

The Debate Between Personal Information and Data in Chinese Law: Definition and Distinction of Information, Data, and Their Related Concepts

Authors: Jin Yulu, Jin Yulu

Date: 2024-04-08T00:00:00+00:00

Abstract

The Civil Code of the People's Republic of China stipulates personal information and data as two independent legal objects, and legislation such as the Personal Information Protection Law and the Data Security Law distinguishes between personal information and data. However, due to the lack of clear differentiation between the concepts of information and data in both legislation and theoretical research, personal information and data, as well as personal information and personal data, are used interchangeably, leading to continuous disputes over the intension and extension of data property rights. Through multidisciplinary, multi-contextual, and multi-level semantic and normative analyses of the concepts of information and data from factual, normative, and theoretical research perspectives, it is found that the view of information and data as two sides of the same coin and substantially identical cannot be established; information and data are not essentially equivalent and possess distinct normative significance. Information refers to all intelligence, messages, and content in human society, while data constitutes an electronic or other form of recording of information; data represents a novel form of expression of information under artificial intelligence and big data technologies, and data belongs to the category of information. Concepts such as personal information, personal data, personal information property rights, and data property rights in Chinese legal texts and theoretical research also carry differentiated legal significance.

Full Text

The Dispute Between Personal Information and Data in Chinese Law: Definition and Distinction of Information, Data, and Related Concepts

(Tsinghua University School of Law, Beijing 100084)

Abstract

The Civil Code of the People's Republic of China stipulates personal information and data as two independent objects, a legislative approach that also appears in laws such as the Personal Information Protection Law and the Data Security Law, which distinguish personal information from data. However, because the concepts of information and data have not been clearly differentiated in legislation and theoretical research, personal information and data, as well as personal information and personal data, are used interchangeably and conflated, leading to continuous disputes over the connotation and denotation of data property rights.

Through multidisciplinary, multi-contextual, and multi-level semantic and normative analysis of the concepts of information and data from factual, normative, and theoretical research perspectives, this paper finds that the view that information and data are two sides of the same coin and substantially identical cannot hold. Information and data cannot be equated in essence and have distinct normative significance. Information refers to all information, messages, and content in human society, while data is a form of electronic or other recording of information. Data represents a new form of expression of information under artificial intelligence and big data technology, and data belongs to the category of information.

In Chinese legal texts and theoretical research, concepts such as personal information, personal data, personal information property rights, and data property rights also have distinct legal meanings.

Keywords: information, data, personal information, personal data, personal information property rights, data property rights

I. The Problem

In the mid-20th century, with the deployment of the first electronic computers and the development of transistors and large-scale integrated circuits, humanity began to transition from a purely material world into an information world, where information became an important type of resource and wealth. Just as information resources are accessible and utilizable information that constitute part of information, information property is that portion of information that has economic value and can be accessed and utilized. The essence of information

property is information, which differs from traditional “objects” in that its existence and utilization depend on certain carriers. With continuous advances in information technology and deepening research on intellectual property, people have recognized that information and information carriers are distinct objects of transaction. For instance, a work as intellectual information is independent of its carrier, paper. The concept of information and the relationship between knowledge and information have been extensively discussed elsewhere, and this paper will not repeat those elaborations.

With the prevalence of the internet and the exponential development of electronic technology following Moore’ s Law, particularly the emergence of big data technology, cloud computing, and generative AI large models, the concept of data has increasingly appeared in domestic legislation and academic research. However, scholars have expressed diverse and inconsistent opinions regarding the concepts of information and data and their relationship.

At the legal level, Article 111 of the Civil Code of the People’s Republic of China (hereinafter referred to as the “Civil Code”) provides that personal information is protected as a personality right, while Article 127 stipulates that “where the law provides for the protection of data and virtual property, such provisions shall apply.” The Civil Code’ s different legal provisions for personal information and data render them two independent objects of legal protection and provide a Chinese legal basis for distinguishing information from data. In domestic legislation, laws such as the Personal Information Protection Law and the Data Security Law have emerged to differentiate personal information from data. It appears that China has already distinguished personal information from data at the legislative level. However, because current domestic legislation and theoretical research have not clearly distinguished between information and data, the scope of norms between domestic legislation on personal information and data is mixed and unclear, deviating significantly from the rigorous and comprehensive blueprint for personal information and data protection envisioned by law.

In theoretical research on personal information (personal data) protection, sometimes personal information and personal data are used interchangeably to refer to the same thing, while at other times they refer to different content. In foreign legislation, information and data, as well as personal information and personal data, are not distinguished in legislation and research and often refer to the same thing. In domestic theoretical research, from information and data to personal information and personal data, and further to personal information property rights and data property rights, unified conceptual cognition of connotation and denotation has not been formed, with different definitions appearing in different academic papers. This has created a phenomenon where “there are a thousand Hamlets in a thousand people’ s eyes” regarding the legal concepts involving information and data.

Whether to distinguish between the concepts of information and data, and the understanding of whether they have a content-carrier relationship, concerns the characterization of personal information and data in the Civil Code and affects

the systematic construction and application of domestic personal information and data norms. It also influences whether the binary distinction viewpoint in academic research—protecting human rights interests in personal information and protecting property interests in personal data—can hold, as well as the conceptual distinction and understanding of different data property rights appellations such as personal information and enterprise data, enterprise data property rights, network data, and commercial data property rights. The key and source to solving these problems lies in clarifying and distinguishing the concepts of information and data. Unlike existing research that views information and data as two sides of the same coin or substantially identical, this paper argues that information and data cannot be equated in essence and have distinct normative significance. The following sections will respectively conduct semantic empirical analysis of the concepts of information and data from different disciplinary perspectives, different countries, and different contexts at the factual level, systematic conceptual 梳理 at the normative level, and content understanding and normative distinction at the theoretical research level, in order to provide a foundation for clarifying and constructing China’s personal information and data legislative system in the future and refining the connotation and denotation of data property rights or data property rights concepts in the “Data Twenty Articles,” thereby breaking through conceptual barriers.

II. Definition and Distinction of Information and Data at the Factual Level

(1) Definitions of Information and Data in National Dictionaries

In the Oxford Dictionary’s definition of information, information is facts or details about someone or something; information is content represented by a particular arrangement or order of things; information is data processed, stored, or sent by computers. Data is facts or information, especially when examined and used to find things or make decisions; data is information stored in computers. Thus, in the Oxford Dictionary, when information is stored or processed by computers, it becomes data, and data is electronically stored information.

The German Duden Dictionary defines information as: everything worth knowing; details about individuals, or statements or implications; message content composed of code characters. It defines data as: characters, instructions, and information stored electronically. In the Duden Dictionary, data is information stored electronically, the scope of information includes data, information can be composed of code characters, but also includes more content-level things.

In Chinese dictionaries, information is news, messages, and “the object of transmission and processing in communication systems, generally referring to events or material data.” The Modern Chinese Dictionary also defines information as: in information theory—information refers to reports transmitted by symbols. Data is defined as values used in various statistics, calculations, and scientific research, or as the object of processing by electronic computers. “Early com-

puters were mainly used for scientific calculations, so the objects processed were mainly numbers representing numerical values. Modern computers have increasingly broad applications, and the objects they can process include numbers, text, letters, symbols, documents, images, etc.” Thus, both information and data definitions include being objects of computer processing. Beyond this, the scope of information also includes the meaning of news or messages, with a broader scope of application, closer to the life level, while data is more used in computer technology and scientific fields.

In essence, current dictionaries of various countries have different definitions for the two terms of information and data. Definitions of information focus more on the message and content level, with diverse forms of expression. Data is coded instructions or electronically stored information. Information includes data, with a scope larger than data, and data is one form of expression of information.

**** Definitions of Information and Data in National Dictionaries

(2) Definitions of Information and Data in Different Disciplines

Different disciplines have different understandings and interpretations of the concepts and content of information and data from their respective perspectives.

Information as a scientific term was first used by Hartley in 1928. In information management, information generally refers to objects that can be obtained, transmitted, stored, and utilized, which can eliminate uncertainty and increase effectiveness. For example, in the 1950s, Shannon, the founder of information theory, defined information in “A Mathematical Theory of Communication” as: information is the uncertainty of the information source, something used to eliminate uncertainty. Information is also called information entropy. The founder of information cybernetics defined information as: information is the content that people perceive about the objective world and exchange and communicate. In information management, information is broadly defined as signals, symbols, materials, knowledge, intelligence, or message content. Information is not matter, nor energy. Information is information itself, is the representation symbol of things, is the signal sent by the information source, or is the content contained in information. Some scholars believe that information is stock, is a reflection of the objective world, and information has value because of exchange and transmission. This view emphasizes the objective existence of information, believing that information is the sum of content and its material shell.

In information economics, Kenneth Arrow defined information in 1977 as the degree of knowledge in transmission, and believed that information is an economic good, with indivisibility and non-competitiveness, meaning information is difficult to monopolize or occupy exclusively. Information has characteristics of hierarchy, indivisibility, and shareability. Some information economists believe that information has universality, objectivity, derivability, and transmissibility. Derivability also means infinity, referring to the fact that other information gen-

erated from information can form an endless derivative chain, and the result of information derivation can even produce an information explosion effect. This view emphasizes more the content, meaning, and subjective value of information.

In information ethics research, some scholars point out that the ontological existence of information has an evolutionary process, beginning to appear in a “cyborged” form with some human agents (wireless networks, smartphones, tablets, etc.). Information technology has caused a digital transformation of the information ontology. Floridi distinguishes information from data and proposes the concept of the infosphere, an organic body of human social informatization composed of information, data, knowledge, etc. This view both recognizes the objective existence of information and understands the significance of information’s dependence on carriers.

Data began to be widely used in the field of computer science in the 20th century. In computer science, data is usually defined as: the general term or collection of symbols that can be recognized and processed by computers. Data is the raw material for computer processing and is strings in computer systems. Data is also considered “symbolic representation of objective things, and physical symbols that carry information.” In computer science, data has broad meaning, and text, images, sounds, etc., that can be input into computer systems for encoding can all be classified as data. Early on, some scholars believed that information is data, and the two were conflated. Later, with the application of computers, some scholars believed that data is a distinction of physical state, and different combinations of data may convey some information or may not convey information. Some scholars believe that information is data with specific meaning, data is unprocessed factual description, and describe the relationship between data and information as that between raw material and product. Other scholars believe that data is the result of computer recording or measurement of things in the real world through code.

Overall, in the current field of information science, information and data are distinguished from each other but can be converted into each other. Data is the physical symbol or computer code that carries information and is arranged according to certain rules. “Information is the content carried by data.” For example, the father of modern management, Drucker, defined information as: “data endowed with relevance and purpose.” He believed that raw data itself has no practical meaning; only when interpreted and affecting entity behavior does it have meaning, and at this point, data becomes information. Therefore, data can serve as the basic material for information; through processing and refining, data can evolve into meaningful and valuable specific sequences of data combinations—information. He believed that continuing to extract regular useful content from information is knowledge. He proposed the progressive DIKW system (Data-Information-Knowledge-Wisdom), where the relationship between data, information, knowledge, and wisdom is that information is processed and refined from data, knowledge is regularized and abstracted from information,

and wisdom is repeatedly summarized from knowledge.

Thus, in the field of information science, the concepts of data and information are distinguished from each other. Data is considered the electronic form of information and the basic element or symbol that constitutes information.

**** Definitions of Information and Data in Various Disciplines

III. Definition and Distinction of Information and Data at the Legal Normative Level

(1) Identification and Distinction of Information and Data by International Organizations

The International Organization for Standardization (hereinafter referred to as ISO) defines data in the standard “Information Technology Vocabulary” as: data is a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Data can be processed manually or automatically. From the repeated mention of “access information or data,” and the definition of information processing as “systematically performing information operations, including data processing, which may include data communication and office automation operations,” and “information processing shall not be used as a synonym for data processing,” it can be seen that in ISO’s definition, information is distinguished from data, and data is an expression of information in a specific era and specific technical context. The scope of information is obviously larger than data.

The Organisation for Economic Co-operation and Development (OECD) defines personal data in the “Guidelines on the Protection of Privacy and Transborder Flows of Personal Data” as: “any information relating to an identified or identifiable individual (data subject).” Personal data is information directly or indirectly related to natural persons. Both personal information and personal data concepts appear in the Guidelines, but no detailed distinction is made. According to its definition of personal data, the essence of personal data is information. In the Guidelines, information and data are not distinguished and are used interchangeably. In November 2020, ten ASEAN countries and China, Japan, South Korea, Australia, and New Zealand, totaling 15 countries, signed the Regional Comprehensive Economic Partnership (RCEP). RCEP defines personal information as: “personal information means any information, including data, relating to an identified or identifiable natural person.” Thus, in RCEP, information and data are two different concepts, with information including data.

Thus, in standards, agreements, and guidelines issued by international organizations, there are cases where information and data are used without distinction, as well as provisions that distinguish information and data, holding that data belongs to information and is a re-expression of information.

**** Definitions of Information and Data by International Organizations

(2) Identification and Distinction of Information and Data in National Legal Norms

At the legal normative level, different countries do not uniformly use concepts such as data, information, information privacy, and materials in the field of personal information protection, based on their national practices and judicial characteristics. The EU is accustomed to using the concept of “personal data,” the United States uses “personal information,” China’s Taiwan, Hong Kong, and Macao regions mostly use “personal data” (the term “data” is a translation of “Data”), while mainland China uses the concept of personal information. In legislation related to information and data, the EU mostly uses the data concept for legislation, while mainland China and the United States use both information and data simultaneously. For example, Article 2 of the EU Digital Markets Act (DMA) stipulates that data refers to digital behaviors, facts, or information, and compilations of behaviors, facts, or information, including in the form of sound, video, or audiovisual recordings. From the EU DMA’s provisions, it can be seen that data is digital information and compilations of information. Data can be regarded as a form of expression of information. The GDPR defines personal data as: “personal data means any information relating to an identified or identifiable natural person (data subject).” The EU GDPR defines the essence of personal data as information. Thus, EU legal norms are accustomed to using the data concept, with data serving as an expression of information in the new era.

In the U.S. Open Government Data Act of 2017, it is stipulated that data refers to recorded information, regardless of the manner and medium in which the data is recorded. The U.S. California Consumer Privacy Act (CCPA) of 2018 uses the concept of personal information and stipulates: “personal information means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.” From the definitions of personal data in the EU GDPR and personal information in the U.S., both refer to the same thing, especially since the EU GDPR defines the attribute of personal data as information, showing that data belongs to information. From the definitions of data in the EU DMA and the U.S. Open Government Data Act, data can be seen as electronic or other forms of recording or compilation of information.

(3) Identification and Distinction of Information and Data in Mainland China’s Legal Norms

The first distinction between information and data at the normative level in China’s laws appeared in the Cybersecurity Law. Article 76 of the Cybersecurity Law, implemented in 2017, defines “personal information” as: “various kinds of information recorded in electronic or other forms that can identify a natural person’s personal identity, either alone or in combination with other

information.” It defines “network data” as: “various electronic data collected, stored, transmitted, processed, and generated through networks.” From these provisions, it can be seen that at the normative level, Chinese legislators’ definition of personal information tends to be content-based, with forms including electronic or other recording methods; while the definition of data emphasizes “electronic data,” highlighting the electronic characteristics of data, with data being the digital expression form of information.

From the beginning, the distinction between information and data at the normative level in China has been relatively clear. Subsequently, during the compilation of the Civil Code, the “General Principles of Civil Law (First Draft)” Article 108 once listed “data information” as an object of rights. The appearance of the term “data information” shows that Chinese legislators have been contemplating the distinction between information and data and have been hesitant on this issue. Ultimately, Article 127 of the Civil Code, implemented in 2021, lists data as an object of civil rights, while Article 1034 defines the concept of personal information and the specific content of legal protection of personal information. From the Civil Code’s distinguishing provisions on personal information and data, Chinese legislation recognizes the significance of distinguishing information and data at the normative level.

In particular, China has successively issued the Data Security Law and the Personal Information Protection Law, which still distinguish between data and information. Article 3 of the Data Security Law defines data as: “data refers to any record of information in electronic or other forms.” Article 38 stipulates: “data such as personal privacy, personal information, trade secrets, and confidential business information learned in the performance of duties shall be kept confidential according to law.” Article 53 mentions “data processing activities involving personal information.” These provisions both clarify that data is the electronic or other forms of recording of information and stipulate that personal information belongs to data.

Thus, after 梳理 ing international legal norms that stipulate information and data, it can be seen that the international community has distinguished between information and data at the normative level. Although there is conflation in the definitions of personal information and personal data in the EU and the United States, overall, data belongs to information, data is the electronic or other forms of recording of information, and data is an expression of information in the era of artificial intelligence and big data. In China’s legislation related to information and data, there are separate Personal Information Protection Law and Cybersecurity Law, and the Civil Code recognizes the distinction between information and data, equating the essence of data and information, and holding that data is an electronic or other form of recording of information.

**** Definitions of Information and Data in National Legal Norms

IV. Definition and Distinction of Information and Data in Theoretical Research

Current research holds three main views on the identification and distinction of information and data: first, that in the current era, there is little significance in distinguishing information and data, either factually or at the legal normative level, and no distinction is necessary; second, that there are factual differences between information and data, but little significance in distinguishing them at the normative level; and third, that information and data should be distinguished at both the factual and normative levels.

(1) The No-Distinction Theory of Information and Data

Scholars holding the no-distinction theory believe that information and data have homogeneity and symbiosis, with no essential difference. This view is supported by the current use of different terms such as personal information, personal data, personal materials, and information privacy in various countries. Jeffrey Ritter and Anna Mayer argue that there is no need to distinguish between information and data, holding that data refers to any information recorded electronically or digitally, is retrievable, and is perceivable by humans and machines. From this perspective, data that should be legally protected has become conflated with information, and there is no normative significance in distinguishing them.

(2) The Normative-Level Distinction Theory of Information and Data

Some research holds that information and data have semantic and connotative differences in reality. However, under current digital technology, information and data have high interoperability and symbiosis. Therefore, distinguishing information and data conceptually is not very significant. Although there are many discussions on distinguishing data and information in fact or usage, because data has become the dominant form of information in the digital age and data and information can directly correspond, in practice, the two refer to each other and do not cause confusion. Therefore, there is no need to distinguish between data and information in usage. This view is confirmed by the use of different terms in current worldwide legal norms. For example, the “data” concept used in Europe and America, the “information” concept used in the United States, and the “materials” concept used in some regions, although different in 称谓, essentially refer to the same content.

Yale Law School Dean Robert C. Post also believes that when discussing information and data issues, especially when different countries currently use different terms such as personal information, personal data, and information privacy, there is no need to dwell too much on 称谓 issues, but rather to focus on the specific things behind or pointed to by the different 称谓. Currently, when countries talk about protecting personal information and personal data, they refer to the same thing and do not cause confusion.

Therefore, scholars supporting this view believe that distinguishing information and data only at the 称谓 or usage level is not very significant, and that the distinction between data and information issues should be made at the legal normative level. For example, Mei Xiaying believes that the legal significance of distinguishing data and information lies in distinguishing different legal issues. Legal issues concerning data media and carriers without involving content, such as enterprise data protection, should be classified as pure data problem types, while issues concerning information content protection, such as personal information protection, belong to the information problem type. The distinction between information problems and data problems can better apply different legal norms.

(3) The Factual and Normative-Level Distinction Theory of Information and Data

There is also a view in academic research that information and data should be distinguished at both the factual and normative levels. Lawrence Lessig divided network regulation into three levels: content layer, code layer, and physical layer. Information corresponds to the content layer, where the information adjusted by law is content with certain meaning; while data corresponds to the code layer, where the data adjusted by law is code existing in cyberspace in the form of binary bits of 0 and 1, and is the symbol and carrier of information. Some scholars believe that when information and data are mentioned in the current world, they are basically understood and defined from three aspects: information content (semantic) level, symbol (syntactic) level, and entity level. Some scholars believe that the distinction between information and data is clear both in reality and in legal provisions. Currently, data as a carrier is already protected at the entity level; while collections of data such as databases, as symbols, are already protected by copyright law; personal information, trade secrets, etc., at the information (semantic) level have not yet obtained property rights protection but are protected through defense mechanisms.

Other scholars believe that the distinction between data and information is significant, otherwise unnecessary misunderstandings will arise when discussing concepts such as personal data property rights. The essence of data is the carrier of information, is the digital description of information, is the object of technical processing, and is the electronic recording of information. Especially at the normative level, distinguishing information and data at the object level is the key to data ownership issues. Data is a symbol, a carrier, and something stored by computers; while information emphasizes the content reflected or carried by symbols or carriers. Accordingly, data can become the object of property rights protection, while the information it carries is protected by personality rights. It is pointed out that at the normative level, copyright law has already achieved property rights protection for both information and data, such as the protection of rights in works and databases.

In current research, the mixed use of “information” and “data” makes it difficult

for different studies to conduct research under a unified discourse system. Research named with the same “personal data property rights” or “data property rights” discusses rights subjects, objects, and content that are far apart. Some scholars believe that at the level of unity of legal discourse system, conceptual distinction between data and information is significant. Information and data have distinctions in concept, essence, and normative significance. Information is the objective attribute and law of things, and data is the symbol and carrier. Intelligent subjects identify information and record it to form data. The relationship between information and data is: information (objective) –intelligent subject identification (subjective) –data (objective) carrier. Therefore, at the legal normative level, information right holders have rights over information, while data right holders have rights over data. Information cannot be traded, but data has economic value, and data property rights should be established over data. Paul Solove, in his improvement of identifiability theory, points out that “personally identifiable information” includes two parts, “data” and “personally identifiable information,” which should be distinguished.

Overall, current academic views on whether to distinguish between the concepts of information and data cannot be unified. There are three views: the no-distinction theory, the factual distinction but not normative distinction theory, and the distinction at both factual and normative levels.

V. Definition and Distinction of Information, Data, and Related Concepts in Legislation and Legal Research

In the legal field, the distinction and definition of the concepts of information and data relate to the understanding and distinction of the concepts of personal information and personal data. This affects the selection of concepts such as personal information and data, personal information and data, personal information property rights and data property rights, information property and data property, personal information property rights, data property rights, and information property rights in future theoretical research and legal texts, as well as the distinction between these concepts. It will also affect the understanding of the connotation and denotation of personal information rights and data property rights and influence subsequent theoretical research. This paper argues that at the factual level, conceptual usage, and legal norms, “information” and “data” should be strictly distinguished, and based on this, the concepts of personal information and personal data, personal information property and data property, information property and data property, personal information property rights, data property rights, and information property rights should be distinguished, and subsequent Chinese legislation and research should be developed based on this conceptual distinction and definition to achieve uniformity in concept usage in legal texts and theoretical research and avoid confusion in understanding. The conceptual connotation and distinction between information and data are detailed below.

(1) Conceptual Definition of Information and Data

Both information and data have a long history and continuously acquire new meanings with the development of the times. Information in Latin is “Indicium,” referring to information intelligence, information, or evidence. Data in Latin is “Datum,” referring to facts, also meaning known and objective records of facts. Broadly speaking, both information and data in daily life can refer to all messages, information, and content. In the traditional “small data era,” daily references to scientific research data, statistical data, and financial data refer to scientific research materials or scientific research information. At that time, the concepts of data and information were not distinguished in terms of content and category. With the progress of information technology, the application of information increasingly accompanies the use of computers, the internet, algorithms, big data, and other technologies. When people talk about regulating the use of information in daily life or normative significance, data increasingly appears in people’ s vision. The term “data” has become a concept closely related to computer technology, as in various national dictionaries, where data is often explained as electronically stored information or information stored by computers. In information disciplines, data is explained as the electronic form evolved from the information ontology, and data has become a new form of information in the new era.

In the era of artificial intelligence and big data, the concepts and categories of information and data in legal norms worldwide have begun to diverge. For example, in the ISO standard “Information Technology Vocabulary,” information is defined as various symbols, messages, and content, while data is considered a re-expression of information, with information including data. In the EU DMA, the U.S. Open Government Data Act, and China’ s Data Security Law, data refers to information recorded electronically or in other forms, or the recording of information. Therefore, from the perspective of existing international normative documents, in the era of information and big data, information still refers to all information, messages, and content in human society, while data has become an electronic or other form of recording of information. Data is a new form of expression of information under artificial intelligence and big data technology, and data belongs to information. This definition also conforms to the definition of data in Article 3 of China’ s Data Security Law (data refers to any record of information in electronic or other forms).

(2) Similarities and Differences Between Information and Data as Legal Concepts

As legal concepts, information and data are two distinct concepts. As mentioned above, information refers to all messages, information, knowledge, and meaningful content in human society; data is the electronic or other form of recording of information, a form of expression of information. The nature of information and data has similarities and differences, and there is controversy over the scope of the two categories.

First, in comparing the similarities and differences in the legal nature of information and data, the similarities are: first, the essential attributes of information and data are the same, and both can become objects in law. As a form of expression of information in the new era, data belongs to information, and the essential attributes of the two are the same. The Civil Code of China simultaneously stipulates data and personal information, proving that information and data are two distinct concepts, and both data and personal information have objectivity. Second, both information and data are intangible, non-material, and non-consumable, distinguishing them from traditional tangible “objects.” Third, both information and data are reproducible and have commercial value. The differences in the nature of information and data are: first, different forms of expression. The form of information can be text, charts, can be recorded and inscribed on paper, rock walls, stones in the real world, can be recorded as data in cyberspace, and can even be carried by new carriers such as quantum that may appear in technological iterations. Data is one form of expression of information, mostly electronic binary symbols. Second, different identifiable subjects. Information emphasizes meaning to humans, with natural persons as the identifiable subjects, while data emphasizes more machine identifiability, with computers as the identifiable subjects. Third, the basis for commercial value is different. The commercial value of information is based on the value of information itself, while data often has commercial value because of its potential to extract valuable business information from disordered, massive, and fragmented data. Therefore, the greater the potential of data to extract valuable information, the greater its commercial value.

Second, in understanding the legal categories of information and data, currently, there are three views in academia: “information = data,” “information > data,” and “information < data.” Under the concepts of information and data defined in this paper, the category of information is larger than data. Information is a universal concept for human society from the primitive era, agricultural era, industrial era to the information era, artificial intelligence, and big data era. The category of information refers to all messages, information, and content in human society. Information can exist simultaneously in the current real and network worlds. Data is a form of expression of information, a product of the big data era, and a recording and expression of information in electronic or other forms in cyberspace or real space. Therefore, the category of information is larger than data, which is consistent with the definitions of information and data in current domestic and foreign legal normative documents.

(3) Selection and Distinction Between Personal Information and Personal Data Concepts

From the above arguments, the concepts of information and data should be distinguished at both the factual and legal normative levels. The concept of “data” in contemporary times has distinct technical characteristics and morphological features. Different from the technical specificity of “data,” the concept of

“information” is technologically neutral and can be applied under various technical backgrounds, emphasizing the content meaningful to humans expressed by various symbols.

Thus, in the use of personal information and personal data concepts, the concept of personal information has a broader scope. Current national norms and academic research mention concepts such as personal information, personal data, and consumer personal information, conflating personal information and personal data. For example, personal data in the EU GDPR, personal information in the CCPA, and personal data and personal information concepts mentioned by scholars worldwide in academic research have mixed phenomena, referring to the same thing. In China’s personal information protection legislation, among the concepts of personal information, personal data, personal data, and personal information privacy, the concept of “personal information” is chosen. This is because data is only one form of expression of information, and the scope of the personal data concept is smaller than that of the personal information concept. If legislation chooses to enact a “Personal Data Protection Law,” it would only protect part of personal information or one of the many forms of personal information, and the scope of protection would not be comprehensive. The concept of “personal data” is the Chinese translation of the English “personal data” and is essentially no different from “personal data.” The concept of “privacy” is used in American law and by Americans and has different connotation and denotation compared to the privacy concept in the civil law system, especially in China. If “privacy” is used to represent personal information in China, it is easy to confuse it with the traditional concept of privacy. Therefore, strictly speaking, at the normative level in China, China’s Personal Information Protection Law’s use of “personal information” is more rigorous and scientific than using “personal data” or “personal data,” and is consistent with China’s legal tradition and habits. In future theoretical research in China, personal information and personal data should also be distinguished to maintain consistency with Chinese legal provisions. However, special attention should be paid to the fact that when personal data is mentioned in legal texts or judicial practices to respect the original text or customary usage, it is still the concept of personal information, and special explanation should be made when used in this way.

(4) Coordination and Unification of Personal Information and Data Concepts in Legislation

In distinguishing personal information and data, according to incomplete statistics, currently at least 19 laws and 627 departmental regulations in China alternately use the concepts of “personal information” and “data.” As mentioned above, personal information is different from the personal data concept, and the category of personal information is larger than personal data. Therefore, personal information does not belong to data; both personal information and data belong to a category of information. Data is a re-expression of information in electronic or other specific forms. In this sense, the possible overlapping cat-

egory between data and personal information is personal data. Personal data belongs to both personal information and data. Since personal information includes personal data, when protecting and regulating personal information as a whole, the personal data expression form or other expression forms of personal information recorded electronically or in other forms are all within the scope of adjustment. With this understanding, the current academic view that personal information includes both personality rights and property rights can be self-consistent, because the concept of personal information includes personal data. There is no need to dichotomize personal information and personal data to protect personality rights in personal information and property rights in personal data; the two have an inclusive relationship. After clarifying the relationship between information and data, personal information and personal data, and personal information and data, it can be seen that the naming of China's Personal Information Protection Law and Data Security Law is scientific, and the two laws protect different categories. Future Chinese laws, departmental regulations, etc., should also distinguish the usage of personal information, personal data, and data according to this.

In summary, after conducting multidisciplinary, multi-contextual, and multi-level semantic and normative analysis of the concepts of information and data from factual, normative, and theoretical research perspectives, this paper finds that information and data are not two sides of the same coin, and mixed use will cause confusion in legislative systems, legislative normative language, and the scope of norms and research in theoretical research and judicial practice. Information refers to all information, messages, and content in human society, while data is an electronic or other form of recording or expression of information, and data belongs to information. Personal information and data are essentially both information; data cannot include personal information, and the two are independent concepts without overlapping scope. The Civil Code of China separately stipulates personal information and data as objects in civil law in Articles 111 and 127, also indicating that personal information and data are two different concepts. Compared to data, personal information refers to information content with specific subject personal identifiability, while data is the recording and expression form of massive information. Therefore, China's use of personal information rather than personal data or personal data in the Personal Information Protection Law is scientific. Legislation on personal information protection includes the personal data form of personal information. The definition of data in the Data Security Law is also scientific and reasonable. Future Chinese legislation should also define related concepts based on the distinction between personal information and data in these two laws and the Civil Code, strengthening the systematization and unification of personal information and data protection and provisions in legal norms. The terms such as data, enterprise data, government data, and public data mentioned in current Chinese legal norms, judicial practice, and academic research 实质上 refer to electronic or other forms of recording of information, and when mentioning the data concept, it should be noted that it is distinguished from personal information.

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

Source: ChinaXiv – Machine translation. Verify with original.