Research on Privacy Risk Identification in Government Data Sharing

Jiahao Pan(✉)

School of Information Engineering, Wuhan University of Technology, Wuhan, China
1203708976@qq.com

Abstract. This study brings together risk governance theory and proposes a framework of privacy risk identification mechanism for government data openness in order to comprehensively and precisely identify the privacy risks contained in the process of open government data and to help privacy risk governance facilitate the process of open government data. Based on the analysis, this paper proposes an effective mechanism framework combining a directional identification mechanism, a driving mechanism, a regulation mechanism, and the relationships between each mechanism. Taking this approach can reduce the risk of private data being used for illegal purposes by governments, which is worth paying attention to.

Keywords: Government Data Sharing · Privacy Risk Identification · Risk Governance Theory

1 Introduction

The rapid development of Internet and Internet of things technology has led to the exponential growth of information and data resources. Data mining technologies such as big data, cloud computing and artificial intelligence have further integrated, interconnected, cross referenced and deeply created all kinds of data, so as to gradually realize data portraits [1]. All sectors of society generally recognize the importance of data in the modernization of national governance. In order to comply with the trend of data governance, the government has put forward plans such as digital country, smart city and open government data. Government data governance has become an important aspect of the current construction of social governance. Governments around the world have accumulated a large amount of data in the information construction. These data are the digital records of the whole social activities and are non-material wealth that can be reused.

Since 2003, the European public sector has carried out the public sector information disclosure directive (PSI directive). In 2009, then U.S. President Barack Obama signed the open government directive, officially opening the curtain on the practice of open government data [2]. In 2011, the open government partnership program and the 2013 G8 open data charter campaign extended it to the world and received great attention from all countries.

G. Ali et al. (Eds.): ISEMSS 2022, ASSEHR 687, pp. 3255–3260, 2022.
https://doi.org/10.2991/978-2-494069-31-2_383

With the large-scale development of the practice of open government data, the theoretical research of open government data has also attracted the attention of researchers in computer science, management, politics, law and social sciences. Researchers and practitioners both acknowledge that open government data has great potential in promoting national and social governance, and also point out some common challenges - how to effectively govern privacy protection. In fact, the timely and effective provision of open government data services usually involves the sharing and use of data between the government and scientific research institutions, enterprises and other entities, and these valuable data information often have accidental or malicious violations. In addition, large-scale data leakage events will further affect public confidence in the government, and even generate political and social governance risks.

Academic and societal attention has been drawn to the issue of privacy risk identification of open government data as a result of the existing research results, however there are still the following shortcomings: First, the black box of the process of privacy risk identification of open government data has not yet been opened. Secondly, the existing research lacks the support and application of scientific theory, particularly risk identification theory. This paper aims to analyze related privacy risk types, and propose mechanisms that can effectively minimize privacy risks in order to reduce the tension between government data sharing and privacy risk issues.

2 Theoretical Framework

Open government data refers to the government's interactive activities of providing citizens and stakeholders with access to information about the local or national government by publishing and reusing the data generated in daily management. Opening government data can not only improve the operation efficiency of public administration departments, enhance the cooperation between public departments and the public, and improve the transparency and democracy of government management; At the same time, through cooperation and innovation with developers, we can find new uses of specific data, explore value-added services, stimulate economic growth, benefit the public interest, and promote social development and innovation. Privacy protection refers to the protection of users' privacy. Privacy was first proposed by American scholars in the late 19th century.

Evidence suggests that a lag in identifying privacy risks, a lack of effective privacy protections aligned to regulation, and complex data management processes that make it difficult for government to identify privacy risks [3]. In this context, government departments are in urgent need of advanced concepts and scientific knowledge to show the practical way forward for the identification of privacy risks of open government data. On the topic of privacy risk identification of open government data, scholars mainly focus on two sub-topics: risk definition and risk identification system.

In order to provide a scientific basis for the construction of government data open privacy risk identification mechanisms, it is necessary to first comprehend the concept of general mechanisms and their operation principles [4]. Mechanisms are the combinations of various functions that a certain system has, which ensures that the system as a whole operates normally, as well as rules, orders, and stages that make the integrated functions work, such as directional function, and power function, regulation function, and

integration function [5]. This definition addresses the causes (orientation and dynamics) as well as the results (regulation) of development but ignores the logic of that process. As a result, the mechanism needs to be complemented by an "execution" function in order to describe the outcomes. Additionally, the integration function contains the logic of the relationship between causes, processes, and results, which may be hidden in the linkages between functions. Accordingly, the basic functions of a mechanism should include: directional function, driving function, execution function, and regulation function. The designability and the hierarchical nature of the mechanism indicate that it is a dynamic process involving multiple secondary (functional) mechanisms interdependent upon one another [6].

3 Types of Privacy Risks in Government Data Sharing

3.1 Risks Associated with Breach of Related Principles

There are two risks associated with violations of the relevant principles: the risk of collection and the risk of use. The former refers to the risks of misuse or unauthorized use and disclosure that might occur due to a lack of clarity regarding the purpose of collecting personal data or the lack of relevance to government functions. As for the latter, it is the risk resulting from the government's use of personal information that there may be misuse and infringement of privacy rights.

3.2 Risk Associated with Invalid License and Low Quality

The risk of invalidating a permission is a privacy risk associated with a breach of the permission principles, such as collecting, using, and disclosing personal information. A potential collection risk is the loss of control over one's data due to the government collecting personal information where the individual does not have the right to choose whether or not to provide it, as well as the content of the data or the person who uses it. The risk of use is the "accidental" misuse of personal data by the government without the individual's consent [7]. The disclosure risk refers to the possibility that the government will disclose extra personal information without the individual's consent. If a disclosure consent is invalid, it can result in the loss of control for individuals, and may have a negative impact on personal reputation.

A low-quality risk arises from the infringement of quality principles, and includes risks related to data obsolescence, data mutilation, and data errors. Data obsolescence refers to the risk of storing personal information for a longer period of time than is necessary for fulfilling the purpose for which it was collected, which results in obsolescence, errors, and related security risks. Mutilation/error risk results from incomplete or incorrect personal data leading to incorrect decisions, uses, and disclosures that adversely affect a person's privacy rights.

3.3 Non-compliance Risk

A non-compliance risk refers to the possibility of privacy violations in the collection, retention, use, disclosure, access, and correction of data. In cases where the government

collects more data from individuals than anticipated or beyond what is functionally necessary, the likelihood of data breach increases [8]. When personal data is not known for how long it will be retained, it may be used over time, resulting in a breach of privacy. The risk of non-compliance in the use of data is the violation of privacy rights by the government through its inappropriate use of personal information. Government sharing of negative information about individuals (e.g., criminal records) with third parties could damage the reputation of individuals. It is also possible that changes in backups can lead to changes in the way information is retained, which can lead to the possibility of privacy breaches.

4 Risk Identification Mechanisms

As stated above, there are basically four risk identification mechanisms that can be effective in appropriately identifying privacy risks.

Identifying the privacy risks associated with the sharing of government data is the application and presentation of a mechanism's operation principle for a special situation. In terms of the mechanism operation principle, the government data open privacy risk identification mechanism manifests as directional identification mechanism, identification driving mechanism, identification execution mechanism, identification regulation mechanism, and the interactions between these mechanisms [9]. Furthermore, the identification field (environment) affects and constrains its operation. In particular, the directional identification mechanism controls the direction in which the recognition driving mechanism operates and directs the execution mechanism to follow. The recognition drive mechanism ensures that the recognition execution mechanism operates smoothly. The recognition control mechanism constrains the recognition execution mechanism.

It is important to note that while the open government data privacy risk identification mechanism follows general principles, it must take into account its particular context. The identification and implementation mechanisms are centered on identifying privacy risks and determining whether they exist, as well as the identification detection, identification drive, and identification regulation in order to ensure the scientific rationality and feasibility of the identification and implementation mechanisms [10]. On the other hand, according to the theory of risk identification, risk identification calls for the use of scientific tools to identify and define risk types. Accordingly, the implementation mechanism for government data open privacy risk identification consists of establishing identification principles, selecting identification methods, and presenting identification results. The next figure has demonstrated such relationship among the privacy risk identification mechanisms (see Fig. 1).

The privacy risk identification driving mechanism provides sufficient momentum for the overall mechanism to operate and functions efficiently. Identification of privacy risks is driven by both the legal system and the authority. The right to privacy, as a fundamental right of citizens protected by law, requires the government to identify and prevent privacy risks involved in data accessibility. Under the dual assessment standards of data utility and privacy security, the government should be proactive in identifying privacy risks in data openness in order to protect citizens' privacy rights and take advantage of the benefits social data can offer.

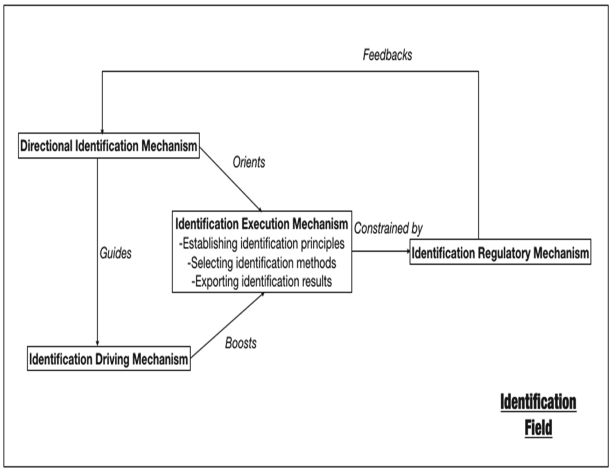

Fig. 1. How identification mechanisms work with each other.

The identification execution mechanism is the most central element of the privacy risk identification mechanism; it must determine the identification principle, select the identification method, and output the identity result [11]. Under this mechanism, relevant identification principles should be established and complied with. Also, effective risk identification methods should be selected in order to output correct risk identification results.

5 Conclusion

The tension between data sharing and privacy security illustrates how privacy risk management has become a challenge for governments to open up their data. Identifying privacy risks has become an increasingly significant part of Data security and confidentiality [12].

First, the government should develop detailed guidelines on privacy risk identification, and comprehensively design and standardize the identification process such as risk identification objects, principles, methods, and updates of identification results. Second, the government should provide sufficient resources for privacy risk identification, including policy, technology, information and human resources. Finally, it should also establish and improve the corresponding supporting system.

References

1. Neves P C, Schmerl B, Cámara J, et al. (2016). Big data in cloud computing: features and issues. International Conference on Internet of Things and Big Data, 307-314.
2. Office of Management and Budget (OMB). (2009). Open Government Directive. http://www.whitehouse.gov/omb/assets/memoranda-2010/m1006.pdf

3. Lee, J., & Jun, S. (2021). Privacy-preserving data mining for open government data from heterogeneous sources. Government Information Quarterly, 38(1), 101544. https://doi.org/https://doi.org/10.1016/j.giq.2020.101544
4. Lindzey, G. (1952). Review of Field theory in social science. The Journal of Abnormal And Social Psychology, 47(1), 132-133. https://doi.org/10.1037/h0052870
5. PASTD, C. (2017). Data Sharing Principles in Developing Countries (The Nairobi Data Sharing Principles). Journal of Global Change Data & Discovery, 1(1), 12–15. https://doi.org/10.3974/geodp.2017.01.03
6. Rose, E. (2006). An examination of the concern for information privacy in the New Zealand regulatory context. Information & Management, 43(3), 322-335. https://doi.org/10.1016/j.im.2005.08.002
7. Jaeger, P., McClure, C., & Fraser, B. (2002). The structures of centralized governmental privacy protection: approaches, models, and analysis. Government Information Quarterly, 19(3), 317-336. https://doi.org/10.1016/s0740-624x(02)00111-9
8. Floridi, L. (2014). Open Data, Data Protection, and Group Privacy. Philosophy & Technology, 27(1), 1-3. https://doi.org/10.1007/s13347-014-0157-8
9. Simpson, R. (2001). How can we keep private data private?. Nursing Management (Springhouse), 32(5), 12-13. https://doi.org/10.1097/00006247-200105000-00006
10. Wang, H., & Lo, J. (2016). Adoption of open government data among government agencies. Government Information Quarterly, 33(1), 80-88. https://doi.org/10.1016/j.giq.2015.11.004
11. IEEE Security & Privacy. (2017), 15(4), c4-c4. https://doi.org/10.1109/msp.2017.3151335
12. Data security and confidentiality. (1985), 4(1), 75-81. https://doi.org/10.1016/0167-4048(85)90010-0