risks from five sources are revealed, as shown in Table 1 . The frequency of each type of risk occurring at every stage of the OGD lifecycle is calculated. At different stages, the distributions of the 14 types of risks are shown in Figure 3 [Figure 3: see original paper]. Among the five stages, the data collection stage is risk intensive. This implies that an appropriate risk management plan should be made before the OGD program begins.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 Table 1. The lifecycle distribution of open government data (OGD)-related risks.

Collection Organiza- Release Utilization Mainte- nance Data risks Risks Risks to Utilization risk Technology risks Data management Content legitimacy Data quality Data value Platform support Information security Management risks Organizational Environment risks support Resource input Institutional support Business process Imperfect regulations and standards Scarce external resources External pressures Figure 3. The lifecycle distribution of 14 types of open government data (OGD)-related risks.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊

6.2 The Occurrence Stages of 14 Types of Risks

A radar chart is made to show the number of the occurrence stages of 14 risks. As shown in Figure 4 [Figure 4: see original paper], the risks that may occur at all five stages include those related to data value, data management, platform support, information security, resource input, institutional risk, and flawed regulations and standards. This indicates the importance of a strong technical platform and a sound management system for the success of the OGD program. The risks that may occur at four stages include organizational and business process risk, which still imply the importance of OGD project management. The risks that may occur at three stages are those related to content legitimacy, data quality, scarce external sources and external pressures.

Among the risks that occur at all five stages, only the risk related to data value

is a risk from OGD. This means that, except their natural characteristics, the value of the open data should be improved at every stage of the OGD lifecycle. The other three types of risks from OGD emerge at three or fewer stages. The risk of data use only occurs at the utilization stage. This implies that a specific type of risk that may be brought by OGD should be controlled or avoided at a few specific stages. Most of the risks to OGD could be mitigated by improving project management and a few of them could be avoided by strengthening regulations and technical standards, e.g., the enforcement of EU GDPR.

Content legality External pressures Scarce external resources Imperfect regula(cid:3)on & standard Business process Ins(cid:3)tu(cid:3)onal support Data quality Data value Data management Pla(cid:2)orm support Resource input Informa(cid:3)on security Organiza(cid:3)onal support Occurrence frequency Figure 4. The number of the occurrence stages of 14 types of risks.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊

7. RISK MANAGEMENT STRATEGIES FOR OGD PROJECTS

Risks from OGD and risks to OGD have different sources and consequences and are distributed at different stages in the OGD lifecycle. Strategies in response to them are suggested as follows:

7.1.1 Data Governance Strategy for the Risk in Content Legitimacy

The legitimacy of the open content is the first concern of many government departments, and has even become the default explanation for any deficiency in data openness. In response to this kind of risk, the government must accelerate the enactment of related laws and regulations (e.g., privacy protection act, and data security act), establish examination criteria for privacy-involved data and build an effective data governance system to ensure the legitimacy of open contents. In May 2018, the GDPR was enforced by the EU, empower people with more control over their personal data collected by all companies operating in the EU [45]. In the Nanhai district of Foshan city in Guangdong province, China, a data governance committee was established to determine the legitimacy of open data [30].

7.1.2 Value Appraisal and Feedback Mechanism for the Risk in Data Value

Among the departments that are inactive to data openness, some are unconvinced of the value of the data to be released. To ensure the value of data that have yet to be disclosed, it is necessary to establish a mechanism for data value

appraisal mechanism. Further, a feedback mechanism should be set up to inform the data providing departments about the results of data utilization, thus giving them sufficient information to determine the priorities for future releases.

7.1.3 Quality Management and Data Provenance Strategy for the Risk in Data Quality

In response to the risk in data quality, guidelines and evaluation standards for stable data formats and metadata, secure platform and data provenance are needed to ensure the authenticity, integrity and usability of data at each stage of the lifecycle of an OGD program. For instance, the OPEN Government Data Act of the USA requires federal agencies to publish their information online based on an underlying open standard that is maintained by a standard organization [57].

7.1.4 Legislation Strategy for the Risk in Data Utilization

In response to data utilization risks, it is necessary to clearly stipulate the purpose, means, scope, and results of data utilization by speeding up the legislation, and mete out legal penalties for the malicious use of data.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊

7.2.1 Through the Lifecycle of the Risk Itself

During the different periods of the risk lifecycle, different strategies are needed to warn, evaluate and respond appropriately. [58] proposed the establishment of a data risk warning mechanism, an internal control mechanism and the fostering of risk coping ability. (1) During the period of latent risks, government departments must actively analyze and identify potential risks, make efforts to avoid and transfer them, and establish reserved countermeasures for use in times of crisis. In this phase, although the government of Shanghai had foreseen possible adverse consequences at the early stages of their OGD program, they were not able to formulate a complete risk response plan because they lacked an understanding of OGD-related risks. (2) In the risk occurrence phase, it is necessary to initiate the emergency plan immediately and take effective measures to minimize losses as much as possible, including timely cut-loss, communication enhancement and winning public understanding. Due to the failure of taking measures in time to pledge public understanding since 2 million publicity funds were not invested as originally planned, the care.data initiative of UK had to be stopped under public pressures [50]. (3) After the risk occurs, an emergency plan should be initiated immediately to minimize any adverse effects. Due to its limited capacity for crisis management and poor inter-departmental communication, the IRS failed to spot and solve the problem in time after the hacker invasion, and allowed the data-theft to last undetected for three months [59].

7.2.2 Through the Lifecycle of an OGD Project

At all stages of the lifecycle of an OGD project, it is necessary to implement corresponding risk control strategies. 1). At the stage of data collection, a complete OGD plan is needed. A detailed handbook should be developed to set the data open scope, conditions and technical means, especially the security level of data involving secrets. 2). At the data organization stage, a set of operable technique standards is needed so as to ensure trustworthy data quality, such as metadata, data preprocess, data catalogue, and data documentation. 3). At the stage of data release and utilization, detailed regulations should be made to ensure proper data use. Besides, emergency plans should be formulated in response to possible privacy leakage, malicious use, hacker and virus and weak publicity. 4). At the maintenance stage, plans should be made to prevent unsustainable resource investment to ensure the smooth progress of the project.

8. CONCLUSION AND DISCUSSION

Risk aversion is a potential factor of resistance to OGD, which stems largely from the lack of understanding rather than their uncontrollability. The construction of a taxonomy model of risks can help government departments understand them better, and thus avoid conservative inaction. By conducting a cross-case analysis on three cases of OGD in the UK, the USA, and China, this study identified 14 types of risks associated with OGD. According to the risk source, the 14 types of risks are classified into five classes:

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 data, technology, management, utilization, and environmental risks. According to the risk consequence, the 14 types of risks are classified into two groups: risks to OGD and risks from OGD. A holistic taxonomy model of OGD-related risks is then constructed. Based on the model, the distribution of each type of risk across five stages of the OGD lifecycle is analyzed. It is found that the stage of data collection is risk-intensive and seven types of risks may emerge at all five stages. It is also found that most of the risks to OGD could be avoided or mitigated by improving risk management throughout the lifecycle of an OGD project. In response to the risks from OGD, it is necessary to improve data governance, data value appraisal and feedback, quality management, and data provenance and speed up legislation.

Among the 14 types of identified risks, some have been discussed in previous studies, such as privacy leakage [60, 61], low data quality [17, 62], and improper use [63], but beyond these, this study also identifies several new risks associated with OGD, including those concerning content legitimacy, data value, weak data management, excessive external pressures, institutional flaws, and external resource scarcity. Furthermore, this study also makes the following contributions to this field: firstly, it distinguishes the risks to OGD from the risks from OGD, and thus differentiates the legitimacy of OGD from its smooth

realization. This has a theoretical implications on deepening the understanding of OGD-related risks and supporting the rationality of an OGD initiative. Secondly, this study identifies OGD-related risks at different stages of the lifecycle of an OGD program from three cases in different countries and builds a holistic taxonomy model after clustering them. This fills the gap left by the deficiencies of previous studies that focused on a single country and presented the risks fragmentarily. Third, this study analyzes the distribution of each type of risk over five stages of the lifecycle of OGD and suggests specific strategies in response to them. This has practical implications for government departments that are promoting OGD.

The limitations of this study include: (1) Due to the authors' geographical limitations, the data for the cases of the UK and the USA are mainly collected from online sources. Although Google renders a comprehensive search, the absence of interviews with related government officials has prevented the researchers from directly assessing the attitudes of government staff on this point. To make up for this limitation, several rounds of in-depth interviews were conducted in the case of Shanghai. (2) The taxonomy model of OGD-related risks is built with a bottom-up induction approach from three cases. Although it can deepen the understanding of the OGD-related risks, it might have not involved all the potential risks that face other government departments. In the future, we will extend our approach to improve the taxonomy model of OGD-related risks with more cases and to empirically test the correlation between risk management and OGD performance.

AUTHOR CONTRIBUTIONS F. Wang (wangfangnk@nankai.edu.cn) designed the whole framework, examined the results of data analysis and wrote the final paper. A. Zhao (zhaoanstudy@163.com) collected most of the data of the first two cases and made preliminary data analysis. H. Zhao (zhaohong@mail.nankai.edu.cn) analyzed part of the data and revised the draft. J. Chu (chujun620@163.com) took part in all the interviews and transcribed the records.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 ACKNOWLEDGEMENTS We would like to express special thanks to the reviewers and editors for their valuable comments, as well as the interviewees and news commentators for their insightful viewpoints. We are grateful to Jiayue Ma, Wei Zhao, the graduate students, Xiaoyu Wang, Weichong Zhang, Jing Yang, the doctoral students of Nankai University, and Yichen Zhang, the graduate student of University of California, San Diego, for their helps with data collection and figure drawing.

This work has been funded by the project of National Engineering Laboratory of Big Data Application Technology for Improving Government Governance Capability: "The Large-scale Intelligent Government Document Processing Technology Based on Nature Language Processing and Deep Learning" and "Improving the Governance Capability of the Government with Big Data" ; the project of National Social Science Fund of China "Network Society Governance

in China" (No. 14ZDA063), and the project of National Natural Science Fund of China: "Research on the Organization and Mode of Modern Social Governance" (No. 71533002).

REFERENCES

S. Dawes. Stewardship and usefulness policy principles for information-based transparency. Government Information Quarterly 27(4)(2010), 377–383. doi: 10.1016/j.giq.2010.07.001.

J. Attard, F. Orlandi, S. Scerri, & S. Auer. A systematic review of open government data initiatives. Govern- ment Information Quarterly 32(4)(2105), 399–418. doi: 10.1016/j.giq.2015.07.006. [3] M. Janssen, & A. Zuiderwijk. Infomediary business models for connecting open data providers and users.

Social Science Computer Review 32(5)(2014), 694–711. doi: 10.1177/0894439314525902. [4] OECD. Open government data, 2017. Available at: http://www.oecd.org/gov/digital-government/open- government-data.htm. [5] OGP. Global summit, 2016. Available at: https://en.ogpsummit.org/osem/conference/ogp-summit. [6] DMG. China Open Data Index, 2018. Available at: http://ifopendata.fudan.edu.cn/.

SPARC 2019. Passed into Law: OPEN Government Data Act (S.760/H.R. 1770). Available at: https:// sparcopen.org/our-work/open-government-data-act/. [8] C. Martin. Barriers to the open government data agenda: Taking a multi-level perspective. Policy & Internet 6(3)(2014), 217–240. doi: 10.1002/1944-2866.POI367. [9] P. Conradie, & S. Choenni. Exploring process barrier to release public sector information in local govern- ment. In: Proceedings of the ICEGOV, 2012, pp. 5–13. doi: 10.1145/2463728.2463731. [10] J. Zhang, S. Dawes, & J. Sarkis. Exploring stakeholders' expectation of the benefits of barriers of e-government knowledge sharing. The Journal of Enterprise Information Management 18(5)(2005), 548–567. doi: 10.1108/ [11] A. Zuiderwijk, K. Jeffrey, & M. Janssen. The potential of metadata for linked open data and its value for users and publishers. Journal of Electronic Democracy and Open Government 4(2)(2012), 222–244. doi: 10.29379/jedem.v4i2.138. [12] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, & R.S. Alibaks. Socio-technical impediments of open data.

Electronic Journal of e-Government 10(2)(2012), 156–172. Available at: http://www.ejeg.com/issue/ download.html?idArticle=255.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 [13] M. Janssen, Y. Charalabidis, & A. Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. Information Systems Management 29(4)(2012), 258–268. doi: 10.1080/10580530.2012.716740. [14] Y. Zhao, & B. Fan. Exploring open government data capacity of government agency: Based on the resource- based theory. Government Information Quarterly 35(1)(2018), 1–12. doi: 10.1016/j.giq.2018.01.002. [15] S. Martin, M. Foulonneau, S. Turki, & M.

Ihadjadene. Risk analysis to overcome barriers to open data.

Electronic Journal of e-Government 11(1)(2013), 348–359. Available at http://www.ejeg.com/issue/ download.html?idArticle=296. [16] R. McLaren, & R. Waters. Governing location information in the UK. The Cartographic Journal 48(3)(2011), 172–178. doi: 10.1179/000870411X13044121958902. [17] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, & Y.L. Traon. Comparison of metadata quality in open data portals using the analytic hierarchy process. Government Information Quarterly 35(1)(2018), 13–29. doi: 10.1016/j.giq.2017.11.003. [18] A. Vetrò, L. Canova, M. Torchiano, C.O. Minotas, & F. Morando. Open data quality measurement framework:

Definition and application to open government data. Government Information Quarterly 33(2)(2016), 325–337. doi: 10.1016/j.giq.2016.02.001. [19] L. Zheng, & F. Gao. Assessment on China's open government data platforms: Framework, status and problems. In: Proceedings of the 17th International Digital Government Research Conference on Digital Government Research, 2016, pp. 408–414. doi: 10.1145/2912160.2912213. [20] Y. Cao. Government open data survival status: Investigation report on 19 local governments (in Chinese).

Library and Information Service 60(14)(2016), 94–101. doi: 10.13266/j.issn.0252-3116.2016.14.011. [21] A. Whitmore. Using open government data to predict war: A case study of data and systems challenges.

Government Information Quarterly 31(4)(2014), 622–630. doi: 10.1016/j.giq.2014.04.003. [22] S. Choenni, J. van Dijk, & F. Leeuw, Preserving privacy whilst integrating data: Applied to criminal justice.

Information Polity 15(1–2)(2010), 125–138. doi: 10.3233/IP-2010-0202. [23] X. Zhu. The failure of an early episode in the open government data movement: A historical case study.

Government Information Quarterly 34(2)(2017), 256–269. doi: 10.1016/j.giq.2017.03.004. [24] P. Conradie, & S. Choenni. On the barriers for local government releasing open data. Government Informa- tion Quarterly 31(S1)(2014), S10–S17. doi: 10.1016/j.giq.2014.01.003. [25] H.J. Wang, & J. Lo. Adoption of open government data among government agencies. Government Informa- tion Quarterly 33(1)(2016), 80–88. doi: 10.1016/j.giq.2015.11.004. [26] BBC. Government data site user details leak, 2017. Available at https://www.bbc.com/news/technology- [27] BBC. Fitness app Strava lights up staff at military bases, 2018. Available at: https://www.bbc.com/news/technology-42853072. [28] J. Höchtl. Open Government Data –Security risk or mean for threat prevention, 2012. Available at: https://www.slideshare.net/jhoechtl/open-government-data-security-risk-or-mean-for-threat-prevention. [29] C.A. Williams, M. Smith, & P.C. Young. Risk management and insurance. New York: McGraw-Hill Publishing Corporation, 1995, pp. 10–80. [30] F. Wang, & F. Chen. Openness and exploitation of government big data in the course of sate governance (in Chinese). Chinese Public Administration 31(11)(2015), 6–12. doi:10.3782/j.issn.1006-0863.2015.11.01.

[31] J. Tao, & L. Mei. Big data technology embedding government regeneration risk and its control. Journal of Tianshui College of Administration 17(1)(2016), 17-20. [32] K. Granickas. Ethical and responsible use of open government data. European Public Sector Information Platform Topic Report, 2015, No.2. [33] T. Scassa. Privacy and open government. Future Internet 6(2)(2014), 397-413. doi: 10.3390/fi6020397.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 [34] L. Weng, & Y. Li. The opening and sharing of governmental big data: A study on the conditions, obstacles and basic principles (in Chinese). Comparative Economic and Social Systems 32(2)(2016), 113-122. [35] C. Jiang. Big data risk assessment of national cyber security (in Chinese). China Information Security 6(5) (2015), 53-54. doi: 10.3969/j.issn.1674-7844.2015.05.041. [36] J. Kucera, & D. Chlapek. Benefits and risks of open government data. Journal of Systems Integration 5(1) (2014), 30-41. doi: 10.20470/jsi.v5i1.185. [37] R. Davidson. Open data, big data, public trust and risk at the Office for National Statistics. In: Beyond Infotopia: Contextualising Risk, Openness and Transparency in the Information Age, Society for Risk Analysis Europe Conference, 2016, pp. 1-9. [38] R. Kitchin. Big data and official statistics: Opportunities, challenges and risks, 2015. Available at: http:// eprints.maynoothuniversity.ie/7231/1/PC. [39] S. Dawes. A realistic look at open data, 2012. Available at: https://www.w3.org/2012/06/pmod/pmod2012_ submission_{38}.pdf. [40] X. Liu, W. Sun, & L. Zheng. Potential risks of government open data and countermeasures: Talk Shanghai as an example (in Chinese). E-government 77(9)(2017), 22-30. doi: 10.16582/j.cnki.dzzw.2017.09.003. [41] National Audit Office. Managing risks to improve public services, 2004. Available at: http://www.nao.org.uk. [42] COSO. Internal control-integrated framework executive summary, 2013. Available at: https://www.coso.org/Documents/990025P-Executive-Summary-final-may20.pdf. [43] J. Liu. Tax risk control in big data era (in Chinese). Journal of Hunan Tax College 27(6)(2014), 19-21. doi: 10.3969/j.issn.1008-4614.2014.06.006. [44] J. Titcomb. Data protection bill: How will the new laws affect you, 2017. Available at: http://www.telegraph. co.uk/technology/0/data-protection-bill-will-new-laws-affect/. [45] GDPR. General Data Protection Regulation (GDPR). Available at https://gdpr-info.eu/. [46] A. van Veenstra, & T. van den Broek. A community-driven open data lifecycle model based on literature and practice. In: I. Boughzala, M. Janssen, & S. Assar (eds.) Case Studies in e-Government 2.0. Cham, Switzwerland: Springer, 2015. doi: 10.1007/978-3-319-08081-9_{11}. [47] Y. Lu, & J. Lu. Project risk management. Beijing: Tsinghua University Press, 2001. [48] R. Huang, & T. Lai. Barriers of open government data in China from the perspective of data lifecycle (in Chinese). Information studies: Theory & Application 41(2)(2018), 7-13. doi: 10.16353/j.cnki.1000- [49] R.K. Yin. Case study research. Thousand Oaks, CA: Sage Publications, 1994. [50] British Journal of Healthcare Computing. More care.data problems raised, 2014. Available at: http://www.

hitcentral.eu/british-journal-healthcare-computing/more-caredata-problems-raised. [51] B. Goldacre. Care data is in chaos: It breaks my heart, 2014. Available at: https://www.theguardian.com/ commentisfree/2014/feb/28/care-data-is-in-chaos. [52] L. Presser, M. Hruskova, H. Rowbottom, & J. Kancir. Care.data and access to UK health records: Patient privacy and public trust, 2015. Available at: https://techscience.org/a/2015081103/. [53] G. Sollazzo, & D. Miller. Open data in the health sector: Users, stories, products and recommendations,

2017. Available at: http://openhealthcare.org.uk/open-data-in-the-health-sector/.

[54] D. Gray. Life science sector welcomes new guidance on NHS medical data sharing, 2016. Available at: https://www.medicalplasticsnews.com/news/life-science-sector-new-guidance-on-nhs-medical-data/. [55] J. Heckman. IRS: Frequent data breaches make it 'fundamentally more difficult' to verify taxpayers, 2018.

Available at: https://federalnewsnetwork.com/cybersecurity/2018/09/irs-frequent-data-breaches-make-it- fundamentally-more-difficult-to-verify-taxpayers/.

Data Intelligence Building a Holistic Taxonomy Model for OGD-Related Risks: Based on a Lifecycle Analysis ChinaXiv 合作期刊 [56] A. Rappeport. Up to 100,000 taxpayers compromised in Fafsa Tool Breach, 2017. Available at: https://www. nytimes.com/2017/04/06/us/politics/internal-revenue-service-breach-taxpayer-data.html. [57] OPEN Government Data Act 2019. Available at: https://www.datacoalition.org/open-government-data-act/. [58] Y. Xia. Analyzing on the risks and risk management of open government data (in Chinese). Journal of the China Society for Scientific and Technical Information 36(1)(2017), 18–27. [59] E. McKee. IRS data breach allows hackers to steal $30 million from taxpayers, 2017. Available at: https://www.atr.org/irs-data-breach-allows-hackers-steal-30-million-taxpayers. [60] L. van Zoonen. Privacy concerns in smart cities. *Government Information Quarterly* 33(3)(2016), 472–480. doi: 10.1016/j.giq.2016.06.004. [61] J.J. Zhao, & S.Y. Zhao. Opportunities and threats: A security assessment of state e-government websites. *Government Information Quarterly* 27(1)(2010), 49–56. doi: 10.1016/j.giq.2009.07.004. [62] B. Fan, & Y. Zhao. The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data. *Government Information Quarterly* 34(3)(2017), 396–405. doi: 10.1016/j.giq.2017.08.006. [63] S. Dzazali, A. Sulaiman, & A.H. Zolait. Information security landscape and maturity level: Case study of Malaysian Public Service (MPS) organizations. *Government Information Quarterly* 26(4)(2009), 584–593. doi: 10.1016/j.giq.2009.04.004.

AUTHOR BIOGRAPHY

Fang Wang is a Professor of Library and Information Science at the Business School and Director of the Center for Network Society Governance, Nankai University, China. She received her PhD degree from Peking University. She has presided over more than 20 projects funded by the National Natural Science Foundation of China (NSFC) and other foundations, and has published more than 100 papers in Chinese and English as well as 10 books. Her research interests include government information management and knowledge discovery.

An Zhao is currently working in a government agency in Beijing. She received her Master' s degree in Archive Science from Nankai University, China. Her research interest is e-government.

Hong Zhao is currently a PhD student at the Department of Information Resource Management, Business School, Nankai University, China. He received his Master' s degree in Information Science in 2008 from Nankai University, China. His research interests include government information resource management and smart information processing.

Jun Chu is currently working in a government agency in Tianjin. She received her Master' s degree in Archive Science in 2018 from Nankai University. Her research interest is government data sharing.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.