Personal Information Protection in the Big Data Era: A Research Review (Postprint)

Authors: Panpan Jiang

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Through a systematic review of research literature on personal information protection in the big data era, this study provides references for future research in this domain.

[Method/Process] Utilizing the CNKI database, we systematically collected domestic core journal papers on personal information protection in the big data era, analyzed their content, summarized research themes, and conducted detailed discussions.

[Results/Conclusion] Existing research findings are classified into five thematic categories: domestic and international practical experiences, fundamental theoretical issues, personal information protection mechanisms, personal information security risks and countermeasures, and relevant laws and regulations. We review the limitations of existing research and propose directions for future research.

Full Text

Preamble

Vol. 63 No. 15, August 2019
A Review of Personal Information Protection Research in the Big Data Era
Jiang Panpan
School of Law, Jilin University, Changchun 130012

Abstract

[Purpose/Significance] This study systematically reviews the literature on personal information protection in the big data era to provide a reference for future research in this field.

[**Method/Process**] Using the CNKI database, we systematically collected domestic core journal papers on personal information protection in the big data era and analyzed their content to identify and discuss research themes.
[**Result/Conclusion**] Existing research findings were categorized into five thematic areas: studies on domestic and international practical experiences, basic theoretical issues, exploration of personal information protection mechanisms, personal information security risks and countermeasures, and relevant laws and regulations. The paper reviews the shortcomings of current research and proposes directions for future inquiry.

Keywords: Personal Information; Big Data; Privacy; Legal Attributes

2 Content Analysis of Literature

2.1 Research on Domestic and International Practical Experience

Research on personal information protection in China has been significantly influenced by legislation such as the U.S. Privacy Act, the EU General Data Protection Regulation (GDPR), Germany's Federal Data Protection Act, and Japan's Personal Information Protection Act. From the perspective of rights evolution, privacy protection in the United States and Germany has transitioned from traditional privacy domain theories to information privacy and information autonomy, with "information autonomy" or "information self-determination" once forming the theoretical foundation for Chinese scholars discussing personal information protection. The EU GDPR, which took effect on May 25, 2018, has provided substantial inspiration for Chinese research on personal information protection in the big data era, particularly regarding its provisions on personal data definitions, the right to be forgotten, the right to erasure, data portability, and the data protection officer system, offering valuable references for improving China's personal information protection legislation. In recent years, Chinese scholars' research focus has evolved from introducing foreign personal information protection legislation to proposing adaptations based on China's national conditions, and ultimately to 本土化 theories, protection mechanisms, legislative models, rights attributes, and value foundations—achieving a shift from "macro to micro" and "concrete to abstract."

Specific issues examined in studies of foreign practical experiences within the big data context include:

2.1.1 Comparative Research on Rights Evolution The Decision of the Standing Committee of the National People's Congress on Strengthening Network Information Protection (hereinafter referred to as the "Decision") was implemented on December 28, 2012. The Decision states that "the state protects electronic information that can identify citizens' personal identities and involves citizens' personal privacy." However, it does not explicitly define "personal electronic information," creating practical challenges for protecting such information in the big data era, where personal portraits in video surveillance

systems, basic personal information held by social management or service departments, online traces, and computer data are all at risk of leakage. Chen Xuan et al. [1] argue that personal electronic information constitutes privacy. Through comparative analysis of the evolution of privacy protection from traditional privacy to information privacy and information autonomy in the United States and Germany, they ultimately define privacy rights as encompassing both private domains and information autonomy, while opposing the establishment of an independent information autonomy right beyond privacy rights. For the protection of personal electronic information, network service providers and other enterprises must adhere to principles of information autonomy and voluntariness while fulfilling corresponding confidentiality obligations.

2.1.2 Comparative Research on Personal Information Protection Legislation Innovative big data technologies and cloud storage have created new threats to personal information protection, such as the involuntary uploading of personal information online, authorization issues regarding personal information use, and data tracking problems. Existing legislative frameworks cannot adequately meet the needs of personal information protection, necessitating reference to more comprehensive legislative approaches in other countries. Yang Yuan [2], after examining the EU's data protection legislative reforms and U.S. privacy protection regulations, suggests that China should accelerate the improvement of legislation related to internet personal information protection, promote the opening of personal information by governments and other institutions, and establish data application development platforms to encourage individual participation in open information applications. Ji Leilei [3] argues that personal information in the big data era possesses dual attributes of personality interests and property interests. Drawing inspiration from the EU's risk management approach to personal information protection in specific contexts, the U.S. approach of moving beyond the "informed consent" framework, and Russia's localization of personal information, she recommends that China's personal information protection legislation should introduce context-specific risk management, break free from the constraints of the informed consent framework, localize personal information, establish personal information protection regulatory agencies independent of the government, and adopt differentiated relief and compensation models. Wang Shaohui et al. [4] suggest that personal privacy protection in the big data era should draw lessons from New Zealand's personal privacy protection system, which features a comprehensive legal framework (including the Privacy Act), specialized privacy management agencies, and a privacy protection regulatory model. They recommend that China accelerate the introduction of the Personal Information Protection Law, develop cross-border privacy protection cooperation, and promote inter-departmental collaboration in privacy protection. Wang Min et al. [5], based on a comparative analysis of the latest privacy protection regulations in China and the United States, identify differences in quantity and detailed rules, as well as divergent philosophies—China emphasizes dignity while the United States emphasizes value. They propose

establishing ethical norms for legal exceptions and adopting a "small data" mindset to counter big data pollution.

Recent research on foreign legislative experience has shown two main characteristics: a shift from introducing the experience of a single country to presenting the experiences of multiple countries (e.g., from the EU GDPR and U.S. Privacy Act to the personal information protection legislation of the United Kingdom, Japan, New Zealand, Germany, and Russia), and a transition from macro-level introductions to micro-level examinations of specific topics (e.g., from improving the overall personal information protection legal system to refining specific principles such as informed consent and personal information protection regulatory models).

2.1.3 Comparative Research on Personal Information Security Certification Mechanisms for Software Products in the Big Data Era Computer software products represent an emerging strategic industry in the big data era, defined as products created by combining software works with physical materials or electronic information materials to meet people's daily needs. However, all stages of the software industry lifecycle—including development, production, and sales—can impact user personal information security, yet computer software product security certification mechanisms have not received adequate attention. Chen Xing [6] examined and compared the U.S. privacy certification system (initial certification, ongoing supervision, and dispute resolution), Japan's personal information protection evaluation system, and Dalian's personal information protection certification system in China, identifying several differences: (1) differences in evaluation objects—the U.S. system primarily targets online evaluations without industry restrictions, Japan's system applies to all enterprises (not limited to websites), while Dalian's system focuses mainly on software and information service industries with certain industry restrictions; (2) differences in evaluation scope—the U.S. system has developed into a global certification system, Japan's system is a national-level certification applicable only to Japanese enterprises (not foreign enterprises), while Dalian's system is limited to Dalian's software and information service industry but is gradually expanding nationwide. Overall, China's personal information security certification system should be constructed from two aspects: software product personal information security certification standards and certification processes, building a "firewall" for personal information protection in the big data era.

2.1.4 Domestic Practical Experience Research Domestic empirical research on personal information protection in the big data era primarily addresses the rights-based approach and its dilemmas, the current status of protective development and utilization of personal information, countermeasures, privacy concerns regarding personal data, and empirical studies on personal information security and privacy protection. Luo Jiao [7] employs comparative and case study methods to analyze the dilemmas of the single rights-based approach to personal information protection in the big data era, recommending a shift

away from this mindset toward establishing rights such as notification, decision-making, confidentiality, access, correction, deletion, transmission, and blocking, along with corresponding data management policies from the perspective of information control. Liu Yaqi [8] uses questionnaire surveys to analyze the demand for protective development and utilization of personal information in big data environments, identifying security obstacles, model obstacles, and imperfect legal systems as barriers to personal information development and utilization, and recommends improving the legal framework for personal information protection, constructing sound market models to ensure secure information circulation, and technically supporting legal protection systems and market operation models. Wang Zhong et al. [9] employ questionnaire surveys and in-depth interviews to analyze privacy concerns regarding personal data in the big data era, categorizing these concerns into three dimensions: collector credibility, personal awareness, and post-incident relief. They recommend raising privacy protection awareness, establishing reporting mechanisms, and improving relevant laws and regulations. Kuang Wenbo et al. [10] conduct an empirical analysis of personal information security and privacy protection from the perspective of innovation diffusion theory's application to big data, arguing that big data applications must consider human nature and focus on satisfying human needs. As domestic empirical research on personal information protection in the big data era remains at the stage of summarizing research status, future research should gradually move toward systematization and systematic development.

2.2 Basic Theoretical Issues

Research on basic theoretical issues concerning personal information protection in the big data era covers: conceptual differentiation among personal information, personal data, and personal privacy; analysis of the legal attributes of personal information; the informed consent principle in personal information protection; balancing interests between personal information protection and utilization; characteristics and attributes of personal data rights; and rights conflicts and legal regulations in big data ethics. Key points include:

2.2.1 Conceptual Differentiation Among Personal Information, Personal Data, and Personal Privacy Personal data, personal privacy, and personal information are closely related yet easily confused concepts, necessitating clear distinctions in research on personal information protection in the big data era.

First, the distinction between personal information and personal data. Mei Xiaying [11] argues that the scope of personal information is broader than that of personal data, which possesses dual attributes as both personal information ontology and personal information medium, thereby differing from personal information that must be separated from transmission media. Yu Chong [12] defines personal information as information that can directly or indirectly identify a specific individual's identity, while personal data refers to personal information

that has been databased and can be retrieved through computer searches to form personal databases. He contends that personal information is more controllable than personal data. Yang Weiqin [13] suggests that personal information emphasizes substantive content, whereas personal data emphasizes neutral formal carriers, representing a relationship between content and form. Chu Jiewang et al. [14] view personal data as a macro concept encompassing all messages, facts, and records, while personal information refers to personal data that has been screened and systematically arranged, essentially remaining personal data. Shi Weimin [15] notes that personal data primarily applies to technical fields, while personal information applies to legal fields. These studies clarify the essential identity between personal information and personal data: they represent a content-form relationship where personal information manifests its identifiable function through personal data as a carrier. In the internet era, big data has shifted personal information's social space to cyberspace, where personal data merely reflects personal information online. Consequently, the controllability of personal information has gradually weakened, and de-identification has blurred the conceptual distinction between personal information and personal data, effectively making them equivalent. This fundamental change requires a more dialectical view of their conceptual differentiation, as the increasing difficulty in distinguishing them leads to comprehensive personal information protection in the big data era, posing practical challenges to traditional protection approaches. The identity of personal information and data is dominated by cyberspace interactions, compelling us to emphasize the structural characteristics of personal information in cyberspace and adopt corresponding protection measures—a concern that warrants further attention in basic theoretical research on personal information protection in the big data era.

Second, the distinction between personal information and personal privacy. Wang Liming [16] argues that personal information embodies both personality and property interests, while personal privacy embodies only personality interests. Personal information emphasizes identity identifiability and relates to national security, whereas personal privacy emphasizes confidentiality and does not relate to national security. Personal information requires carriers, while personal privacy does not. Personal information employs comprehensive protection with emphasis on ex-ante prevention and remedies through both spiritual and property damage compensation, while personal privacy primarily employs legal protection focusing on ex-post relief through spiritual damage compensation only. Feng Yuan [17] views personal information as dynamic and open, without separating personal living space from external space, while personal privacy is static and closed, with personal living space being independent from and unassociated with external space. Zhou Dong [18] suggests that personal information and personal privacy have an intersecting relationship: some personal privacy manifests as personal information, and some personal information containing personal privacy components or content constitutes personal privacy, with confidential personal information belonging to the category of personal privacy. Wen Yanyan [19] argues that personal information has a broader scope

than personal privacy, personal privacy is more sensitive than personal information, and the consequences of infringing upon personal privacy are more severe than those of infringing upon personal information. Ling Pingping et al. [20] contend that personal information possesses objectivity, with only "identity-identifying" and "activity-reflecting" personal information entering the scope of criminal law, whereas personal privacy possesses subjectivity, requiring subjective judgment for criminal law protection. Criminal law protection of personal information is proactive, while protection of personal privacy is reactive. Han Xuzhi [21] argues that personal information emphasizes identification and information flow freedom, while personal privacy emphasizes confidentiality and restricts information flow. These distinctions between personal information and personal privacy are primarily based on "direct or indirect identifiability" for personal information and "confidentiality" for personal privacy. The General Principles of Civil Law places personal information protection in the chapter on civil rights, separately stipulating personal information rights and privacy rights, clarifying them as two distinct independent concepts—privacy rights under Article 110 as personality rights of civil subjects, and personal information rights under Article 111. This has objectively created the misconception that personal information rights are subordinate to privacy rights, leading to conceptual confusion. In summary, personal information and personal privacy have their own essential characteristics and cannot be conceptually equated.

After reviewing these conceptual distinctions, personal information in the big data context can be defined as: various information that can identify personal identity either alone or in combination with other information, and that may affect personal and property safety. Understanding personal information in the big data era requires dynamic and contextual approaches while adhering to the core criterion of "identifiability." The conceptual understanding should shift from the narrow "identifiability" to an open, result-oriented "harm risk" approach, representing a transition from formal to substantive understanding. Using "harm risk" as the main thread for understanding personal information can reasonably balance the interests between protection and utilization. Insisting on "identifiability" without considering "harm risk" leads to imbalanced interests. Therefore, personal information concepts should be understood dynamically, based on comprehensive protection principles, starting from the necessity and appropriateness of protection, and using substantive interpretation to bridge the gap between the ideal and reality of personal information protection.

2.2.2 Analysis of Legal Attributes of Personal Information Major theoretical perspectives on the legal attributes of personal information include: personality rights theory, privacy rights theory, personality rights plus property rights theory, ownership theory, new civil rights theory, and framework rights theory. Zhang Li'an et al. [22] view personal information rights as an independent personality right, possessing both the negative defensive aspect of personality rights and the positive aspect of directly controlling personal information due to its property value. Wang Xuehui et al. [23] consider personal

information essentially a form of privacy, with legal protection defining the scope of rights, suggesting the concepts of privacy and personal information can be used interchangeably. Zhang Ping [24] argues that the legal attributes of personal information should prioritize personality attributes while considering property attributes. Cheng Cheng [25] views personal information as having property attributes with dual aspects: protection emphasizes private interests, while utilization emphasizes public interests. Li Weimin [26] argues that personal information rights are neither constitutional rights, property rights, personality rights, nor intellectual property rights, but rather independent new civil rights, essentially private rights and another new civil right following equity and intellectual property rights. Ren Longlong [27] considers personal information neither a thing nor privacy, with personal information rights essentially being framework rights that serve a "catch-all" function.

After reviewing these perspectives, the legal attributes of personal information can be understood as follows: debates on legal attributes all treat personal information as a civil right, with the evolution from personality rights to new civil rights reflecting a conceptual shift from protection-only to protection-and-utilization. In the big data era, personal information is fully utilized, especially in commercial fields, where data mining and analysis generate economic and social value, benefiting the public. Consequently, the private rights attribute of personal information has weakened while its social attribute has strengthened. While personal information undoubtedly possesses both personality and property attributes, we cannot simply conclude that it is both. Instead, we should view the issue dialectically. The personality rights theory ignores property attributes; the privacy rights theory is inappropriate as personal information and privacy differ in connotation and extension; the personality rights plus property rights theory, though widely accepted, is problematic because not all personal information has personality attributes (e.g., merchants' personal information databases primarily have property attributes), and the ownership theory's treatment of personal information as civil law objects that can be possessed, used, disposed of, and profited from is also questionable. The new civil rights theory and framework rights theory fail to reveal the essence of personal information's legal attributes, remaining at the level of treating personal information as a civil right. Understanding should not be limited to private interests but must consider the public interest attributes of personal information in the big data era.

2.2.3 The Informed Consent Principle in Personal Information Protection The informed consent principle is fundamental to personal information protection, but in the big data era, it faces dilemmas. Large-scale processing, multi-party sharing, and uncertain purposes of personal information increase the difficulty of obtaining valid consent, challenging the principle's legitimacy. Tian Ye [28] analyzes the dilemmas of informed consent in the big data era using biobanks as an example, arguing that the principle must be reshaped to meet big data application needs. He proposes a new informed consent

principle that balances protection and utilization, shifting from holistic consent to layered consent based on information classification and specific contexts, and from one-time consent to continuous information disclosure and dynamic consent, allowing conditional broad consent plus opt-out rights. Lin Huanmin [29] argues that informed consent should not be formalistic but substantive, seeking balance between enterprises' utilization needs and individuals' expectability. He suggests not treating informed consent as the sole legitimate basis for enterprises processing personal information and introducing personal information anonymization to obtain exemptions from consent. Jiang Panpan [30] draws on EU legislative experience regarding consent in personal information protection law, recommending strict limitations on implied consent, adoption of explicit consent as the principle, substantive consent with a "four-element" effective consent model, and establishing effective verification mechanisms for minors' consent. These analyses reveal that while informed consent remains the cornerstone of personal information protection, significant room for improvement exists, particularly regarding substantive consent and consent method transformations, ultimately aiming to balance protection and utilization.

2.2.4 Balancing Personal Information Protection and Utilization Research on balancing protection and utilization covers: balancing rights between information subjects and controllers, balancing network service providers' interests with personal data protection demands, balancing personal information protection with big data rights attribution, and balancing relationships among individuals, information practitioners, and the state. Yu Xiaoyao [31] analyzes the balance among individual, information practitioner, and state rights from the perspective of author personal information protection and utilization in the big data era, proposing recommendations for balancing these three interests. Zhu Xinli et al. [32] use the economic concept of public goods to analyze the balance between personal data utilization and protection in the big data era, viewing the essence of this balance as defining data circulation scope and proposing a "resource access model": enterprises processing personal data must have information security capabilities and circulate information only among qualified enterprises; personal data is reversible and can become public goods for qualified enterprises; unified data resource platforms should be built with private rights rules, incentive rules, and administrative supervision rules. Zhang Shuqing [33] analyzes the boundaries between big data rights and personal information protection, summarizing a "three-step" rule for big data subject rights attribution: exclusive self-ownership → private law autonomy → proportionality principle. Zhao Lili et al. [34] propose solutions to conflicts between personal data protection and big data services: establishing rules for exchanging personal data for big data services, distinguishing information types under big data applications, and refining rules for personal data reuse lifecycle. Jiang Bo et al. [35] suggest that principles for reasonable use of personal information in the big data era should draw on copyright law's fair use principle, incorporating data security management, personal information risk assessment, and individual participation and

control rules. Wang Yulin [36] analyzes the rights balance between information controllers and subjects, discussing definitions, classifications, legal attributes, and rights and restrictions of personal information in the big data era. This review reveals that in the big data era, the public interest attribute of personal information has become increasingly prominent. As a data asset, personal information should be mined for value while ensuring protection, strengthening free flow and improving balance rules between protection and utilization.

2.3 Exploration of Personal Information Protection Mechanisms

Research themes include traceability mechanisms, governance mechanisms, transaction licensing mechanisms, and reporting mechanisms for personal information protection.

2.3.1 Traceability Mechanisms for Personal Information Protection

Traceability mechanisms refer to establishing product and data traceability systems during data mining, analysis, and processing due to personal data's dual product and data attributes. Huang Guobin et al. [37] identify personal data characteristics as: identifiability of data subjects, richness of content, contextual generation and application, and emphasis on data subjects' existence. In the big data context, personal data privacy associated with these characteristics has special connotations: data in text, images, or video that can identify an individual or group should be considered personal data privacy, all possessing confidential characteristics. Wang Zhong et al. [38] argue that personal data privacy traceability mechanisms stem from personal data's dual product and data attributes in the big data era. The product attribute means personal data used by enterprises exists in both raw and processed forms; the data attribute means personal data is the carrier of personal information, ultimately mined, analyzed, and processed in data form. Therefore, personal data possesses both product and data traceability. Building personal information traceability mechanisms can enhance personal information protection levels through: traceability mechanism standard systems → information registration systems → traceability supervision systems → traceability information reward and punishment systems. Yin Jianli et al. [39] propose establishing a comprehensive personal data traceability management system in the big data era, including: a technical support system (data security and confidentiality technology, data tracking and traceability technology), a policy and regulation guarantee system (sales licensing mechanisms, protection regulations, registration systems), and a tracking and traceability management platform. As big data technology advances, research on personal information protection mechanisms has deepened, with research directions becoming more refined. For instance, personal information protection traceability mechanisms can effectively prevent personal information leakage risks.

2.3.2 Governance Mechanisms for Personal Information Protection

Governance mechanisms refer to classifying stakeholders involved in big data

era personal information protection, defining each stakeholder's role, and ultimately forming a multi-stakeholder collaborative governance framework. Wang Zhong et al. [40] study personal data privacy governance mechanisms in big data environments from a stakeholder perspective, dividing stakeholders into five roles: individuals, personal data collectors, personal data processors, personal data users, and supervisors, involving six major interest groups: individuals, personal data enterprises, government, media, third-party privacy protection organizations, and NGOs. Through power-interest matrix analysis, these subjects are categorized as core stakeholders, important stakeholders, indirect stakeholders, and direct stakeholders. The study recommends building a multi-stakeholder collaborative governance mechanism: establishing competitive governance mechanisms for core stakeholders, incentive governance mechanisms for direct and indirect stakeholders, and leveraging the leading role of important stakeholders. In the big data era, personal information protection and utilization involve multiple stakeholders, and reasonably balancing their rights through targeted multi-stakeholder collaborative governance mechanisms is significant, though further refinement of specific governance mechanism design and practice is needed.

2.3.3 Transaction Licensing Mechanisms for Personal Information Protection Licensing mechanisms treat personal data as valuable assets, where relevant administrative agencies legally grant specific administrative counterparts the right to engage in personal data transactions through licenses or permits. Wang Zhong [41] analyzes the trend of opening personal data transactions and the value of personal data transaction licensing mechanisms in the big data context from the standpoint of promoting free personal data flow. The value includes separating personal data's dual privacy and asset attributes, constraining enterprise behavior when personal data control rights are separated from subjects, and forming competitive industry supervision. He proposes a personal data transaction licensing mechanism involving: issuing transferable licenses, adopting auction granting methods (determining total license numbers, industry allocation quotas, and auction revenue usage), clarifying licensed data types (by data source and content), establishing exit mechanisms, and building supporting mechanisms (personal data leakage reporting mechanisms, personal data leakage traceability mechanisms). This mechanism, as a refinement of personal information protection and utilization based on multi-stakeholder balance, increases transparency in personal information processing and promotes free personal information flow.

2.3.4 Leakage Reporting Mechanisms for Personal Information Protection Reporting mechanisms address the high risk of personal information leakage in big data environments by establishing personal information protection leakage reporting systems. Wang Zhong [42] constructs game models to compare the effectiveness of reporting behaviors by individuals and privacy protection associations under reward and non-reward scenarios, analyzing different

reporting mechanisms. The study concludes that without reporting rewards, association supervision is more effective than individual supervision; reporting rewards are effective incentives that improve supervision efficiency for both individuals and associations; and reasonable reward intensity can maximize the effectiveness of both. Given limited government and market capacities to address externalities in personal information protection in the big data era, establishing personal information protection leakage reporting mechanisms with positive incentives can promote multi-stakeholder participation and ensure free personal information circulation in the market.

2.4 Personal Information Security Risks and Countermeasures

This area covers big data investigative surveillance, smart libraries, social networks, and government management.

2.4.1 Risks and Countermeasures for Personal Information Security in Big Data Investigative Surveillance Big data investigation applies big data technology to investigative activities to ascertain facts and predict crimes. Yu Yang et al. [43] argue that big data investigative surveillance models present a triangular information monitoring structure of "investigation surveillance," "network surveillance," and "commercial surveillance." This model expands the scope of entities infringing upon citizens' personal information, increases personal information leakage risks, and correspondingly increases harmful consequences. They recommend strengthening internal controls and external supervision of investigative agencies to ensure effective protection of citizens' personal information during investigative activities.

2.4.2 Risks and Countermeasures for Personal Information Security in Smart Libraries Applying big data technology to libraries has transformed traditional libraries into smart libraries, but user personal information security risks exist during library information resource utilization. Li Aiguo et al. [44] use literature collection and in-depth interviews to analyze information security risks users face when utilizing library information resources in the big data era, recommending that libraries protect personal information (especially sensitive information or privacy) through legal collection, proper processing and storage, reasonable use, and controlled sharing to enhance mutual trust between libraries and users.

2.4.3 Risks and Countermeasures for Personal Information Security in Social Networks Big data commercialization in social networks is gradually becoming a trend. Meng Xiaoming et al. [45] identify five models for social network big data commercial development: operator model, third-party model, enterprise-owned model, government-guided model, and interest-driven model. Due to user personal reasons and information security technology factors leading to personal information leakage, they propose specific recommendations including improving relevant laws and regulations and adopting technical prevention

and control measures. Wang Xiwei et al. [46] construct an influencing factor model for social network personal information security behavior based on social cognition theory and protection motivation theory, testing the model through questionnaires and structural equation modeling. Results show that response efficacy is the most important factor affecting personal information security protection willingness in social networks, followed by perceived threat and self-efficacy having positive impacts, while avoidance behavior negatively impacts security protection willingness. User personal information security protection willingness positively influences security protection behavior. Therefore, social network personal information security risks can be mitigated through specific measures such as personal information security protection tools and threat assessment. For example, Tan Chunhui et al. [47] propose optimizing personal information protection policy tools by enhancing substantive effectiveness, increasing operability, optimizing structural proportions, and building policy tool networks.

2.4.4 Risks and Countermeasures for Personal Information Security from Government Management Perspective Wang Shaohui et al. [48] analyze personal information protection issues from a government management perspective, proposing specific recommendations including improving supervision mechanisms, establishing cross-border collaboration mechanisms for big data personal information protection, and building multi-governance mechanisms for personal information protection in big data environments. Wang Xiuzhe [49] argues that in the big data era, governments should actively assume responsibility for both negative non-infringement and positive protection of personal privacy rights, thereby promoting the rule of law in personal information protection. As power holders for public security maintenance, governments should fully exercise specific functions including security review, supervision, and management to address the impacts and challenges of big data technology on personal information protection in the public domain.

2.5 Research on Relevant Laws and Regulations

Research on laws and regulations concerning personal information in the big data era primarily focuses on: (1) legislative models for personal information protection, and (2) legislative improvements for personal information protection.

Regarding legislative models, Sun Zhengwei [50] analyzes China's existing personal information protection legal system and the theories of "right to be forgotten" and "information self-determination," finding that China's legal system has not incorporated big data factors into model selection. Since big data technology can integrate low-value-density information into high-value-density information, he recommends constructing a tort law system with the ultimate information user as the responsible subject, centered on secondary information dissemination and utilization. Regarding public-private law protection models, Wu Weiguang [51] critiques private rights protection for personal data information under big

data technology, arguing that private law protection cannot meet personal information subjects' control needs and that personal data information should be treated as public goods, with governance based on public law rather than private law for public interest and security purposes to promote free sharing of personal data information.

Regarding legislative improvements, content covers civil law, criminal law, and other fields. In civil law, Chen Xing [52] proposes that personal information protection should be explicitly stipulated in the general provisions of the personality rights section of the Civil Code, with personal information rights established as an independent chapter in the personality rights section. In criminal law, improvements primarily focus on the crime of infringing upon citizens' personal information, discussing the legal interests of the crime, behavior types, determinations of "serious circumstances" and "information quantity," and interpretations of "violating state regulations."

3 Analysis and Outlook

Research on personal information protection in the big data era spans multiple disciplines, primarily involving law, economics, library and information science, and journalism and communication. Research directions have gradually refined, depth has increased, and methodologies have diversified, injecting new vitality into future research. However, three prominent issues remain: (1) weak basic theoretical research with insufficient legal theoretical justification, leading to conceptual confusion among terms like personal information, personal data, personal privacy, personal information privacy, and personal data privacy; (2) lack of systematic research, showing "fragmented" and "scattered" characteristics, mainly reflected in unreasonable disciplinary distribution and suboptimal research content structure—for instance, law-dominated research exhibits insufficient interdisciplinary integration, requiring future emphasis on cross-disciplinary perspectives; (3) from a content structure perspective, research concentrates on basic theoretical issues like concepts and legal attributes, while lacking empirical research, first-hand survey data, and effective empirical testing, resulting in inadequate practical investigation and statistical support and ineffective standards, which hinders advancement of basic theoretical research. Overall, research content is not closely integrated with big data, failing to fully reflect research characteristics of the big data era.

Future research on personal information protection in the big data era should combine theoretical and practical research with innovative big data technology elements, excavating and analyzing the deep-level impacts of big data technology on personal information protection, identifying effective protection approaches from impact mechanisms. In basic theoretical research, emphasis should be placed on clarifying the relationships and differences among personal information, personal data, and personal privacy to avoid conceptual confusion in future research. Theoretical research should be localized, drawing on advanced foreign basic theories while extracting nutrients suitable for China's national conditions.

In practical research, theory and practice should be closely integrated. Besides continuing current interdisciplinary research trends, integration with big data technology is essential, ensuring that theoretically proposed solutions consider practical feasibility and implementability to avoid overly idealistic tendencies. Therefore, future research should emphasize empirical analysis and effectively validate research findings in practice.

References

[1] Chen Xuan, Li Yan. The Definition of "Personal Electronic Information" in the Big Data Era: A Comparative Study of Rights Derivation[J]. International Press, 2013, 35(12): 20-31. [2] Yang Yuan. Research on Personal Privacy Protection and Information Application in the Big Data Context[J]. Credit Reporting, 2014, 32(8): 24-26. [3] Ji Leilei. Comparative Study on Legislative Paths for Personal Information Protection[J]. Library Construction, 2017(9): 19-25. [4] Wang Shaohui, Du Wen. Progress in New Zealand's Personal Privacy Protection in the Big Data Era and Its Implications for China[J]. E-Government, 2017(11): 65-71. [5] Wang Min, Jiang Zuosu. Comparative Study on Personal Privacy Protection Between China and the US in the Big Data Era: Based on Comparative Analysis of Latest Privacy Protection Regulations[J]. Press Circles, 2016(15): 55-61. [6] Chen Xing. Construction of Personal Information Security Certification Mechanisms for Software Products in the Big Data Era[J]. Journal of Chongqing University of Posts and Telecommunications (Social Science Edition), 2016, 28(2): 39-45. [7] Luo Jiao. Research on Legal Issues of Personal Information Protection in Big Data Environments[J]. Library, 2018(5): 31-36. [8] Liu Yaqi. Analysis and Countermeasures for the Current Status of Protective Development and Utilization of Personal Information in Big Data Environments[J]. Library Science Research, 2015(15): 67-76. [9] Wang Zhong, Zhao Hui. Research on Personal Data Privacy Concerns in the Big Data Era: Based on Survey Data Analysis[J]. Information Studies: Theory & Application, 2014, 37(11): 26-29. [10] Kuang Wenbo, Tong Wenjie. Empirical Research on Personal Information Security and Privacy Protection: From the Perspective of Big Data Application Based on Innovation Diffusion Theory[J]. Wuhan University Journal (Humanity Sciences), 2016, 69(6): 104-114. [11] Mei Xiaying. Legal Attributes of Data and Its Positioning in Civil Law[J]. Social Sciences in China, 2016(9): 164-183, 209. [12] Yu Chong. Legal Interest Attributes and Criminalization Boundaries of "Citizens' Personal Information" in the Crime of Infringing Upon Citizens' Personal Information[J]. Political Science and Law, 2018(4): 15-25. [13] Yang Weiqin. Value Dimension Examination of Personal Information Ownership Models: From the Perspective of Interest Attribute Analysis[J]. Law Review, 2016, 34(4): 66-75. [14] Chu Jiewang, Li An. Research on Personal Information Privacy Protection Under New Situations[J]. Modern Information, 2016, 36(11): 21-26. [15] Shi Weimin. Realistic Dilemmas and Path Selections for Personal Information Protection in the Big Data Era[J]. Journal of Intelligence, 2013, 32(12): 154-159. [16] Wang Liming. On Legal Protection of Personal Information Rights: Centered on the Distinction Between Personal

Information Rights and Privacy Rights[J]. Modern Law Science, 2013(4): 66-68. [17] Feng Yuan. Distinction Between "Personal Information" and "Data" as Emerging Right Objects in the General Principles of Civil Law[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2018, 32(3): 81-88. [18] Zhou Dong. Personal Information and Privacy in the Big Data Era: A Comparative Study Based on Extraterritorial Law[J]. Graduate Law Review, 2015, 30(4): 135-143. [19] Wen Yanyan, Peng Yan. Research on Personal Information Protection Mechanisms[J]. Journal of Intelligence, 2018, 37(7): 127-131. [20] Ling Pingping, Jiao Ye. Re-analysis of Criminal Law Interests in the Crime of Infringing Upon Citizens' Personal Information[J]. Journal of Soochow University (Philosophy & Social Science Edition), 2017, 38(6): 66-71. [21] Han Xuzhi. Legal Doctrinal Analysis of Personal Information Concepts: Centered on Article 76(5) of the Cybersecurity Law[J]. Journal of Chongqing University (Social Science Edition), 2018, 24(2): 154-165. [22] Zhang Li'an, Han Xuzhi. Private Law Attributes of Personal Information Rights in the Big Data Era[J]. Legal Forum, 2016, 31(3): 119-129. [23] Wang Xuehui, Zhao Xin. Exploration of Integrated Public-Private Law Protection of Privacy Rights: From the Perspective of Big Data Era Personal Information Privacy[J]. Hebei Law Science, 2015, 33(5): 63-71. [24] Zhang Ping. Legislative Choices for Personal Information Protection in the Big Data Era[J]. Acta Scientiarum Naturalium Universitatis Pekinensis (Philosophy and Social Sciences), 2017, 54(3): 143-151. [25] Cheng Cheng. How to Cut Off Personal Information Crime Chains[J]. People's Tribune, 2017(17): 118-119. [26] Li Weimin. Research on the Nature and Legislative Model of "Personal Information Rights": From the Perspective of Internet New Rights[J]. Journal of Shanghai Normal University (Philosophy & Social Sciences), 2018, 47(3): 66-74. [27] Ren Longlong. Theoretical Foundation for Civil Law Protection of Personal Information[J]. Hebei Law Science, 2017, 35(4): 181-192. [28] Tian Ye. Dilemmas and Solutions of Informed Consent Principle in the Big Data Era: Using Biobanks as an Example[J]. Law and Social Development, 2018, 24(6): 111-136. [29] Lin Huanmin. Dilemmas and Solutions of Informed Consent Principle in Personal Information Protection[J]. Journal of Beijing Administrative College, 2018(6): 1-8. [30] Jiang Panpan. Legislative Experience and Implications of Consent in EU Personal Information Protection Law[J]. Library Construction, 2018(11): 11-16, 22. [31] Yu Xiaoyao. Leakage and Balance of Author Personal Information from a Big Data Perspective[J]. China Editor, 2017(9): 73-77. [32] Zhu Xinli, Zhou Xuyang. Balance Between Personal Data Utilization and Protection in the Big Data Era: Proposing the "Resource Access Model"[J]. Journal of Zhejiang University (Humanities and Social Sciences), 2018, 48(1): 18-34. [33] Zhang Shuqing. Footprints and Paths: Personal Information Protection and Big Data Rights Attribution[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2018, 31(3): 13-21. [34] Zhao Lili, Jin Xu. On the Contradiction and Coordination of Exchanging Personal Data for Services in the Big Data Era[J]. Journal of Intelligence, 2018, 37(12): 156-161. [35] Jiang Bo, Zhang Yanan. Principle of Reasonable Use of Personal Information in the Big Data Context[J]. Journal of Shanghai Jiaotong University (Philosophy and Social

Sciences), 2018(3): 108-121. [36] Wang Yulin. Research on Legal Issues of Personal Information Development and Utilization in Big Data[J]. Information Studies: Theory & Application, 2016, 39(9): 19-24. [37] Huang Guobin, Zhang Shasha, Yan Xin. Research on Conceptual Categories and Basic Types of Personal Data[J]. Library and Information Service, 2017, 61(5): 41-49. [38] Wang Zhong, Yin Jianli. Design of Personal Data Privacy Leakage Traceability Mechanism in Big Data Environments[J]. China Circulation Economy, 2014, 28(8): 117-121. [39] Yin Jianli, Wang Zhong. Research on Personal Data Traceability Management System in Big Data Environments[J]. Information Science, 2016, 34(2): 139-143. [40] Wang Zhong, Yin Jianli. Research on Personal Data Privacy Governance Mechanism in Big Data Environments: Based on Stakeholder Perspective[J]. Journal of Technical Economics & Management, 2014(8): 71-74. [41] Wang Zhong. Research on Personal Data Transaction Licensing Mechanism in the Big Data Era[J]. Theory Monthly, 2015(6): 131-135. [42] Wang Zhong. Research on Personal Data Privacy Leakage Reporting Mechanism in the Big Data Era[J]. Journal of Intelligence, 2016, 35(3): 165-168, 79. [43] Yu Yang, Wei Junbin. Conflict and Reconciliation: Personal Information Protection Under Big Data Investigative Surveillance Models[J]. Journal of Intelligence, 2018, 37(12): 147-155. [44] Li Aiguo, Cao Xiang, Wang Shejiao. Issues and Countermeasures for User Privacy Protection During User Information Resourceization in Libraries[J]. Library and Information Service, 2015, 59(13): 26-30. [45] Meng Xiaoming, He Minwei. Personal Privacy Protection in Social Network Big Data Commercial Development and Utilization[J]. Library Tribune, 2015, 35(6): 67-75. [46] Wang Xiwei, Wang Lei, Jia Ruonan, et al. Empirical Research on Influencing Factors of Personal Information Security Behavior in Social Networks[J]. Library and Information Service, 2018, 62(18): 24-33. [47] Tan Chunhui, Tong Lin. Analysis and Optimization Recommendations for China's Personal Information Protection Policy Tools[J]. Library and Information Service, 2017, 61(23): 67-75. [48] Wang Shaohui, Yin Houjie. Research on Personal Information Protection Issues in Big Data Environments from a Government Management Perspective[J]. Chinese Public Administration, 2015(11): 19-24. [49] Wang Xiuzhe. Government Responsibility for Personal Information Protection in Public Security Fields in the Big Data Era[J]. Theoretical Investigation, 2017(4): 52-56. [50] Sun Zhengwei. Legal Protection Model Selection for Personal Information in the Big Data Era[J]. Research on Library Science, 2016(9): 72-76, 65. [51] Wu Weiguang. Critique of Private Rights Protection Theory for Personal Data Information Under Big Data Technology[J]. Political Science and Law, 2016(7): 116-132. [52] Chen Xing. Establishment and Position of Personal Information Rights in China's Civil Code in the Big Data Era[J]. Journal of Beijing Administrative College, 2016(6): 1-8.

A Review of Personal Information Protection Research in the Big Data Era

Jiang Panpan
School of Law, Jilin University, Changchun 130012

Abstract: [Purpose/significance] By sorting out research literature on personal information protection in the big data era, this study provides reference for fu-

ture research on personal information protection. [Method/process] Using the CNKI database, we systematically collected domestic core journal papers on personal information protection in the big data era, analyzed their content, summarized research topics, and discussed them. [Result/conclusion] Existing research findings were classified into five thematic categories according to topic: domestic and international practical experience research, basic theoretical issues research, exploration of personal information protection mechanisms, personal information security risks and countermeasures, and relevant laws and regulations research. The paper reviews shortcomings of existing research and proposes future research directions.

Keywords: personal information; big data; privacy; legal attributes

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