A Review of Research on Personal Privacy Protection in Domestic and International Government Open Data: Postprint

Authors: Chen Zhaobing, Hao Wenqiang

Date: 2023-04-01T00:00:00+00:00

Abstract

[Objective/Significance] This study systematically reviews domestic and foreign research literature on personal privacy protection in government data openness, aiming to provide references and insights for further research in this field domestically.

[Method/Process] Employing literature survey methodology, along with analytical methods such as induction, comparison, and synthesis, this paper conducts a comprehensive review from aspects including the concept, rationale, logical paradoxes, practical dilemmas, and implementation pathways of personal privacy protection in government data openness.

[Results/Conclusions] The findings indicate that research in this domain has generally undergone three stages: the emergence of privacy issues, personal privacy protection in government information disclosure, and personal privacy protection in government data openness. Research on related topics has yielded substantial achievements, characterized by evident interdisciplinary features, with research perspectives demonstrating both consensus and divergence. It is recommended that future domestic research proceed from four dimensions: expanding the scope of research topics while focusing on key issues; grasping the complexity of research problems and emphasizing systematic thinking; enriching the application of theoretical tools to highlight the processual and dynamic nature of research; and introducing analytical methods such as historical and comparative approaches to strengthen case-based empirical research.

Full Text

Preamble

A Review of Research on Personal Privacy Protection in Government Data Opening at Home and Abroad

Chen Chaobing, Hao Wenqiang

School of Public Administration, Southwestern University of Finance and Economics, Chengdu 611130

Abstract:
[Objective/Significance] This paper systematically reviews the domestic and international research literature on personal privacy protection in government data opening, aiming to provide reference for further research in this field in China. [Method/Process] Using literature research methods and employing analytical approaches such as induction, comparison, and synthesis, this review examines the concept, rationale, logical paradoxes, practical dilemmas, and implementation pathways of personal privacy protection in government data opening. [Result/Conclusion] The findings indicate that research in this field has generally experienced three stages: the emergence of privacy issues, personal privacy protection in government information disclosure, and personal privacy protection in government data opening. Research on related topics has yielded fruitful results, with obvious interdisciplinary characteristics, and research perspectives show both consensus and divergence. It is recommended that future domestic research be conducted from four aspects: expanding the scope of research topics while focusing on key issues; grasping the complexity of research problems and emphasizing the application of systematic thinking; enriching the application of theoretical tools and highlighting the processual and dynamic nature of research; and introducing analytical methods such as historical and comparative analysis while strengthening case-based empirical research.

Keywords: government data opening; government information disclosure; personal privacy; privacy protection; literature review
Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2020.08.016

Since the mid-20th century when human society entered the internet and information age, personal privacy risks and security issues have been a persistent concern. In the 21st-century era of big data, the global movement toward government data opening, while powerfully promoting "democratizing data," has further intensified concerns about privacy protection. Undoubtedly, as countries worldwide undergo a "transformation and upgrading" from government information disclosure to government data opening, privacy protection has become a major practical issue of common concern and high priority for governments and the international community, with privacy leakage, privacy destruction, and privacy infringement becoming increasingly prominent and severe.

Based on attention and response to these realities, the academic community began research on personal privacy protection as early as the 1880s, and has devoted increasing attention to this issue in the context of government data opening in recent decades, accumulating a large body of valuable research findings. How has academic research on personal privacy protection in government data opening evolved? What work has been conducted? What achievements have been made? What deficiencies exist? Answering these questions is a necessary condition for advancing future research in this field, yet few studies have systematically addressed them. This paper aims to summarize and evaluate existing research by reviewing representative domestic and international academic literature, thereby providing assistance for the further development of research in this field.

In terms of literature retrieval and topic selection, on May 6, 2019, we searched the CNKI database using the themes "data opening and privacy" and "information disclosure and privacy" with "precise" matching, and searched the Web of Science database using the themes "freedom of information & privacy" and "open data & privacy," obtaining 55 Chinese documents and 148 foreign-language documents respectively. Based on thematic relevance, we selected 39 Chinese and 36 foreign-language representative documents for in-depth reading and analysis, and on this basis identified the core issues widely concerned and discussed by scholars as the thematic framework for this review.

1. Research Development History

Research on personal privacy protection in government data opening can be traced back to the emergence of privacy protection issues in 1890, and is continuous with research on personal privacy protection in government information disclosure since the 1970s (see Figure 1 [Figure 1: see original paper]). Overall, the development of research on personal privacy protection in government data opening can be divided into three stages: initial attention, preparatory development, and key advancement.

1.1 Initial Attention Stage (1890–1965): Emergence of Privacy Issues

The term "privacy" was first defined in 1890 by the renowned American lawyer S. Warren and Supreme Court Justice L. Brandeis as "the right to be let alone," marking the formal entry of "privacy" as an academic concept into scholarly research [1]. After World War II, personal privacy attracted international attention. The United Nations Declaration of Human Rights, adopted by the UN Security Council in 1948, proclaimed that "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation," establishing privacy as a fundamental right of citizens and further drawing scholars' attention to privacy protection issues [2]. In 1964, U.S. Supreme Court Justice H. Samuel and scholar H. George published the book *The Right of Privacy*, systematically examining the origin and scope of privacy rights, proposing that privacy rights could be traced back

to ancient Jewish law, and citing legal cases from Europe and federal states to define the scope of privacy rights [3].

Although early privacy research did not yet involve the field of government data opening, its discussions on the concept of privacy and privacy rights undoubtedly laid a solid theoretical foundation for subsequent research on personal privacy protection in government data opening.

1.2 Preparatory Development Stage (1966-2008): Personal Privacy Protection in Government Information Disclosure

In 1966, the United States enacted the Freedom of Information Act, stipulating citizens' right to access government information and government agencies' obligation to provide information to the public [4]. Thereafter, government information disclosure, as a development trend in government administration, gradually became a breeding ground for privacy risks. In response, the United States introduced the Privacy Act in 1974, restricting administrative agencies from disclosing personal privacy in the process of opening government information [5], providing legal protection for personal privacy protection in government information disclosure. Meanwhile, scholars such as E. Volokh and F. H. Cate joined the ranks of research on personal privacy protection in government information disclosure [6-7].

During this research stage, personal privacy protection in the context of government information disclosure had a relatively narrow definition of privacy data scope and limited research topics, perspectives, and methods. On the one hand, the object of personal privacy protection was considered to be "sensitive data." In 1993, P. Kumaraguru and L. F. Cranor first used the "medical sensitivity index" to describe the degree of privacy concerns in medical contexts [8], making them among the earliest scholars to focus on privacy sensitivity. Subsequent research consistently defined sensitive data as particularly special information that individuals are more reluctant to share with others [9], such as "intimate information" [10] and "highly intrusive data" [11]. In 1995, the European Union' s Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data specified that sensitive data includes "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and personal medical or sexual life" [12]. On the other hand, scholars' research topics mainly focused on the legal system, data management, and public administration of personal privacy protection in government information disclosure, with disciplinary perspectives concentrated in social sciences such as law, information science, and management, and research methods primarily using qualitative approaches like literature analysis and case studies [13-14]. For example, D. H. Flaherty used British Columbia' s Freedom of Information and Protection of Privacy Act in Canada as a case study to demonstrate how legal system design could resolve conflicts between information disclosure and privacy protection [15].

1.3 Key Advancement Stage (2009-Present): Personal Privacy Protection in Government Data Opening

In 2009, the Obama administration in the United States issued the Open Government Directive, requiring government-held data to be opened to society to a greater extent, in a broader scope, and across more fields, thus launching the global government data opening movement. Since the opening of "data" is far more likely to trigger privacy security and risk issues than the disclosure of "information," government data opening inevitably poses intense shocks and new challenges to the original personal privacy protection system.

In this new context of government data opening, this research stage has expanded the scope of privacy data, with research topics, disciplinary perspectives, and methods showing rich, broad, and diverse characteristics. First, the object of personal privacy protection has expanded from sensitive data to all data that can identify individuals or have the potential to identify individuals. With the emergence of data analysis technologies such as big data and cloud computing, even non-sensitive or even anonymized personal data can be re-identified to infringe upon personal privacy [16]. In 2016, the European Union' s General Data Protection Regulation defined personal data as "any data relating to an identified or identifiable natural person," with a scope including but not limited to "name, home address, electronic information, identity document number, location information, IP address, browsing record IP, mobile device identifier, and data held by hospitals or doctors" [17]. Second, scholars have begun to explore topics such as anonymization technology, data algorithms, and information technology for personal privacy protection in government data opening, with disciplinary perspectives achieving integration between social sciences (law, information science, management) and natural sciences (computer science, communications), and research methods beginning to employ empirical analysis such as experimental design and survey research [18-20]. For example, Professor N. Kshetri of the University of North Carolina at Greensboro verified through survey research that five elements of data—volume, velocity, variety, variability, and complexity—affect personal privacy protection in government data opening [21].

2. The Concept of Personal Privacy Protection in Government Data Opening

2.1 The Evolution of the "Privacy" Concept

Clearly and accurately defining the concept of "privacy"is the logical prerequisite for research on personal privacy protection in government data opening. Academic understanding of the privacy concept has undergone an evolution from significant early divergence to gradual later consensus. In this process, British and American scholars have been pioneers and main contributors, offering many insightful theories.

Before the 21st century, scholars held two different views on the concept of privacy: value-based and behavioral/institutional. Some scholars viewed privacy as a value concept. For instance, British literary scholar S. George noted that "privacy is primarily a modern, Western bourgeois value" [22]; renowned Oxford philosopher T. Charles defined privacy as an inherent characteristic of how individuals in Western society conceptualize self-identity [23]; and Assistant Professor S. Carl of Meadville Lombard Theological School combined privacy with a sense of shame, arguing that privacy is the sense of shame that protects the private sphere from exposure [24]. In contrast, other scholars, from different disciplinary perspectives, viewed privacy as behavioral paradigms and institutional agreements. For example, Professor A. F. Westin of Columbia University' s Department of Political Science defined privacy from a legal perspective as information self-determination as early as 1967, that is, "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [25]; I. Altman from the University of Utah' s Department of Psychology defined privacy from a psychological perspective as a selective control mechanism for individuals or groups [26]; and scholar R. O. Mason from Southern Methodist University noted from a management perspective that privacy issues are the primary ethical concerns in information management [27].

After the 21st century, with the rise and gradual stabilization of privacy research, researchers have continuously reached consensus on the concept and connotation of privacy, generally recognizing privacy as an institutional product of human society. As researcher D. J. Solove of the University of Aberdeen stated: "Privacy is not a natural thing, but an institution built by humans based on a series of paradigms and laws" [28]. Regarding the understanding of privacy rights, the vast majority of scholars agree with A. F. Westin' s view that privacy refers to natural persons' right to control personal information [29, 30]. Some scholars have also proposed new perspectives that privacy is a group right rather than an individual right, such as Oxford researcher L. Floridi, who argued that "privacy is a group right, a right owned by the group as a whole, not a right owned by each member individually" [31].

2.2 The Rationale for Personal Privacy Protection in Government Data Opening

Why should personal privacy in government data opening be protected? Scholars have offered three different explanations: the "legal rights theory," the "harm avoidance theory," and the "openness guarantee theory."

(1) **Legal Rights Theory.** Some scholars, from a legal perspective, point out that protecting citizens' rights is the legal basis for personal privacy protection in government data opening. However, regarding the specific content of citizens' rights, different scholars hold different views. Some believe that the rights foundation for personal privacy protection in government data opening is citizens' privacy rights. For example, T. Jaatinen,

Secretary of the Finnish Data Protection Board, noted that the right to protect personal privacy data is a fundamental right stipulated in the Charter of Fundamental Rights of the European Union, and personal privacy data must be processed fairly and lawfully [32]. Others believe the rights foundation is freedom of speech, such as American scholar B. C. Newell, who argued that privacy is necessary to ensure that people can freely express their opinions and take actions within the scope permitted by legislation and social norms [33]. Still others believe the rights foundation includes both citizens' privacy rights and freedom of speech, such as American scholar D. H. Flaherty, who cited British Columbia' s Freedom of Information and Protection of Privacy Act in Canada, pointing out that maintaining and realizing the public' s privacy and information freedom rights and interests are the dual goals of public institutions in protecting personal privacy [15].

(2) **Harm Avoidance Theory.** Some scholars, from the perspective of personal privacy protection goals and functions, believe that avoiding privacy harm is the practical need for personal privacy protection in government data opening. In the view of R. Calo, Senior Research Fellow at Stanford University' s Center for Internet and Society, privacy harm caused by government data opening can be divided into subjective and objective aspects: subjective harm is reflected in individuals' lack of control over their own information, while objective harm is reflected in the accidental or coercive use of information about a person [34]. Guo Weijia and Guo Shaoyou expressed the public' s privacy harm concerns as: "If people suspect that their personal data will be stored in public institutions, they may be less willing to contact public sector agencies, especially those with questions about disease, pregnancy, drugs, economic difficulties, or suicidal thoughts, who may not seek help"[35]. Further, Professor T. P. Keenan of the University of Calgary pointed out that one adverse consequence of privacy infringement is the improper use of data and the generation of unfairness: "Personal data opened by the public sector may be reused by data brokers for new purposes, resulting in unfairness" [36]. It is precisely based on these potential harms to personal privacy that may be caused by government data opening that scholars believe measures are necessary to protect personal privacy in government data opening.

(3) **Openness Guarantee Theory.** Many scholars, from the perspective of the government data opening process and its benefits, believe that implementing personal privacy protection is an essential requirement for ensuring the smooth implementation of government data opening and improving its comprehensive benefits. It is not difficult to understand that if government data opening infringes upon citizens' personal privacy, it will greatly damage the benefits of government data opening [37-38]. In this regard, researchers M. S. Bargh et al. from the Ministry of Security and Justice in The Hague, Netherlands, pointed out that personal privacy risks in government data opening may have a contractionary effect

on government transparency, leading to citizens potentially losing trust in government [39]. Huang Ruhua and Liu Long also noted that personal privacy security issues are related to public recognition and acceptance of government data opening; if personal privacy security cannot be properly protected, government data opening will be difficult to sustain [40]. In summary, only by effectively protecting citizens' privacy in government data opening can the process be smooth and the benefits optimal.

3. Logical Paradoxes and Practical Dilemmas of Personal Privacy Protection in Government Data Opening

Scholars have noted that although personal privacy protection has value and necessity in government data opening, it inevitably falls into two embarrassing situations: first, logical conflicts and paradoxes with data opening; and second, multiple dilemmas in practice.

3.1 The Logical Paradox Between "Data Opening" and "Privacy Protection"

Logically, there is an irreconcilable tension between "data opening" and "privacy protection," meaning that simultaneously achieving both is almost impossible. Scholars have mainly explained this from two perspectives.

First, improving data utility comes at the cost of sacrificing data privacy. Oxford researcher B. Zevenbergen et al. believe that data utility and data privacy are usually inversely proportional; controlling privacy leakage risk is typically achieved by significantly reducing data utility, while a small increase in data utility often requires disclosing more personal information [41]. M. Andrei of Delft University of Technology in the Netherlands pointed out that releasing raw data means maximum data utility and no privacy, while not releasing data means complete personal privacy protection and zero data utility [20].

Second, the combinability and memorability characteristics of data are the root causes of the logical paradox between data opening and personal privacy protection. Dutch scholar M. S. Bargh et al., based on the combinability characteristic of data, pointed out that open data may not appear to be personal data on the surface, but by combining it with other publicly available data, it may become personal data and lead to privacy leakage. To prevent privacy leakage, it is necessary to eliminate privacy-sensitive attributes, but this may negatively affect the usability of open data and damage the benefits of government data opening [39]. Tian Xinling and Huang Xiaozhi, based on the memorability characteristic of data, noted that big data has memory; even if data is deleted from user data graphs during the data opening process, it cannot guarantee that personal data will not be reused [42].

3.2 Practical Dilemmas of Personal Privacy Protection in Government Data Opening

Following the thread of "privacy scope-privacy risk-privacy protection," scholars have identified three practical dilemmas of personal privacy protection in government data opening: difficulty in defining privacy scope, easy occurrence of privacy risks, and difficulty in achieving privacy protection, with emphasis on explaining the reasons behind each dilemma.

3.2.1 Difficulty in Defining Privacy Scope Due to the blurred mixture of public and private data, the ease of re-identifying data, and the involvement of special groups, defining the scope of personal privacy data in government data opening is quite difficult. Tian Xinling and Huang Zhixiao, analyzing from the perspective of data storage methods, believe that privacy data and public data are blurred and mixed in the cloud, making it difficult to determine what belongs to the scope of personal privacy protection [42]. Zou Dongsheng, from the perspective of the data identification process, pointed out that the complexity of defining personal privacy lies in the fact that even if government-opened data initially contains no personal privacy or personal information has been anonymized, individuals can be re-identified by linking this open data with other data, thereby threatening personal privacy [43]. I. Szekely of Central European University in Hungary, taking the personal privacy definition of civil servants as an example, pointed out that as public officials, the boundary between their personal privacy and information disclosure is difficult to determine, and civil servants' privacy changes with position promotion and working time changes [44].

3.2.2 Easy Occurrence of Privacy Risks Some scholars have analyzed the reasons for the easy occurrence of personal privacy risks in government data opening from the perspectives of technology and data elements. G. Navarro and V. Torra from the Artificial Intelligence Research Institute at the Autonomous University of Barcelona, Spain, pointed out that re-identification technology can re-identify individuals using government-opened de-identified personal data through multi-channel information acquisition and correlation analysis [45]. Professor N. Kshetri of the University of North Carolina at Greensboro believes that data volume, velocity, variety, variability, and complexity are all factors leading to privacy leakage [21]. Chen Mei noted that factors that may cause personal privacy risks in urban government data opening include directly identifiable data, indirectly identifiable data, metadata, address data, geographic coordinates, unstructured data, and subsets of sensitive data [46].

Other scholars have discussed the occurrence mechanisms of personal privacy risks in government data opening from the perspectives of data collection, data acquisition, and data use processes. T. Jaatinen, Secretary of the Finnish Data Protection Board, believes that the lack of individual autonomous choice rights in the data collection process can easily lead to privacy risks: "Usually, data

subjects do not have a real choice about whether to disclose their personal data to public sector agencies" [32]. S. Gupta and P. Kumaraguru from the Indian National Institute of Technology believe that the diversity of data acquisition channels is an important cause of privacy risks [47]. Australian scholars Z. Pingo and B. Narayan believe that using information data for purposes other than approved purposes during data use also leads to privacy risks [48].

3.2.3 Difficulty in Achieving Privacy Protection From the institutional, organizational, technical, and literacy perspectives, scholars have identified several reasons why personal privacy protection is difficult to achieve in government data opening.

From the institutional design perspective, scholars believe that the lag and imperfection of personal privacy protection systems are the core reasons for the difficulty in achieving personal privacy protection in government data opening. Specifically, the lag and imperfection of personal privacy protection systems are manifested in the lack of policy regulations for cross-border data flows [49], unclear definition of personal privacy concepts and scope [50], and inapplicability of personal privacy protection principles [31], among other aspects. Some scholars, addressing the difficulties of personal privacy protection in China' s government data opening, pointed out that "the rights and responsibilities that relevant stakeholders should enjoy in China' s open government data are not clearly defined, and the concept and scope of personal privacy are not clearly defined, which seriously affects the process and level of personal privacy protection in China" [51].

From the technical capability perspective, scholars have found that technical security risks and insufficient technical capabilities are important causes of difficulties in personal privacy protection in government data opening. Researcher P. T. Jaeger et al. from Florida State University demonstrated from three aspects—information source, channel, and destination—that information technology progress has made personal privacy protection more difficult than before: in terms of information source, technological development provides more convenient means for collecting public information; in terms of information channel, technology makes information dissemination more convenient; and in terms of information destination, technology makes information utilization methods and subjects more complex than before [52]. Professor N. Kshetri of the University of North Carolina at Greensboro pointed out that governments may lack the technical capability to securely store large amounts of data and manage unstructured data during peak data traffic periods, thus causing privacy leakage risks [21].

From the psychological and cultural perspective, scholars believe that public individuals' privacy concepts, privacy disclosure psychology, and privacy infringement prevention capabilities largely affect the achievement of personal privacy protection in government data opening. B. Jacobs from the Institute of Computing and Information Sciences at Radboud University in the Netherlands

believes that the lack of privacy protection awareness is an important reason hindering personal privacy protection. He pointed out that social network services provide people with ways to be seen online and share experiences in real time, and for many people, "the desire to be seen is stronger than the desire to protect private information" [53]. S. Trepte et al. from the Department of Media Psychology at the University of Hohenheim in Germany believe that one main cause of privacy risk occurrence is the public' s lack of emphasis on, basic cognition of, and correct application of privacy protection tools [54].

4. Implementation Pathways for Personal Privacy Protection in Government Data Opening

The ultimate purpose of theoretical research is to scientifically guide practice. Research on implementation pathways directly relates to the guidance of personal privacy protection practice in government data opening and thus has received widespread attention from the academic community. In general, scholars have mainly discussed implementation pathways for personal privacy protection in government data opening from four levels: institutional design, organizational setup, technical application, and literacy cultivation.

4.1 Strengthening Personal Privacy Protection Institutional Design

Institutional design is the first line of defense for personal privacy protection in government data opening. Scholars have focused their discussions on legal formulation, institutional content, and institutional tools.

Regarding legal formulation, scholars have generally provided three approaches: First, formulate national-level personal privacy protection laws. For example, Zhang Xiaojuan et al. believe that China should formulate a Personal Information Protection Law to improve the legal system for government data opening and personal privacy protection [55]. Second, while formulating national personal privacy protection laws, international personal privacy protection laws are also needed. For example, D. Banisar, Senior Legal Counsel for the Global Freedom of Expression Campaign, taking Nigeria' s government data opening personal privacy protection as an example, pointed out that the country has both the Nigerian Constitution, which stipulates citizens' right to privacy and protects their homes, communications, telephone conversations, and telegraph communications, and international laws such as the African Charter on the Rights and Welfare of the Child, which protects children' s privacy rights [56]. Third, conduct joint legislation on information data freedom and personal privacy protection. For example, I. Szekely of Central European University pointed out that information privacy and information freedom are interrelated and interdependent concepts, and the goal of personal privacy protection in government data opening can be achieved through joint or at least interrelated legislative models in these two fields [44].

Regarding institutional content, scholars have proposed viewpoints such as fol-

lowing personal privacy protection principles, implementing privacy data review, and improving website privacy policies. For example, M. S. Bargh et al. from the Ministry of Security and Justice in The Hague, Netherlands, taking the Dutch Privacy Protection Act as an example, pointed out that personal privacy protection principles include the finality principle, legality principle, proportionality principle, subsidiarity principle, transparency principle, and rights principle [39]. Similarly, Huang Ruhua and Liu Long proposed that China' s government data opening personal privacy protection policy content should include strict data review, formulation of unified data desensitization processing standards and norms, improvement of data opening standardization and normalization, and reduction of the possibility of privacy risk occurrence [41]. Furthermore, Feng Changyang, taking China' s government data opening portal as an example, required the formulation of relatively comprehensive website privacy policies, which need to include user information content, purpose and method, information collection, personal information protection, online comments, browsing information collection, website security, external links, disclaimer, and policy adjustment [57].

Regarding institutional tools, scholars have studied the policy tool of privacy impact assessment from the perspectives of assessment content, methods, and implementation. Regarding privacy impact assessment content, S. Gupta and P. Kumaraguru from the Indian National Institute of Technology believe it includes multiple dimensions such as degree of harm, repeatability, usability, degree of impact, and discoverability [47]. Regarding privacy impact assessment methods, Dutch scholar M. Andrei believes that threshold rules, sample weighting, heuristics, and record linkage can be used [20]. Regarding privacy impact assessment implementation, Huang Ruhua and Liu Long believe that a privacy impact analysis mechanism should be established throughout the entire data lifecycle, strengthening privacy impact judgment and review at each stage of government data collection, processing, organization, opening, and use [40].

4.2 Establishing Personal Privacy Protection Organizations

Although scholars agree that organizational setup should be an implementation pathway for personal privacy protection in government data opening, there is considerable controversy over whether to establish independent personal privacy protection agencies and whether inter-departmental relationships should be cooperative or independent.

Regarding the establishment of personal privacy protection agencies, most scholars believe that specialized personal privacy protection agencies should be established. For example, P. T. Jaeger et al. from Florida State University proposed that in the context of government data opening, the government should adopt a centralized management agency to protect personal privacy [52]. Huang Ruhua and Liu Long also believe that China should establish a specialized agency for personal privacy protection, which should be legally independent and not subject to the constraints of general government departments, being only responsible to

a specific agency [58]. Different from the above views, some scholars believe that establishing an independent personal privacy protection agency is not realistic at China' s current stage, and the responsibility for personal privacy protection could be assigned to existing open data security management agencies [47].

Regarding inter-departmental relationships, most scholars believe that a cross-functional departmental government data opening personal privacy protection mechanism should be established. For example, some scholars believe that co-operation and communication between different departments and agencies at the same level of government help improve the effectiveness of personal privacy protection in government data opening [59]. However, some scholars believe that centralized data management in government data opening activities poses security risks and that information silos should be recreated. For example, M. Janssen and J. Hoven from the Faculty of Technology, Policy and Management at Delft University of Technology in the Netherlands believe that past efforts mainly focused on dismantling silos, but now silos should be recreated to prevent gradual information concentration and establish mechanisms for protecting personal privacy [60].

4.3 Applying Advanced Personal Privacy Protection Technologies

In research on technical pathways for personal privacy protection in government data opening, anonymization technology is a widely favored technology type. For example, scholars represented by Research Associate C. Perera of the UK' s Open University have proposed numerous anonymization strategies applicable to personal privacy protection in government data opening, including minimization design, pseudonymization, accepting transparency and openness, hiding, separation, aggregation, and control [61].

However, with the transformation and deepening development from government information disclosure to government data opening in the big data era, the increasing possibility of data re-identification may render anonymization technology insufficient for personal privacy protection. Therefore, many scholars have begun to research the application of information entropy, blockchain, and other emerging technologies in personal privacy protection in government data opening. For example, S. H. Kim et al. from the Korea Advanced Institute of Science and Technology in Daejeon proposed an information entropy-based personal privacy protection model through experimental design. This model uses the S-attribute set as a basic unit to measure organizational entropy and can obtain data entropy values from all organizational risks, achieving more accurate re-identification risk on a fine-grained basis, thereby protecting privacy while ensuring data usability [19]. Similarly, American scholar F. Bonomi et al. proposed fog computing technology for personal privacy protection, pointing out that fog computing is a paradigm that extends cloud computing and services to the network edge. Its unique fog characteristics—being close to end users, dense geographical distribution, and support for mobility—have the effect of reducing potential privacy-infringing behaviors [62].

4.4 Cultivating Citizens' Privacy Protection Literacy

Achieving personal privacy protection in government data opening is not solely the responsibility of the government as the subject, but also depends on the participation and cooperation of citizens as the objects of privacy protection. Among these, citizens' privacy protection literacy is of considerable importance. As Deng Shengli and Wang Ziye pointed out: "Privacy protection literacy is a key factor affecting privacy protection in government data opening. Improving citizens' privacy protection literacy plays an important role in privacy protection in government data opening" [63].

Regarding the issue of citizens' privacy protection literacy in government data opening, scholars have focused on discussing pathways for cultivating citizens' privacy protection literacy. Some scholars believe that improving citizens' privacy protection literacy mainly relies on education and training. For example, Australian scholars Z. Pingo and B. Narayan believe that education is important for helping citizens make reasonable and proactive decisions when disclosing personal information for various purposes, and can also help citizens make effective decisions on privacy issues [48]. Other scholars believe that citizens' autonomous learning is the main pathway to improving their privacy protection literacy. For example, Israeli scholar M. Weinberger et al. pointed out that most information about online privacy protection tools is now publicly available, and these tools are easy to learn and use. Users should consciously explore and learn methods and skills to protect their personal privacy [64].

In addition to the above four pathways, some scholars have proposed other approaches such as setting personal privacy, protecting group privacy, and collective action by stakeholders to achieve personal privacy protection in government data opening. For example, Researcher B. V. D. Sloot of the Institute for Information Law at the University of Amsterdam proposed that it is necessary to introduce personal privacy settings for third-party reuse of citizen data, allowing everyone to register their privacy settings with the government, and allowing citizens to independently choose data release objects, scope, and purposes, incentivizing them to join by providing them with a certain proportion of profits from third-party reuse of their personal data [65]. Similarly, Oxford researcher L. Floridi believes that individual privacy protection in government data opening should be achieved by protecting group privacy, which is like "individual sardines may think that the fishing net is trying to catch it. In fact, the fishing net is trying to catch the entire shoal. Therefore, if you want to save the sardines, you need to protect the shoal. Sometimes, the only way to protect an individual is to protect the group to which the individual belongs" [31]. Furthermore, N. Kshetri of the Oxford Internet Institute proposed from a stakeholder perspective that "in the process of privacy protection in government data opening, the government must adopt reasonable regulatory measures and formulate higher privacy standards; enterprises should proactively formulate risk management systems and deeply understand the root causes of risks; citizens can exert pressure on the government through rights protection, requiring it to open and

use data legally and reasonably" [21].

5. Research Summary and Future Prospects

As can be seen from the above, personal privacy protection is a hot topic that has attracted close and sustained attention from domestic and international academic circles for a long time, and has particularly attracted more scholars' research and attention in the new context of government data opening. Among them, researchers such as M. Andrei and M. Janssen from Delft University of Technology' s Faculty of Technology, Policy and Management, L. Floridi and B. Zevenbergen from the University of Oxford' s Oxford Internet Institute, P. T. Jaeger from Florida State University' s College of Information Studies, and Huang Ruhua from Wuhan University' s School of Information Management are identifiable research forces in this field. Existing research has made concentrated discussions on the concept, rationale, logical paradoxes, practical dilemmas, and implementation pathways of personal privacy protection in government data opening, not only achieving relatively fruitful research results but also preliminarily forming a thematic research system on the "what,""why,""how,"and "what to do" of personal privacy protection in government data opening, as specifically shown in Figure 2 [Figure 2: see original paper].

Overall, existing research exhibits three characteristics: First, **transformative and continuous research periods.** Research in this field has experienced two periods—government information disclosure and government data opening —with inherent transformative and continuous relationships between them. Second, **interdisciplinary research perspectives.** The research has not only been discussed by social sciences such as information science, law, political science, and public management but has also attracted attention from natural sciences such as computer science and communications. Third, **consensus and divergence in research viewpoints.** There is basic consensus on aspects such as the rationale for privacy protection, the tension between privacy protection and data opening, and privacy protection dilemmas and pathways, but there remain significant controversies on issues such as the concept and connotation of privacy, the rights foundation of privacy protection, and the establishment of privacy protection agencies.

Existing research still has certain limitations and shortcomings: In terms of research topics, discussions on basic privacy issues (such as the meaning, characteristics, boundaries, and scope of privacy, as well as the occurrence conditions, evolution processes, and influencing factors of personal privacy risks in government data opening) are still insufficient, and the exploration of personal privacy protection issues is also relatively limited. In terms of research objects, the grasp of the systematic and complex nature of the research object—personal privacy protection in government data opening—is not yet deep enough, and systematic and complex thinking has not been effectively applied. In terms of research tools and perspectives, there is a lack of innovative introduction of relevant disciplinary theoretical tools, and research perspectives have not transcended

outcome-based and static perspectives. In terms of research methods, the research remains at the level of normative analysis, with obviously insufficient empirical research, which cannot meet the needs of scientific research.

Standing at the emerging background of government data opening, focusing on the current theoretical guidance needs for personal privacy protection in government data opening [67], and examining it with a broader, more comprehensive, and holistic research perspective, we believe that future domestic research in this field should focus on strengthening and breaking through the following four aspects:

(1) **Further expand the scope of research issues in personal privacy protection while focusing on key and major issues.** The next step in research needs to expand relevant research issues in personal privacy protection in government data opening, including the meaning, characteristics, and nature of privacy, the boundaries and scope of personal privacy in government data opening, the occurrence mechanisms, evolution processes, and influencing factors of personal privacy risks in government data opening. In addition, research issues worth exploring also include the impact, harm, and assessment of privacy, classification of privacy risks and problems, value balancing and conflict management in personal privacy protection, typical experiences and lessons in privacy management and protection, and systematic operational strategies for privacy protection mechanisms and institutions. Finally, for the many issues already studied and listed above, it is necessary to further identify key and major issues among them, such as the occurrence conditions and mechanisms of privacy problems and systematic strategies for privacy protection, and then invest substantial research energy and resources to achieve breakthroughs and realize value-focused research.

(2) **Grasp the complexity of personal privacy protection issues in government data opening and emphasize the application of systematic thinking in research.** The public nature of government data opening determines that personal privacy protection issues within it are complex, manifested in that the value orientation and goal positioning of personal privacy protection in government data opening are not singular and consistent (e.g., there are value orientations and goal positioning with tension between privacy protection and transparent openness), and the involved elements are comprehensive and complex (e.g., involving government systems, public policies, social needs, public power, and government responsibility). Therefore, when facing personal privacy protection issues in government data opening, researchers must comprehensively apply multidisciplinary knowledge from information science, political science, administrative science, management, law, etc., and simultaneously use systematic, complex, holistic, and relational thinking to strengthen the grasp of personal privacy protection issues in government data opening. For example, regarding the question of how to implement personal privacy

protection in government data opening, it is necessary to simultaneously consider the government, enterprises, social organizations, professional institutions, news media, and individual citizens in terms of subjects, and simultaneously consider data management mechanisms, administrative legal regulation mechanisms, administrative ethics guidance mechanisms, and government responsibility constraint mechanisms in terms of mechanisms.

(3) **Enrich and innovate the application of relevant theoretical tools, highlighting process and dynamic perspectives in research.** For example, privacy issues arise throughout the entire lifecycle of government data opening, so data lifecycle theory can undoubtedly provide an effective analytical framework for analyzing personal privacy protection in government data opening—that is, deeply analyzing the occurrence factors of privacy risks and then proposing prevention and resolution measures sequentially from each stage of the data lifecycle. Similarly, the issue of personal privacy protection in government data opening is actually a process of game-playing among three main stakeholders—the government, enterprises, and citizens—so stakeholder theory and game theory can both provide appropriate theoretical foundations and offer new perspectives for explaining and solving issues such as privacy risk occurrence and privacy protection realization. While enriching and innovating the application of relevant theoretical tools, this field should also shift more toward and promote research from process and dynamic perspectives to compensate for the current research bias toward outcome-based and static perspectives. For example, strengthen process research on the inducing factors and formation mechanisms of personal privacy risk generation in government data opening, and strengthen dynamic research on personal privacy management and protection mechanisms in government data opening.

(4) **Introduce analytical methods such as historical analysis and comparative analysis, and simultaneously strengthen the conduct of empirical research such as case studies.** Privacy protection is a long-standing issue that contains certain experiences and lessons, as well as specific evolution and development patterns, over a long historical period. At the same time, privacy protection is a common issue faced by governments of different countries, regions, and levels, and privacy protection has necessarily accumulated practices full of commonalities and differences among these different governments. This means that historical analysis and comparative analysis have good applicability for current research on personal privacy protection in government data opening and should be introduced and supplemented as important methods for current normative analysis in this field. Moreover, because personal privacy protection is highly practical, empirical research on personal privacy protection in government data opening can select case studies as an important empirical research method to change the current research situation that emphasizes normative analysis over empirical research.

For example, typical cases of personal privacy protection in government data opening at home and abroad can be explored to provide empirical support for theoretical propositions on personal privacy protection in government data opening.

References

[1] WARREN S, BRANDEIS L. The right to privacy [J]. Harvard law rev, 1890, 4(5): 193-220.
[2] The universal nations declaration of human rights [EB/OL]. [2019-05-06]. https://www.un.org/en/universal-declaration-human-rights/index.html.
[3] HOFSTADTER S H, HOROWITZ G. The right of privacy: with a reprint of "Right to Privacy" by Samuel D. Warren and Louis D. Brandeis [M]. New York: Central Book Company, 1964.
[4] What is FOIA? [EB/OL]. [2019-05-06]. https://www.foia.gov/about.html.
[5] The privacy act of 1974 [EB/OL]. [2019-05-06]. https://www.archives.gov/about/laws/privacy-act-1974.html.
[6] VOLOKH E. Freedom of speech and information privacy: the troubling implications of a right to stop people from speaking about you [J]. Stanford law review, 2000, 52(5): 1049-1124.
[7] CATE F H, FIELDS D A, MCBAIN J K. The right to privacy and the public' s right to know: the "central purpose" of the freedom of information act [J]. Administrative law review, 1994, 46(1): 41-74.
[8] KUMARAGURU P, CRANOR L F. Privacy indexes: a survey of Westin' s studies [EB/OL]. [2019-05-06]. http://www.cs.cmu.edu/~ponguru/CMU-ISRI-05-138.pdf.
[9] WANG M. Value convergence and cultural divergence: a comparative study of sensitive data perception between Chinese and American "millennial" college students [J]. Journalism & communication review, 2018, 71(2): 28-41.
[10] NISSENBAUM H. Protecting privacy in an information age: the problem of privacy in public [J]. Law and philosophy, 1998, 17(5): 559-596.
[11] PESCIOTTA D T. I' m not dead yet: katz, jones, and the fourth amendment in the 21st century [J]. The case western reserve law review, 2012, 63(2): 187-255.
[12] Data protection directive, article 4 [EB/OL]. [2019-05-06]. http://ec.europa.eu/justice/policies/privacy/doc
46-ce/dir1995-46_{{{part1}}{{en}}}.pdf.
[13] COMEAU P A, OUIMET A. Freedom of information and privacy: Quebec' s innovative role in north america [J]. IRB, 1995, 23(4): 13-21.
[14] SHEINKOPF C M. Balancing free speech, privacy and open government: why government should not restrict the truthful reporting of public record information [J]. Ucla L. rev, 1997, 44(5): 1567-1611.
[15] FLAHERTY D H. Balancing open government and privacy protection [EB/OL]. [2019-05-06]. http://www.austlii.edu.au/au/journals/PrivLawPRpr/1999/56.html.
[16] MACHANAVAJJHALA A, REITER J P. Big privacy: protecting confidentiality in big data [J]. XRDS: crossroads, the ACM magazine for students, 2012, 19(1): 20-23.

[17] The general data protection regulation [EB/OL]. [2019-05-06]. http://www.consilium.europa.eu/en/policies/data-protection-reform/data-protection-regulation/.
[18] ZHANG W. The essence and approach of personal data protection legislation [J]. Jiangxi social sciences, 2018, 38(6): 169-176.
[19] KIM S H, JUNG C, LEE Y J. An entropy-based analytic model for personal data privacy protection [EB/OL]. [2019-05-06]. https://arxiv.org/pdf/1312.2784.pdf.
[20] ANDREI M. Publishing privacy-sensitive open data [EB/OL]. [2019-05-06]. http://insy.ewi.tudelft.nl/sites/default/files/thesis{final}.pdf.
[21] KSHETRI N. Big data' s impact on privacy, security and consumer welfare [J]. Journal of theoretical and applied electronic commerce research, 2014, 9(3): 32-44.
[22] STEINER G. Literature and post-historicism [M]. London: Cato & Windus, 1971.
[23] TAYLOR C. The crisis of modernity [M]. Cambridge: Cambridge University Press, 1985.
[24] SCHNEIDER C D. Shame, exposure, and privacy [M]. Cambridge: Cambridge University Press, 2000.
[25] WESTIN A F. Science, privacy, and freedom: issues and proposals for the 1970s. Part I—the current impact of surveillance on privacy [J]. Columbia law review, 1966, 66(6): 1003-1050.
[26] ALTMAN I. Privacy regulation: Culturally universal or culturally specific? [J]. Journal of social issues, 1977, 33(3): 66-84.
[27] MASON R O. Four ethical issues of the information age [J]. MIS quarterly, 1986, 10(1): 5-12.
[28] SOLOVE D J. Understanding privacy [J]. Social science electronic publishing, 2008, 59(7): 57-58.
[29] GÁRSES F S. Multilateral privacy requirements analysis in online social networks [J]. Journal of vibration & shock, 2010, 27(9): 139-141.
[30] GUTWIRTH S, POULLET Y, HERT P D, et al. Reinventing data protection? [J]. Identity in the information society, 2010, 2(6): 673-681.
[31] FLORIDI L. Open data, data protection, and group privacy [J]. Philosophy & technology, 2014, 27(1): 1-3.
[32] JAATINEN T. The relationship between open data initiatives, privacy, and government transparency: a love triangle? [J]. International data privacy law, 2016, 6(1): 28-38.
[33] NEWELL B C. Technopolicing, surveillance, and citizen oversight: a neorepublican theory of liberty and information control [J]. Government information quarterly, 2014, 31(3): 421-431.
[34] CALO R. The boundaries of privacy harm [J]. Social science electronic publishing, 2011, 86(3): 1131-1162.
[35] GUO W, GUO S. Research on personal privacy risk types and resolution mechanisms in open government data [J]. Henan science and technology, 2018(4): 21-24.
[36] KEENAN T P. Are they making our private public?—Emerging risks of governmental open data initiatives [C]//IFIP prime life international summer school on privacy and identity management for life. Berlin: Springer, 2011:

1-13.
[37] JANSSEN M, CHARALABIDIS Y, ZUIDERWIJK A. Benefits, adoption barriers and myths of open data and open government [J]. Government information quarterly, 2012, 29(4): 258-268.
[38] RONALD M, PETER C, SUNIL C. Reconciling contradictions of open welfare [M]. Oxford: Pergamon Press, 2014.
[39] BARGH M S, CHOENNI S, MEIJER R. On addressing privacy in disseminating judicial data: towards a methodology [J]. Transforming government: people, process and policy, 2017, 11(1): 9-41.
[40] HUANG R, LIU L. Issues and countermeasures of personal privacy protection in China' s government data opening [J]. Library, 2017(10): 1-5.
[41] ZEVENBERGEN B, BROWN I, WRIGHT J, et al. Ethical privacy guidelines for mobile connectivity measurements FINAL v2.docx [EB/OL]. [2019-05-06]. http://dataethics.github.io/proceedings/EthicalPrivacyGuidelinesforMobileConnectivityMeasurements.pd
[42] TIAN X, HUANG Z. The paradox between "public data opening" and "personal privacy protection" [J]. Journalism research, 2014(6): 55-61.
[43] ZOU D. Government open data and personal privacy protection: the example of Canada [J]. Chinese public administration, 2018(6): 75-82.
[44] SZEKELY I. Freedom of information versus privacy: friends or foes? [M]. Berlin: Springer Netherlands, 2009.
[45] NAVARRO-ARRIBAS G, TORRA V. Data privacy: a survey of results [J]. Studies in computational intelligence, 2015, 10(3): 27-37.
[46] CHEN M. Privacy risks and technical control strategies in urban government open data [J]. Library development, 2018, 290(8): 16-21, 27.
[47] GUPTA S, KUMARAGURU P. OCEAN: Open-source collation of e-government data and networks—understanding privacy leaks in open government data [EB/OL]. [2019-05-06]. https://arxiv.org/pdf/1312.2784.pdf.
[48] PINGO Z, NARAYAN B. When personal data becomes open data: an exploration of lifelogging, user privacy, and implications for privacy literacy [EB/OL]. [2019-05-06]. https://link.springer.com/content/pdf/10.1007%2F978-3-319-49304-6_1.pdf.
[49] BAUMER D L, EARP J B, POINDEXTER J C. Internet privacy law: a comparison between the United States and the European Union [J]. Computers & security, 2011, 23(5): 400-412.
[50] CHEN M. Research on personal privacy protection in Germany' s open government data [J]. Library, 2018(8): 11-16.
[51] HUANG R, LI N. Research on personal privacy protection in U.S. open government data [J]. Library, 2017(6): 19-24, 76.
[52] JAEGER P T, MCCLURE C R, FRASER B T. The structures of centralized governmental privacy protection: approaches, models, and analysis [J]. Government information quarterly, 2002, 19(3): 317-336.
[53] JACOBS B. Two of the grand changes through computer and network technology [M]. Berlin: Springer, 2013.
[54] TREPTE S, TEUTSCH D, MASUR P K, et al. Do people know about privacy and data protection strategies? towards the "online privacy literacy scale" (OPLIS) [M]. Berlin: Springer Netherlands, 2015.

[55] ZHANG X, WANG W, TANG C. Research on policies and regulations of government data opening and personal privacy protection in China and the United States [J]. Information studies: theory & application, 2016, 39(1): 38-43.
[56] BANISAR D. Linking ICTs, the right to privacy, freedom of expression and access to information [J]. Social science electronic publishing, 2010, 16(1): 124-154.
[57] FENG C. Comparative study on privacy policies of government open data portals [J]. Digital library forum, 2016(7): 52-56.
[58] HUANG R, LIU L. Research on personal privacy protection in UK government data opening [J]. Library development, 2017(10): 47-52.
[59] Strategies for protecting privacy in open data and proactive disclosure | Scassa | Canadian journal of law and technology [EB/OL]. [2019-05-06]. https://ojs.library.dal.ca/CJLT/article/view/8480/7286.
[60] JANSSEN M, HOVEN J V D. Big and open linked data (BOLD) in government: a challenge to transparency and privacy? [J]. Government information quarterly, 2015, 32(4): 363-368.
[61] PERERA C, RANJAN R, WANG L. End-to-end privacy for open big data markets [J]. IEEE cloud computing, 2015, 2(4): 44-53.
[62] BONOMI F, MILITO R, ZHU J, et al. Fog computing and its role in the internet of things [EB/OL]. [2019-05-06]. http://conferences.sigcomm.org/sigcomm/2012/paper/mcc/p13.pdf.
[63] DENG S, WANG Z. A review of foreign research on online privacy literacy [J]. Digital library forum, 2018(9): 66-72.
[64] WEINBERGER M, ZHITOMIRSKY-GEFET M, BOUHNIK D. Factors affecting users' online privacy literacy among students in Israel [J]. Online information review, 2017, 41(1): 655-671.
[65] SLOOT B V D. On the fabrication of sausages, or of open government and private data [J]. Social science electronic publishing, 2013, 3(2): 136-151.
[66] LI Y. Research on information privacy protection issues in the big data era [J]. Henan social sciences, 2017, 25(4): 67-73, 124.
[67] HOU S. Legal regulation of data and information collection in the big data era [J]. Party and government research, 2018(2): 22-28.

Author Contributions:

Chen Chaobing: Designed the paper framework, proposed the research plan, and was responsible for paper writing and revision.

Hao Wenqiang: Collected and sorted literature, wrote the first draft, and participated in later revisions.

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