Public Data Security Risk Prevention Research

Yingxue Zhao[1,2,a], Ge Hu[3,b], Jianhua Zhu[4,c], Nan Li[5,d], Denghua Hou[6,e*]

[1]The Fifth Procuratorate Department, Huairou District People's Procuratorate, Beijing, China
[2]School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing, China
[3]School of Grammar, University of Scienceand Technology Beijing, Beijing, China
[4]School of Civil and ResourceEngineering, University of Science and Technology Beijing, Beijing, China
[5]China Automotive Battery Research Institute, Beijing, China
[6]University ofScience and Technology Beijing, Beijing, China

[a]zhaoyingxue1122@126.com, [b]724672165@qq.com,
[c]swinejianhua@163.com, [d]408038473@qq.com,
[e*]396556388@qq.com

Abstract. This study analyzes and compares the domestic and foreign public data open-sharing cases involving the confirmation of rights, privacy, intellectual property rights, and other types of cases by examining the concept of public data open sharing and the current status of legislation, as well as the links and roles of the data life cycle they belong to, and analyzes the risks and challenges faced in depth. It also proposes countermeasures for public data security risk regulation from five perspectives, including targeted individuals, enterprises, governments, countries, and global governance.

Keywords: Public Data; Security Risk Prevention; Open Sharing of Public Data

1 Introduction

Recently, there has been a significant focus on open sharing of public data. In September 2015, the State Council issued *the Action Plan for Promoting the Development of Big Data*, which pointed out to accelerate the opening and sharing of government data, promote the integration of resources and enhance governance capabilities. On June 10, 2018, the General Office of the State Council issued *the Further Deepening of "Internet + Government Services" and the Implementation Plan for Promoting "One Network + One Door + One Time"* Reform of Government Services. In March 2021, *the Outline of the 14th Five-Year Plan of the National Economic and Social Development of the People's Republic of China and Vision 2035* was released, stating: "Establishing a sound national public data resource system helps to ensure public data security, and promote the convergence and integration and deep use of data across sectors, levels, and regions."

X. Ding et al. (eds.), *Proceedings of the 2023 4th International Conference on Big Data and Social Sciences (ICBDSS 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 12,
https://doi.org/10.2991/978-94-6463-276-7_14

Although the nation is paying more and more attention to the construction of the open-sharing mechanism of public data, it is also intended that through data sharing, the "dormant" public data resources can be fully developed and harnessed to give full play to their real value. However, the mechanism under the data element market in China is not yet sound, and there are still more public data security risks. For example, the "Health Code", which came into being as a result of the COVID-19 pandemic, relies on data resource aggregation, digital technology support, and product thinking to enable multi-party, dynamic, and accurate digital governance that drives the flow of people, money and materials in the real world through "data flow" and "data sharing". However, while opening and sharing, it also faces the risk of excessive collection of personal information, inaccurate information, and information leakage.

In response to the above-mentioned issues in the field of data sharing jurisprudence, domestic issues mainly focus on the protection of personal information privacy, data rights protection, data identification, and benefit distribution, use of data and unfair competition, and research on the basic rules of data sharing and transfer. For example, in Legal Analysis and Response of Personal Data Sharing—A Perspective of Internal Sharing in Financial Holding Group, the scholar Jin analyzed in depth the legal research of personal data sharing within financial holding groups; [1]the scholar Huang pointed out in Configuration Rules of Personal Data Ownership in the Era of Big Data that through in-depth analysis from two levels personal data ownership in different situations, the data ownership configuration of data industry and data subjects;[2] the scholar Li pointed out in Reflection on the Protection of Public Data Empowerment and Legal Remedies of Internet Platforms in the Era of Big Data that "when using the anti-unfair competition law to protect public data of Internet platforms, we must change the thinking concept of data protection, abandon the infringement mode of judicial practice, and return to judging the legitimacy of the data competition behavior of Internet platforms themselves." [3]

Thus, compared with the above-mentioned existing studies, it is very necessary to research the public data open-sharing and analyze its risks and challenges.

2 The concept of open sharing of public data and the current status of legislation

At the current stage, data sharing is highlighted by the open sharing of public data resources in the public sector such as the government, while large enterprises are also increasing their efforts in data open sharing. Public data, as an important object of consideration under the market-oriented allocation of data elements, is being continuously explored and used to promote data circulation, activate the data market, and enhance the value of data resources in society as a whole.

Data compliance legislation and standards on the open sharing of public data are steadily emerging and regulated with the introduction of relevant data laws and regulations.

Under the extraterritorial perspective, foreign legislation was introduced earlier and more frequently, and countries and regions with a high degree of public data

openness, such as the United States and Europe, have introduced special public data openness legislation. The U.S. introduced data-sharing legislation earlier, stipulating relevant subjects' data opening and sharing obligations. In 2009, the world's first freely accessible and interactive data-sharing platform Data.gov.us was launched. In February 2013, it was clarified in *the Memorandum on Improving Access to the Results of Scientific Research* funded by the Federal Government that unclassified scientific research data generated by federal funding, in whole or in part, shall provide free and maximum access to the public. In January 2019, *the Open Government Data Act* was passed to clarify that all general government data within the scope of open access should keep open. The EU also continues to enhance the boundaries of data openness and provide for the secondary use of public data with the introduction of *the Open Data and Public Information Reuse Directive* (EU2019/1024) in June 2019, which focuses on the secondary use of public data resources. In addition, the UK government launched data.gov.uk in 2010, which contains 30,000 datasets from various UK government departments, and Japan has built the national open data website Data.go.jp.

From the domestic perspective, domestic data policies and legislation have accelerated, and domestic data legislation has been introduced and ramped up from the policy and regulation level to continuously define and expand the operational value of the data element market to cope with various problems that may arise. At the central level, the Fourth Plenary Session of the 19th CPC Central Committee in the *Decision of the CPC Central Committee on Several Major Issues on Adhering to and Improving the Socialist System with Chinese Characteristics and Promoting the Modernization of the State Governance System and Governance Capability* for the first time clarified the participation of data as a factor of production in social distribution; in the *Opinions of the State Council of the CPC Central Committee on Building a More Perfect Institutional Mechanism for Market-Based Allocation of Factors*, it was proposed to accelerate the cultivation of data elements market, promote the open sharing of government data, enhance the value of social data resources, and strengthen the integration and security protection of data resources, etc. At the legal level, *the Network Security Law* in June 2017, the *Data Security Law* on September 1, 2021, and the *Personal Information Protection Law* on November 1, 2021, have been implemented and introduced one after another. Meanwhile, the *Civil Code*, which will take effect in 2021, stipulates that the handling of personal information includes the collection, storage, use, processing, transmission, provision, and disclosure of personal information. In addition, the *Civil Code* clearly states that personal information processors must disclose the rules of information processing and express the purpose, manner, and scope of information processing.

3 The data lifecycle link in which public data open sharing takes place

Open data sharing is an inexorable requirement to release the value of data, so it is necessary to explicitly study the link in the data life cycle of open data sharing.

First, an introduction to the life cycle of public data matters. The data life cycle is divided into six main stages: collection, storage, transmission, use, sharing, and destruction.

The second is the value of the link in which public data is openly shared. The important value of data opens sharing active market data element market. The development and use of data elements have the characteristics of synergy between supply- and demand-side economies of scale so that the continuous cycle provides more data resources for the data elements market, which in turn continuously generates new information knowledge, enriching the market source supply and activating the data elements market.[4]

Third, it is about the connection between the front and back of public data opening and sharing. Data sharing plays a role in the use of data. Before data can be shared, it needs to be analyzed. The data sale is the last part of the data life cycle, which is a means to eliminate low-value data when the data storage reaches a certain time limit or in the use or sharing process and a way to achieve data security.

4 Risks and challenges to open sharing of public data

The meeting or combination, utilization, and reuse of large amounts of data and data collide to create more useful information and greater value of data. While the open sharing of public data leads to greater opportunities for benefit sharing, it also faces greater risks and challenges.

Currently, China's application of data-sharing mechanisms is mainly focused on administrative affairs such as population census, resource and environmental monitoring data sharing, basic food safety data sharing, transportation regulatory information interaction, and labor and employment situation information sharing. Various administrative departments exchange and share data through data-sharing platforms and exercise data management authority at their level of government through the establishment of special data-sharing management agencies. 2020 witnessed the worldwide dissemination of the COVID-19 pandemic, and China made numerous attempts to share data, achieving a major leap forward in a very short time. The integration of the data collection and shared use such as Healthbot and trip codes made a great contribution to the epidemic attack in China. In addition, the sharing of data by large Internet enterprises to government departments has become an important innovative channel to enrich the pool of data resources for the whole society in the data-sharing mechanism. However, the risks of open data sharing also come with it.

First, it is the source of public data collection: the leakage of relevant personal information violates the privacy rights of individuals. Public data is formed based on thousands of individual data. However, driven by interests, the theft, trafficking, and exploitation of personal data are becoming increasingly rampant. in August 2018, a hotel group's hotel chain user data was sold on the dark web, which included all the registration information, stay information, etc. of customers. There is no information about individuals associated with each piece of information, and the biggest crisis of public data, just corporate public data, being leaked will be the person the data was

originally collected from and pose a huge challenge to individual privacy.[5]

Second, the data storage link is subject to the risk of attack. In April 2016, a major data breach occurred in Turkey. A hacker attacked its national identity information database, resulting in the personal information of nearly 50 million citizens being exposed online.

Third, the data transmission and use link faces data sharing compliance issues among enterprises. Due to data ownership issues, enterprises face data permission and legal compliance issues when sharing data openly. They will also face compliance issues regarding access to data, the collection process, sharing, and allowing third-party use.[6] In addition, the increasing international exchange and cooperation have also put forward higher requirements for Chinese enterprises to go globally. The core data and assets of Chinese enterprises overseas will face the regulation of sovereign countries. Cross-border data governance should focus on how to comply with the relevant laws and regulations of sovereign countries while maximizing the interests of Chinese enterprises.

Fourth, the data sharing link comes across the data black industry surging. As early as 2014, the total size of the underground industry chain of dumping personal data in China exceeded ten billion.

Fifth, amid the data destruction link, public data face the dilemma of the inability for thorough destruction. Once the data is open, especially when it is stored in the cloud, it is easy to delete the data in its original location because it may be shared in multiple data centers. However, when the shared data is moved and deleted, it is difficult to ensure that the data remains, and if the data is still accessible after deletion, it will pose unknown risks.

In addition, asset disputes caused by data sharing have raised many legal issues. The reasons for this are mainly the following four aspects: First, there is a serious problem of data silos in the opening of government data. Relevant policies and laws have been introduced at the national level to promote the sharing and circulation of government information and public data, but the results are not yet obvious, and the problems of duplicate construction and data redundancy still exist. Data sharing includes data sharing among government departments, information sharing among governments across administrative regions, cooperation and sharing of data between government and enterprises, and data sharing among enterprises and institutions, and China has a complete mechanism for opening and sharing the aforementioned data. Second, the regulation of intermediaries is unclear. There exist intermediaries for data sharing and circulation to process the data, enhance the value of the data, and sell it to the users of the data. The processing methods of these intermediaries are different, and the data sources are not fully compliant. The value of processed data sharing and circulation has a significant impact, and without regulation, it may violate personal privacy from time to time. [7]Third, the infrastructure is not yet perfect. Data sharing is inseparable from the construction of a data sharing infrastructure—a data sharing platform. China's data-sharing platform is still in its initial stage, and it still needs to accumulate practical experience from the beginning to the mature application. Fourth, the data standards and classification of data sharing are not yet uniform. More preparatory work is needed before data sharing, such as sharing categories, authority and

responsibility lists, shared data formats, exchange processes, and standards, [8]all of which need further refinement before an enforceable data-sharing mechanism can be established.

5 Domestic and foreign open-sharing policy exploration and practice research

In terms of open data sharing legislation, countries with a high degree of public data openness, such as the US and Europe, have introduced special open public data legislation. The data opening and sharing obligations of relevant subjects are clearly defined, and the scope of open and shared data is continuously expanded, such as the access and sharing of 30% of data in the EU in 2017; This was followed by the publication of the European Guidelines on Data Sharing between Enterprises in July 2017, which provides recommendations to promote data sharing between enterprises. The report Guidance on Private Sector Data Sharing in the European Data Economy was released in April 2018, which examines and analyzes data sharing patterns and constraints, and proposes countermeasures to address them. The Open Data Directive introduced in June 2019 focuses on the reuse of public data resources. Compared with China, foreign countries with a high level of open data sharing have built unified open government data sharing platforms shared by multiple departments to promote data sharing and accumulation of data resources in important public areas. The power of government access to enterprise data is stipulated in the legislation of many countries to promote the fair and reasonable sharing of data generated by the private sector and realize the comprehensive integration and reuse of social public data resources.

Regarding cross-border data open sharing, two main approaches exist extraterritorial. For example, Interpol's data sharing and cooperation. Through data sharing and cooperation, Interpol maximizes the detection of various transnational crimes and protects people's personal and property safety; another example is the data sharing of the Five Eyes Alliance formed by the intelligence agencies of the United States, the United Kingdom, Canada, Australia, and New Zealand. These five countries were formed by the intelligence spy alliance within the interconnection of intelligence information, and stolen commercial data from other countries are shared among government departments and companies and enterprises in these countries. Second, the data control model. Although most countries are more active in facing the cross-border flow of e-commerce data, the regulation of data flow in the political sphere is still relatively stringent.[9]

The relevant domestic legislation is mainly focused on the Data Security Law, which will be implemented on September 1, 2021. It insists on maintaining data security and promoting data development and utilization and encourages and supports the innovative application of data in various industries and fields. In particular, the fifth chapter "government data security and Openness" clearly states in Article 42: "The state is obliged to develop an open directory of government data, build a unified standard, interoperable, secure and controllable government data open platform, and promote the open use of government data." It also highlights the basic path and over-

all goal of establishing and improving China's government affairs data-sharing mechanism. In addition, China's localities have issued corresponding normative documents to promote and regulate the open sharing of public data. Overall, China's current data-sharing mechanism relies on the construction of national data infrastructure and is an important part of the national data governance strategy, which is dedicated to promoting the joint participation of relevant departments, industry organizations, enterprises, individuals, and others in data security protection and forming a collaborative governance system for open data sharing across society.

6 Security risk prevention measures for public data sharing and opening

First, from the perspective of individuals, concerning the most preliminary data collection and use links of data sharing, whether the user's permission is obtained, and whether there is any infringement of the user's rights, more attention ought to be given to the protection of personal privacy, especially children's privacy.[10]

Second, from the perspective of enterprises, it is important to focus on the issue of data ownership and ensure data sharing compliance. The key reason why the ownership of data cannot be clearly defined is that data has multiple attributes, and its attributes are highly dependent on specific scenarios. Under the law of digital market economy development, market orientation should be adhered to, and data should be treated as a production factor. At the same time, it matters to follow a scenario-based rule-making approach to data ownership.

Third, from the government's perspective, efforts should be stressed to increase support for technical research and the development of data-sharing links. It is crucial to further clarify the scope of data opening and sharing under market elements, build a data sharing standard specification system, and strengthen the supervision of data authenticity and sharing standardization.

Fourth, from a national perspective, the ongoing improvement of laws and regulations is vital as data sharing becomes more prevalent. From a jurisprudential point of view, in terms of the legislative concept for data sharing mechanism, it is essential to strike a balance between efficiency and security of data opening; in terms of legislative structure, the whole process of opening up should be the main line, and the core link of opening up should be the focus. A legal system should be in place for open data confidentiality review and security management, to regulate the construction and management of open data platforms and ensure law-based open data sharing.

Indeed, there is a greater need to establish a more robust multiparty data-sharing system for governments, enterprises, and users. Data can be considered both personally owned, owned by the platform, shared by the individual and the platform, and considered public data in the Internet space. Therefore, in turn, the governance and protection of data sharing can be carried out by the government, enterprises, and users at the same time. Users authorize data to enterprises, and enterprises conduct proper market economic activities under the supervision and management of the government to jointly promote the establishment of a data governance system.

Fifth, from the global governance perspective, a comprehensive and multi-level Chinese solution to cross-border data sharing and governance should be established, promoting China's data sharing and governance system, and forming a "Chinese standard" for data flow.

7 Conclusions

During the period of the most preliminary data collection and data sharing, individuals need to protect them from the beginning. Enterprises need to ensure data-sharing compliance. The government needs to strengthen the supervision of data authenticity and sharing standardization. The country also needs to focus on regulating the construction and management of open data platforms and ensuring law-based open data sharing. And from the perspective of global governance, we need to promote China's data sharing and governance system and form a "Chinese standard" for data flow.

Acknowledgment

This paper is a project of the Beijing Social Science Foundation: Global Cross-Border Data Governance and Sovereignty Game, No. 20FXB007.

References

1. Kim, J. (2021). Legal Analysis and Response to Personal Data Sharing - A Perspective on Internal Sharing in Financial Holding Groups. Journal of Yibin College, 21(05): 61-70. Doi: 10.19504/j.cnki.issn1671-5365.2021.05.007.
2. Huang P. (2021). Rules for the Allocation of Personal Data Ownership in the Era of Big Data. Law Science Magazine, 42(01): 99-110. Doi: 10.16092/j.cnki.1001-618x.2021.01.010.
3. Li, X. (2021). Rethinking the Protection of Public Data Empowerment of Internet Platforms in the Era of Big Data and the Way Forward for Legal Remedies. Intellectual Property, (02): 33-48. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=ZSCQ202102004&DbName=CJFQ2021.
4. Li, J. (2021). Data Elements, Digital Economy, and the World of Digital Power High Tech and Industrialization. Law Science Magazine, 27(02): 40-42. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=GKFC202102018&DbName=DKFX2021.
5. Sandra, S.B. (2023). European Union · EDPB Adopts updated Personal Data Breach Notification Guidelines under GDPR: The End of the One-Stop-Shop Reporting Mechanism for Non-EU Establishments. European Data Protection Law Review, 8(4): 517-520. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SJLXC7B2CE637637B6FEA34B545EAF11D3A9&DbName=GARJ2021_4.

6. Friederike, K., Iheanyi, N. (2023). Practitioner's Corner · Managing Data Protection Compliance through Maturity Models: A Primer. European Data Protection Law Review, 8(4): 536-543. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SJLXB1530F9DA54E72D80ADAE998D4DCCA4E&DbName=GARJ2021_4.
7. Schaefer, R. (2019). Consumer Protection: Recalls, Data Security and Congressional Issues. Nova Science Publishers, Inc, New York. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SBCNBC53C743ECD836C77524381CD163301F&DbName=GARBLAST.
8. Ma, L., Zhang, L. (2021). Helping Government Governance under High-Quality Development with Big Data Sharing Platform. Jiangnan Forum, 373(03): 7-9. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JLLT202103003&DbName=CJFQ2021.
9. Robert, H. (2020). Blockchain, GDPR, and Fantasies of Data Sovereignty. Law, Innovation and Technology,12(1): 156-174. doi:10.1080/17579961.2020.1727094.
10. González, E.G., De Hert, P. (2019). Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles. ERA Forum, 19(4): 597-621. doi:10.1007/s12027-018-0546-z.