Postprint: An Analysis of GDPR' s Applicability and Role in Personal Data Protection for Open Scientific Data Sharing

Authors: Sheng Xiaoping, Yang Shaobin

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

Abstract

[Purpose/Significance] By analyzing the relevant provisions of the EU' s General Data Protection Regulation (GDPR), this study aims to provide reference for the protection of personal data in the open sharing of scientific data in China. [Method/Process] Through textual analysis and a review of research on GDPR and personal data protection, this paper analyzes the applicability and role of GDPR in personal data protection for scientific data open sharing and its implications for China. [Results/Conclusion] GDPR has many normative effects on personal data protection in scientific data open sharing, including clarifying the basic concepts and scope of protected objects, main principles, main rights of data subjects, and main responsibilities and obligations of data controllers and processors, and establishing a legal foundation for personal data processing. The implications for China are: we should establish and improve China' s personal data protection legal system, strengthen risk management of personal data in scientific data open sharing, build a dynamically linked and traceable scientific data open sharing system, thereby achieving personal data protection in scientific data open sharing in China.

Full Text

Analysis of the Applicability and Functions of GDPR in Protecting Personal Data in Open Sharing of Scientific Data

Sheng Xiaoping, Yang Shaobin
School of Library, Information and Archives, Shanghai University, Shanghai 200444

Abstract:
[Purpose/Significance] This paper provides reference for protecting personal

data in China' s scientific data open sharing by analyzing relevant provisions of the EU General Data Protection Regulation (GDPR). [Method/Process] Using textual analysis, this paper reviews research on GDPR and personal data protection, then analyzes GDPR' s applicability and functions in protecting personal data in scientific data open sharing, and its implications for China. [Result/Conclusion] GDPR has many normative functions for personal data protection in scientific data open sharing, including clarifying basic concepts and scope of protection objects, main principles, primary rights of data subjects, and main responsibilities and obligations of data controllers and processors, as well as establishing the legal basis for personal data processing. GDPR' s enlightenment for China is that we should establish and improve China' s personal data protection legal system, strengthen risk management of personal data in scientific data open sharing, and build a dynamically linked and traceable scientific data open sharing system, thereby achieving personal data protection in China' s scientific data open sharing.

Keywords: scientific data; open sharing; data protection; personal data; GDPR

With the development of big data, the Internet, cloud computing, and the Internet of Things, data—especially personal data—are increasingly collected and transmitted with greater convenience while facing new threats [**?**]. To address technological developments and the inadequacy of existing protection frameworks, the EU officially implemented the General Data Protection Regulation (GDPR) on May 25, 2018 [**?**], creating the EU model for personal data protection. Subsequently, Brazil and California drew on this regulation to formulate the General Data Protection Law and the California Consumer Privacy Act [**?**]. In March 2020, China' s updated and improved "Information Security Technology - Personal Information Security Specification" aimed to further curb illegal collection, misuse, and leakage of personal information, and maximize protection of legitimate rights and interests of individuals and social public interests [**?**]. China' s "Personal Information Protection Law" has also entered the legislative stage [**?**]. Meanwhile, people have recognized that scientific data sharing and reuse have important value in promoting scientific progress and innovation, saving R&D costs, and reproducing scientific research [**?**]. Scientific data often contain personal information, especially in fields such as psychology, clinical medicine, anthropology, and genetics [**?**]. How to protect personal data while opening and sharing scientific data has become an important issue in data governance.

As the most widely adopted personal data protection model and general rules [**?**], is GDPR applicable to personal data protection in scientific data open sharing? What functions does it have? How can China draw on GDPR to strengthen personal data protection in scientific data open sharing? Analyzing and answering these questions will help promote the development of China' s scientific data open sharing practice.

1 Literature Review

In recent years, numerous studies have been conducted on GDPR and personal data protection both domestically and internationally. In addition to introducing GDPR' s background [?], evolution [?], application scope [?], technical and organizational requirements [?], key business aspects [?], and structural chapters and content [?], these studies have also addressed the following themes:

First, the impact of GDPR on personal data protection. GDPR aims to use the distinction between "unambiguous consent" and "explicit consent" as an important means of protecting personal data, explicitly requiring that personal data processing needs to obtain the data subject's "unambiguous consent," while sensitive data requires "explicit consent"[?]. One year after GDPR's implementation, EU member states'data protection authorities have been striving to enforce GDPR' s core principles of responsible and transparent processing and protection of personal data [?]. Undoubtedly, GDPR has set a good example for other countries' personal data protection legislation, with many positive implications: it affects judicial jurisdiction, expands extraterritorial effect of data protection laws, provides rule guidance for international consensus on cross-border data transmission, and promotes sound data protection legislation in other parts of the world [?]18-30; it expands the scope of personal data protection, increases data subjects' rights, imposes heavier obligations on data controllers and processors, and strengthens supervision and relief mechanisms for personal data [?]; it fills some legal loopholes, enabling individuals to better control their personal data while also benefiting the creation of a good competitive environment among enterprises; it ensures more efficient cooperation among member states' regulatory agencies, facilitates communication between enterprises and regulatory agencies, and provides individuals with effective ways to protect their rights when data are infringed upon. However, GDPR also has problems and controversies: some systems (such as the informed consent system) have defects, some issues are not clearly defined, the effectiveness of GDPR' s implementation remains questionable, and there are issues about how to balance personal data protection with other legal interests [?].

Second, key areas or themes of GDPR implementation for personal data protection. GDPR involves all aspects of personal data protection, but focuses on strengthening data subjects' control over data, namely protection of data personality rights and property rights [?]; advocates establishing a Data Protection Officer system to implement personal data protection [?]; and regulates core issues such as cross-border data flow protection [?], data portability right [?], and right to be forgotten [?].

Third, applied research on GDPR in personal data protection across different fields. This includes patient privacy protection [?], cross-border data circulation and protection [?], data protection impact assessment [?], data protection issues in scientific research data publishing [?], smart library user data privacy protection [?], and IoT data protection for data controllers and processors [?].

Fourth, GDPR' s implications for China' s personal data protection. Main suggestions include: China should draw on GDPR' s legislative philosophy to build a dual-track personal information protection system from legal, social governance, and information industry dimensions [**?**]; China should establish a scenario-based and technological data protection concept based on the principle of effectiveness, reasonably introduce a "consent withdrawal" model, and give play to ex-ante supervision and relief functions [**?**]; China' s legislation should build a personal information rights and interests system shared by four parties —information subjects, other natural persons, state organs, and enterprises—to properly resolve conflicts of rights and interests among various parties [**?**]; China should construct a GDPR-based enterprise self-assessment indicator system for personal data protection [**?**]; we should recognize that although GDPR regulates EU personal data protection, sharing personal data between the EU and China is challenging, currently limited to anonymous data or data with data subject consent [**?**]. In summary, GDPR is a legal framework for personal data protection that specifies key principles and safeguards that must be adopted when processing personal data within its scope, and has important reference value for China' s personal data protection. However, there is currently little research on GDPR in personal data protection for scientific data open sharing, although people have found that GDPR has had widespread impact in areas such as genetic data sharing [**?**] and clinical medical research [**?**].

2 Applicability Analysis of GDPR to Personal Data Protection in Scientific Data Open Sharing

Scientific data open sharing aims to promote unrestricted access to and reuse of scientific data to maximize scientific data value and facilitate open research and open innovation. It places scientific data in an open environment, requiring individuals or institutions to implement data protection according to corresponding norms and appropriate measures when opening scientific data. Data protection refers to implementing appropriate administrative, technical, or physical measures to minimize the risk or damage of unauthorized or accidental data leakage [**?**]. Personal data protection aims to protect citizens' rights, personal data integrity, and personal privacy [**?**], involving not only relevant laws, regulations, and policies, but also technologies and systems for collecting, storing, and processing data [**?**]. Is GDPR applicable to personal data protection in scientific data open sharing? To answer this question, we need to understand GDPR' s structure and content.

2.1 GDPR Structure and Content

GDPR is a massive and complex data protection law, structurally divided into recitals (173 articles) and main text (11 chapters, 99 articles), totaling 88 pages and approximately 55,000 words. GDPR provides a comprehensive legal framework for protecting Europeans' personal data and promoting responsible data processing for a range of legitimate purposes. It thoroughly rectifies how or-

ganizations collect, use, and share personal data [**?**]. The main text includes the following 11 chapters [**?**]: 1) General provisions, clarifying GDPR' s main matters and objectives, scope of application, territorial jurisdiction, and definitions of relevant concepts; 2) Principles, clarifying seven principles of personal data protection—lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability; 3) Rights of data subjects, stipulating seven rights of data subjects; 4) Controllers and processors, specifying responsibilities of data controllers and processors; 5) Transfer of personal data to third countries or international organizations, setting requirements for such transfers; 6) Independent supervisory authorities, specifying their independence, general requirements, powers, tasks, and authority; 7) Cooperation and consistency, specifying cooperation, assistance, and joint operations among lead supervisory authorities and other relevant authorities, and consistency mechanisms; 8) Remedies, liability, and penalties, stipulating data subjects' rights to lodge complaints with supervisory authorities, effective judicial remedy against supervisory authorities, controllers, or processors, rights to compensation, and general conditions for administrative fines; 9) Provisions relating to specific processing situations, specifying processing, freedom of expression and information, and safeguards and derogations for processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes; 10) Delegated acts and implementing acts, specifying exercise of delegation and committee procedures; and 11) Final provisions, clarifying repeal of Directive 95/46/EC, relationship with previously concluded agreements, reporting obligations of the Commission, entry into force, and application.

2.2 GDPR' s Applicability to Personal Data Protection in Scientific Data Open Sharing

GDPR involves all aspects of personal data protection, strengthening data subjects' control over data, clarifying concepts, principles, rights, responsibilities, and regulatory requirements for personal data protection, which are equally applicable to personal data protection in scientific data open sharing. This is mainly reflected in four aspects:

First, personal data in scientific data open sharing falls within the scope of protection objects defined by GDPR. Article 2(1) of GDPR stipulates: "This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system"[**?**]. Here, "personal data"means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person [**?**]. Accord-

ing to GDPR' s definition of personal data, it includes not only information that can directly identify data subjects (such as names), but also information that can indirectly identify data subjects when combined with other information (such as phone numbers, passport numbers, biometric characteristics, etc.) [**?**]. Personal data in scientific data open sharing falls within the scope of personal data defined by GDPR, typically appearing in forms such as experiments, measurements, field observations, survey results, interview records, and images [**?**]. Therefore, GDPR can be used to regulate and protect personal data in open sharing.

Second, scientific data open sharing is a special data "processing" activity that can be included in the "processing" activity category defined by GDPR, thus receiving GDPR protection. Article 4(2) of GDPR defines "processing" as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" [**?**]. Scientific data open sharing is a special type of data processing activity. In scientific data open sharing, whether it is data collection, recording, storage, submission, and uploading by researchers, or access, acquisition, transmission, and reuse by users, all constitute processing of personal data. In other words, personal data processing runs through the entire value chain of scientific data open sharing. Therefore, GDPR applies to personal data protection in scientific data open sharing.

Third, GDPR' s "territorial scope" has exceeded the geographical boundaries of EU member states and applies to open sharing environments. Article 3 of GDPR on "Territorial scope" stipulates [**?**]: "1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not. 2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to: (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or (b) the monitoring of their behaviour as far as their behaviour takes place within the Union. 3. This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law." These three provisions, especially Article 3(3), essentially bring open shared personal data into the scope of protection, regardless of whether the data controller is within the EU.

Fourth, GDPR can provide personal data protection for scientific research, and in essence can also provide personal data protection for scientific data open sharing. GDPR does not contain a formal definition of scientific research, but from a broad research concept, states that "the processing of personal data

for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research, and privately funded research" [**?**]. GDPR mentions regulations related to "research" in 14 places in the recitals (such as Articles 26 and 33) and 6 places in the main text (such as Articles 5 and 89). For example, Recital 156 stipulates [**?**]: "The processing of personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be subject to appropriate safeguards for the rights and freedoms of the data subject in accordance with this Regulation. Member States should provide for appropriate safeguards for the processing of personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes." This means that when processing personal data for scientific or historical research purposes, data subjects' rights and freedoms should be protected, and EU member states should provide appropriate safeguards. Similarly, Article 89(1) of the main text stipulates [**?**]: "Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Such safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation." Article 89(2) stipulates [**?**]: "Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes." These provisions essentially require appropriate safeguards for data processing for scientific or historical research purposes, and allow for derogations from data subjects' rights of access, rectification, restriction of processing, and objection, thereby facilitating the sharing and use of such personal data. Since scientific research broadly includes scientific data open sharing activities, and scientific data open sharing can better promote scientific research, GDPR' s personal data protection provisions for scientific or historical research purposes also apply to scientific data open sharing activities.

3 Analysis of GDPR' s Functions in Personal Data Protection for Scientific Data Open Sharing

As an EU member state law, GDPR applies directly to all enterprises and citizens within the EU and has strong extraterritorial effect. Regardless of whether data processing occurs within the EU, as long as the data involve data subjects within the EU, they are subject to GDPR regulation. GDPR constructs a complete personal data protection framework covering scope of application, data protection principles, data subject rights, responsibilities and obligations of data controllers and processors, cross-border data transfer mechanisms, data

regulatory agencies, and administrative penalties. For personal data protection in scientific data open sharing, GDPR' s functions are mainly reflected in the following aspects:

3.1 Clarifying Basic Concepts and Scope of Protection Objects

Article 4 "Definitions" of GDPR defines 26 core concepts, including personal data, processing, restriction of processing, profiling, pseudonymisation, filing system, controller, processor, recipient, third party, consent, personal data breach, genetic data, biometric data, data concerning health, main establishment, representative, enterprise, group of undertakings, binding corporate rules, supervisory authority, relevant supervisory authority, cross-border processing, relevant and reasoned objection, information society service, and international organization. These definitions establish core concepts and basic ideas for personal data protection in scientific data open sharing. For example, applying GDPR' s defined concept of "personal data" to scientific data open sharing means that personal data protection is not just protection of personal property rights data or personality rights data, let alone narrow personal privacy data protection, but broad protection of "any information relating to an identified or identifiable natural person ('data subject')." A clear definition of "personal data" can help people understand what personal-related data needs to be protected in scientific data open sharing.

As a personal data protection law, GDPR distinguishes four different types of personal data: personal data, special categories of personal data, pseudonymised data, and anonymous data. In GDPR, the concept of personal data has a broad scope. Special categories of personal data, also called "sensitive personal data," include: (1) data revealing racial or ethnic origin; (2) political opinions; (3) religious or philosophical beliefs; (4) trade union membership; (5) genetic data; (6) biometric data; (7) data concerning health; and (8) data concerning a natural person' s sex life or sexual orientation. These sensitive personal data pose a higher degree of risk to data subjects and are generally prohibited from processing, only allowed when there is a lawful basis and the processing meets one of the 10 special conditions listed in Article 9(2) of GDPR [**?**]. Pseudonymisation of personal data means processing personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person [**?**]. Pseudonymised personal data still falls within the scope of personal data defined by GDPR and is considered a security safeguard under the concept of technical and organisational measures, but these technologies cannot be used to circumvent compliance obligations under GDPR [**?**]. GDPR does not define anonymous data, but Recital 26 explicitly states: "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or

to personal data rendered anonymous in such a manner that the data subject is not or is no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes" [**?**]. Thus, GDPR applies to pseudonymised data but not to anonymous data. This means that pseudonymised data in scientific data open sharing should be processed according to GDPR, while processing anonymous data does not trigger personal data protection law [**?**]. This distinction has great guiding significance for implementing personal data protection in scientific data open sharing.

3.2 Clarifying Main Principles of Personal Data Protection

Article 5 of GDPR defines the following principles for personal data processing [**?**]: (1) Lawfulness, fairness and transparency (personal data relating to data subjects shall be processed lawfully, fairly and in a transparent manner). (2) Purpose limitation (collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes). (3) Data minimisation (personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed). (4) Accuracy (personal data shall be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay). (5) Storage limitation (kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject). (6) Integrity and confidentiality (processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures). (7) Accountability (the controller shall be responsible for, and be able to demonstrate compliance with, the above principles). These seven principles are also the main principles that must be followed for personal data protection in scientific data open sharing, because according to GDPR's broad definition of processing, transferring personal data for scientific purposes (including scientific data open sharing) is considered a form of processing and therefore should comply with GDPR' s substantive norms [**?**]. In other words, GDPR establishes the main principles that should be followed for personal data protection in scientific data open sharing.

3.3 Establishing the Legal Basis for Personal Data Processing

The existence of a legal basis for data processing is both an important prerequisite for personal data processing and an important guarantee for personal data protection. Article 6(1) of GDPR stipulates six legal bases for data processing, including consent of the data subject, necessity for contract performance, necessity for legal obligation compliance, necessity for protecting vital interests of the data subject or another natural person, necessity for performing a task carried out in the public interest or in the exercise of official authority, and necessity for legitimate interests pursued. Articles 6(2) to 6(4) more detailedly regulate the conditions that must be met for data processing under Article 6(1). These legal bases also provide legitimate and reasonable grounds for personal data processing in scientific data open sharing.

First, for scientific data open sharing, obtaining consent from data subjects is the most secure and basic way to process personal data legally. In GDPR, "consent" of the data subject means any freely given, specific, informed and unambiguous indication of the data subject' s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her [**?**]. This definition indicates that validly obtaining data subject consent must simultaneously meet the conditions of "freely given," "specific," "informed," and "unambiguous," meaning that data subject "consent" is not "broad consent." Article 6(1)(a) stipulates that personal data can be processed legally if "the data subject has given consent to the processing of his or her personal data for one or more specific purposes" ; while Article 9(2)(a) stipulates that special categories of personal data can be processed legally if "the data subject has given explicit consent to the processing of those personal data for one or more specified purposes." In other words, GDPR requires "unambiguous consent" for processing general personal data and "explicit consent" for processing sensitive personal data [**?**]. In 2019, Google was fined €50 million by the French National Commission on Informatics and Liberties for violating GDPR, with the main reasons being Google' s violation of the controller' s information disclosure obligations and failure to effectively obtain user consent [**?**].

Second, in addition to data subjects' "unambiguous consent" and "explicit consent" as the legal basis for personal data processing, public interest and legitimate interests are also legal bases for processing personal data in scientific data open sharing. However, public authorities such as (non-private) universities and public research institutions performing public tasks cannot rely on "legitimate interests" as their legal basis for personal data processing, but must do so under legal authorization or rely on "necessity for the performance of a task carried out in the public interest or in the exercise of official authority" [**?**].

Third, GDPR provides legal basis for secondary data use and cross-border data transfer closely associated with scientific data open sharing. Normally, the strict "purpose limitation" principle applies to personal data that "shall not be further processed in a manner that is incompatible with the specific and legitimate pur-

poses for which they were initially collected and processed." However, Article 5(1)(b) stipulates that further processing for scientific research purposes shall not be considered incompatible with the initial purposes in accordance with Article 89(1). This means that if the same organization or another organization is conducting further processing (i.e., secondary use) of data, there is a rebuttable presumption that further processing for scientific research purposes is compatible with the initially stated processing purpose (e.g., for medical diagnosis). Where permitted by applicable member state law, the scientific research exception allowing exemptions from data subject rights under Article 89(2) will apply to such situations. This essentially establishes a legal basis for secondary data use and its open sharing.

Fourth, scientific data open sharing involves international (or cross-border) data transfer issues, where GDPR' s provisions on international data transfer can be referenced. According to GDPR, there are five ways to legally transfer personal data outside the EU: (1) transfer based on an adequacy decision (Article 45); (2) transfer with appropriate safeguards (Article 46); (3) transfer based on binding corporate rules (Article 47); (4) transfer or disclosure not authorized by Union law (Article 48); and (5) transfer based on derogations for specific situations (Article 49). Since open sharing itself includes cross-border data transfer or international transfer, according to GDPR, data controllers must inform data subjects when collecting personal data about the controller' s intention to openly publish these personal data and decide which appropriate pathway to adopt. In the absence of an adequacy decision, if the data controller or processor has appropriate safeguards and enforceable data subject rights and effective legal remedies, data can still be openly published or transferred to third countries or international organizations [**?**].

Fifth, Article 89 of GDPR provides specific "safeguards and derogations" for personal data processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes—under the premise of adopting technical and organizational measures (such as anonymization) that ensure compliance with the data minimization principle and safeguard data subjects' rights and freedoms, when the data subjects' rights of access, rectification, restriction of processing, and objection may render impossible or seriously impair the achievement of these special purposes, the above data subject rights may be exempted, but only when such exemption is necessary for achieving these purposes [**?**]. In other words, when processing personal data for scientific research purposes, there are some exceptions: for example, data subject consent is not always required, but other legal bases such as contract, legal obligation compliance, public interest, or legitimate interests pursued by the controller or third party; personal data may be stored for longer than necessary for processing [**?**].

In summary, GDPR provides extensive legal basis for personal data protection in scientific data open sharing by offering six legal bases for data processing, particularly clarifying "consent," secondary data use, cross-border data transfer, and exceptions (exemptions) for scientific research.

3.4 Clarifying Main Rights of Data Subjects

Data subjects refer to anyone whose personal data are being collected, stored, or processed [**?**]. GDPR grants data subjects the following eight basic data rights [**?**]: (1) Right to be informed: data controllers shall inform data subjects when collecting personal data relating to them, including the identity and contact details of the controller, the purposes of the processing, and the legal basis for processing personal data. (2) Right of access: data subjects have the right to obtain from the controller confirmation as to whether or not personal data concerning them are being processed, and, where that is the case, access to the personal data and certain information. (3) Right to rectification: data subjects have the right to have inaccurate personal data concerning them rectified without undue delay. (4) Right to withdraw consent: data subjects have the right to withdraw consent previously given for processing their personal data for certain purposes and require processors to stop processing based on previously provided consent. (5) Right to restriction of processing: data subjects have the right to restrict processing by controllers under specific circumstances. (6) Right to object: data subjects have the right to object to processing of their personal data by controllers based on legitimate or reasonable grounds, such as objecting to processing for direct marketing purposes or processing for scientific or historical research purposes or statistical purposes not necessary for performing a task in the public interest. (7) Right to erasure (right to be forgotten): data subjects have the right to have their personal data erased or deleted without undue delay. (8) Right to data portability: data subjects have the right to receive the personal data concerning them, which they have provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided.

In scientific data open sharing activities, data subjects also have the above rights. In other words, personal data protection in scientific data open sharing must be implemented by respecting and protecting these data subject rights. For example, the right to data portability aims to balance the freedom and regulation of data flow, enabling data subjects to migrate data in a concise manner and better control personal data. It consists of two independent claims: the right of data subjects to obtain copies of their personal data, and the right to require controllers and processors to provide personal data to other parties [**?**]. Implementing this right of data subjects is key to implementing scientific data open sharing. Conversely, denying the right to data portability would create many difficulties and obstacles for scientific data open sharing.

3.5 Clarifying Main Responsibilities and Obligations of Data Controllers and Processors

GDPR not only clarifies main rights of data subjects, but also defines main obligations of data controllers and processors. In GDPR, "controller" means the natural or legal person, public authority, agency or other body which, alone

or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law. "Processor" means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller [**?**].

According to GDPR provisions, controllers and processors need to undertake some common obligations, including: (1) controllers and processors shall cooperate with supervisory authorities (Article 31); (2) controllers and processors shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk (Article 32(1)); (3) controllers and processors shall designate a data protection officer when processing is carried out by a public authority or requires large-scale, regular and systematic monitoring of data subjects, or involves large-scale processing of special categories of data and data relating to criminal convictions and offences (Article 37(1)); (4) controllers and processors shall support data protection officers in performing their duties and provide necessary resources (Article 38(2)).

In addition, data controllers must fulfill the following obligations: (1) controllers shall facilitate the exercise of data subject rights (Article 12(2)); (2) when collecting personal data relating to data subjects, controllers shall provide data subjects with relevant information such as the identity and contact details of the controller, contact details of the data protection officer, purposes of the processing, and recipients of personal data (Article 13(1)); (3) except in special circumstances, controllers are responsible for erasing personal data without undue delay upon reasonable request from data subjects (Article 17(1)); (4) controllers shall implement appropriate technical and organisational measures to ensure compliance with GDPR (Article 24(1)); (5) controllers shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed (Article 25(2)); (6) each controller shall maintain a record of processing activities under its responsibility (Article 30(1)); (7) controllers shall notify the relevant supervisory authority of a personal data breach without undue delay and, where feasible, not later than 72 hours after having become aware of it (Article 33(1)); (8) when the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, controllers shall communicate the personal data breach to the data subject without undue delay (Article 34(1)); (9) where a type of processing is likely to result in a high risk to the rights and freedoms of natural persons, controllers shall assess the impact of the envisaged processing operations on the protection of personal data prior to the processing (Article 35(1)); (10) where a data protection impact assessment indicates that the processing would result in a high risk if the controller does not take measures, the controller shall consult the supervisory authority prior to processing (Article 36(1)).

Additionally, data processors must fulfill obligations including: (1) each proces-

sor shall maintain a record of all categories of processing activities carried out on behalf of a controller (Article 30(2)); (2) processors shall notify the controller without undue delay after becoming aware of a personal data breach (Article 33(2)).

In summary, data controllers and processors have obligations to comply with GDPR principles and provisions, including protecting data subject rights, implementing technical and organisational measures, maintaining processing records, notifying data breaches, and conducting data protection impact assessments. In scientific data open sharing, scientific data producers (such as researchers, research institutions, libraries, data centers, publishers, other enterprises, and governments), organizers (such as libraries, data centers, and governments), publishers (such as researchers, research institutions, libraries, industry associations, publishers, other enterprises, and governments), disseminators (such as libraries, data centers, industry associations, publishers, other enterprises, and governments), and managers (such as libraries, data centers, publishers, other enterprises, and governments) [**?**] may all become data controllers or processors. Therefore, GDPR helps clarify the responsibilities and obligations of stakeholders in personal data protection for scientific data open sharing and provides legal guarantees.

4 GDPR' s Implications for Personal Data Protection in China' s Scientific Data Open Sharing

In summary, GDPR mainly establishes a unified data protection legal framework across the EU from the aspects of scope of application, data protection principles, legal basis for data processing, data subject rights, and obligations of controllers and processors. It not only provides legal protection for personal data protection within and outside the EU, but also better maintains the balance between personal data protection and scientific data open sharing. It can provide the following reference and enlightenment for personal data protection in China' s scientific data open sharing:

4.1 Establishing and Improving China' s Personal Data Protection Legal System

The key to effectively implementing personal data protection in China' s scientific data open sharing lies in establishing and improving China' s personal data protection legal system. Currently, China has not enacted specialized personal data protection laws. The main laws, regulations, and documents related to personal data protection include: "Information Security Technology - Guidelines for Personal Information Protection in Public and Commercial Service Information Systems," "Criminal Law Amendment (IX)," "Tort Liability Law," "Consumer Rights Protection Law," "Decision on Strengthening Network Information Protection," "Several Provisions on Regulating Internet Information Service Market Order," "Provisions on Protecting Personal Information of

Telecommunication and Internet Users," "Cybersecurity Law," "General Principles of Civil Law," "Information Security Technology - Personal Information Security Specification," "Administrative Measures for Scientific Data," and the "Civil Code" to be implemented on January 1, 2021. Although the latest "Civil Code" establishes the clause that "personal information of natural persons is protected by law," requiring that "processing of personal information shall follow the principles of legality, legitimacy, and necessity, shall not be excessive, and shall meet the following conditions: (1) obtaining consent from the natural person or his/her guardian, unless otherwise provided by laws and administrative regulations; (2) publicly disclosing the rules for processing information; (3) clearly indicating the purpose, method, and scope of processing information; and (4) not violating provisions of laws and administrative regulations and agreements between parties" [**?**], compared with GDPR, China' s existing provisions have the following deficiencies [**?**]: (1) Different legal texts use inconsistent terminology and provisions for the same or similar content, with unclear boundaries and overlapping repetitions, which is not conducive to mutual connection and coordinated exercise of rights and obligations. (2) The scope of regulated obligation subjects is not comprehensive, without special emphasis on behavior subjects prone to leakage. (3) Obligation items and content are too general, not specific, clear, and concise enough; provisions on leakage are not targeted and detailed enough. (4) There is no formal distinction between privacy rights and data protection rights. (5) No formal and usable judicial remedies are provided for individuals. Therefore, China should establish and improve its personal data protection legal system, including: (1) promptly enacting the "Personal Information Protection Law" and "Data Security Law," which will not only comprehensively protect the storage and use of personal data, but also provide individuals with rights of access, rectification, and erasure, and introduce accountability models [**?**], establishing a legal framework for personal data protection from the aspects of scope of application, data protection principles, legal basis for data processing, data subject rights, and obligations of controllers and processors. (2) Formulating specialized administrative measures for scientific data management and open sharing, clarifying data subjects' rights over their personal data, such as the right to be informed, right of access, right to rectification, right to erasure, right to restriction of processing, right to data portability, and right to object, and specifying detailed norms for processing and open sharing of personal data. On February 11, 2019, the Chinese Academy of Sciences (CAS) took the lead in formulating a normative document within its system—the "CAS Scientific Data Management and Open Sharing Measures (Trial)." This measure clarifies the responsibilities of scientific data open sharing subjects, data submission requirements for research projects, establishes a paper-related data submission mechanism, and plans CAS scientific data centers [**?**], but does not regulate data rights related to personal data protection, making it difficult to effectively guide and implement personal data protection in scientific data open sharing. Therefore, other domestic institutions should incorporate personal data protection into their scientific data management and open sharing measures to achieve the dual goals of promoting scientific data

open sharing and effectively protecting personal data.

4.2 Strengthening Risk Management of Personal Data in Scientific Data Open Sharing

GDPR is risk-oriented, shifting personal data protection obligations and responsibilities to data controllers and processors, and setting multiple obligations for them, including implementing data breach notification and data protection impact assessment, and establishing a Data Protection Officer system. This can not only actively anticipate, identify, and respond to risks, but also better protect data subject rights [**?**]. Therefore, in the process of personal data protection in China' s scientific data open sharing, especially when data processing methods may pose high risks to natural persons' rights and freedoms, data controllers should conduct personal data protection impact assessments for envisaged processing operations. Such assessments can be conducted following the process of data processing, high-risk identification, consulting data protection officers or data subjects, creating data processing operation inventories, and post-implementation review [**?**]. However, the data protection impact assessment process is not a one-time activity, but a continuous activity that data controllers need to perform when processing conditions and circumstances change, especially when data subjects' rights and freedoms face risks [**?**]. In this case, institutions conducting scientific data open sharing (such as government departments, research institutions, libraries, data centers, publishers, and other enterprises) need to establish and appoint Data Protection Officers to supervise compliance with personal data protection policies by data controllers or processors, participate in data protection impact assessments, and provide consultation and advice, thereby providing security guarantees for personal data protection.

4.3 Building a Dynamically Linked and Traceable Scientific Data Open Sharing System

GDPR grants data subjects an extremely systematic and complete set of data rights, requiring data controllers to take active measures to protect these rights to the greatest extent possible. However, due to differences in risk management capabilities and cognitive abilities among data controllers and processors, it is difficult to achieve consistent and high-level data protection. Moreover, because data controllers and recipients often have different purposes and methods of use, GDPR cannot solve the problem of personal data getting out of control after scientific data open sharing. For example, when data subjects exercise their right to erasure, although controllers will take certain measures to notify data recipients, they cannot guarantee the actual realization of the right to erasure. Currently, China' s scientific data open sharing infrastructure is not yet sound, making it difficult to maintain personal data protection in scientific data open sharing at a high level. Consideration should be given to building a dynamically linked and traceable scientific data open sharing system, including a dedicated

open and linked scientific data sharing platform, a dynamic scientific data open sharing registration system, and mechanisms to prohibit data recipients from re-identifying data and to track data users, thereby achieving the perfect combination of scientific data open sharing and personal data protection.

References

[1] LAMBERT P. Understanding the new European data protection rules[M]. Boca Raton: CRC Press, 2018: 35.

[2] THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)[EB/OL]. [2020-08-04]. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[3] BREITBARTH P. The impact of GDPR one year on[J]. Network security, 2019(7): 11-13.

[4] China Electronics Standardization Institute. National standard "Information Security Technology - Personal Information Security Specification"(2020 edition) officially released[EB/OL]. [2020-08-06]. http://www.cesi.cn/202003/6213.html.

[5] WANG Bixue. Personal information protection law has been included in legislative planning[N]. People' s Daily, 2019-06-05(4).

[6] PASQUETTO I V, RANDLES B M, BORGMAN C L. On the reuse of scientific data[J]. Data science journal, 2017, 16(8): 1-15.

[7] PAUL Q, LIAM Q. Big genetic data and its big data protection challenges[J]. Computer law & security review, 2018, 34(5): 1000-1018.

[8] SULLIVAN C. EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross-border data transfers and protection of personal data in the IoT era[J]. Computer law & security review, 2019, 35(4): 380-397.

[9] LI Mingyang. On the EU General Data Protection System and China' s legal response—taking the General Data Protection Regulation as the entry point[D]. Shanghai: East China University of Political Science and Law, 2019: 12-14.

[10] JIN Jing. EU General Data Protection Regulation: evolution, key points and doubts[J]. European studies, 2018, 36(4): 1-26.

[11] European Data Protection Board, AO Haijing. Guidelines on the territorial scope of the General Data Protection Regulation[J]. Journal of international economic law, 2020(2): 135-158.

[12] SHARMA S. Data privacy and GDPR handbook[M]. Hoboken: John Wiley & Sons, Inc., 2020.

[13] ZIEGLER S, EVE QUOZE, HUAMANI A M P. The impact of the European General Data Protection Regulation (GDPR) on future data business models: toward a new paradigm and business opportunities[M]// AAGAARD A. Digital business models: driving transformation and innovation. Cham: Springer Nature Switzerland AG, 2019: 201-226.

[14] DING Xiaodong. What are data rights?—Looking at data privacy protection from Europe's General Data Protection Regulation[J]. ECUPL journal, 2018(4): 39-53.

[15] WANG Xueqiao. On personal data protection and "consent" subdivision in EU GDPR[J]. Legal methodology, 2019(4): 136-146.

[16] WU Linling. Research on EU 2016 General Data Protection Regulation[D]. Wuhan: Wuhan University, 2017.

[17] LIU Jiangshan. Data Protection Officer system in EU General Data Protection Regulation[J]. China science and technology forum, 2019(12): 173-179.

[18] YANG Xue. Research on personal data protection issues in EU law—taking cross-border data flow as the core[D]. Beijing: China Foreign Affairs University, 2017.

[19] SHEN Yuhao. Research on data portability right in EU General Data Protection Regulation[D]. Shanghai: Shanghai International Studies University, 2019.

[20] POLITOU E, MICHOTA A, ALEPIS E, et al. Backups and the right to be forgotten in the GDPR: An uneasy relationship[J]. Computer law & security review, 2018, 34(6): 1247-1257.

[21] LU Bingyang. Research on the right to be forgotten system in EU General Data Protection Regulation[D]. Shanghai: Shanghai Normal University, 2020.

[22] GENG Xi, GU Cuifeng, MA Junjian. Implications of EU General Data Protection Regulation for patient privacy protection in China[J]. Chinese medical ethics, 2019, 32(8): 1000-1003, 1009.

[23] MULDERT, TUDORICA M. Privacy policies, cross-border health data and the GDPR[J]. Information & communications technology law, 2019, 28(3): 261-274.

[24] BIEKER F, MARTIN N, FRIEDEWALD M, et al. Data protection impact assessment: a hands-on tour of the GDPR' s most practical tool[C]// HANSEN M, KOSTA E, NAI-FOVINO I, et al. Privacy and identity management: the smart revolution. Cham: Springer International Publishing AG, 2018: 207-220.

[25] CORTINA S, VALOGGIA P, BARAFORT B. Designing a data protection process assessment model based on the GDPR[C]// WALKER A, O' CONNOR R V, MESSNARZ R. Systems, software and services process improvement. Cham: Springer Nature Switzerland AG, 2019: 136-148.

[26] XU Xin, MAO Lu. Research on data protection issues in scientific research data publishing—based on the enlightenment of EU GDPR[J]. Journal of information resources management, 2010, 10(2): 99-107.

[27] LU Kang, LIU Hui, REN Beibei, et al. Research on user data privacy protection in smart libraries—based on textual enlightenment from the Cybersecurity Law of the People' s Republic of China and the General Data Protection Regulation[J]. Library theory and practice, 2020(3): 17-21.

[28] LOIDEAIN N N. A port in the data-sharing storm: the GDPR and the Internet of things[J]. Journal of cyber policy, 2019, 4(2): 178-196.

[29] LIN Ling, LI Zhaoyi. Dual-track mechanism for personal information protection: legislative enlightenment from EU General Data Protection Regulation[J]. Journalism research, 2019(12): 1-15, 118.

[30] SHANG Xixue. Beyond private rights attributes of personal information sharing—based on analysis of legitimate interest clauses in EU General Data Protection Regulation[J]. Studies in law and business, 2020, 37(2): 57-70.

[31] XU Jicang, AN Xiaomi, SUN Jiarui, et al. Research on GDPR-based enterprise self-assessment indicator system for personal data protection[J]. Library and information service, 2018, 62(23): 113-120.

[32] DEURSEN S, KUMMELING H. The new silk road: a bumpy ride for Sino-European collaborative research under the GDPR?[J]. Higher education, 2019, 78(5): 911-930.

[33] SHABANI M, BORRY P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation[J]. European journal of human genetics, 2018, 26(2): 149-156.

[34] DEMOTES-MAINARD J, CORNU C, GUERIN A, et al. How the new European data protection regulation affects clinical research and recommendations?[J]. Therapie, 2019, 74(1): 31-42.

[35] DE HERT P, GUTWIRTH S. Privacy, data protection and law enforcement: opacity of the individual and transparency of power[M]// CLAES E, DUFF A, GUTWIRTH S. Privacy & the criminal law. Oxford: Intersentia, 2006: 61-104.

[36] BLUME P. The citizens' data protection[EB/OL]. [2020-08-06]. https://warwick.ac.uk/fac/soc/law/elj/jilt/1998_1/blume/.

[37] WALTERS R, TRAKMAN L, ZELLER B. Data protection law: a comparative analysis of Asia-Pacific and European approaches[M]. Gateway East: Springer Nature Singapore Pte Ltd., 2019: 1-15.

[38] DOVE E S. The EU General Data Protection Regulation: implications for international scientific research in the digital era[J]. Journal of law, medicine & ethics, 2018, 46(4): 1013-1030.

[39] LEENES R. Do they know me? Deconstructing identifiability[J]. University of Ottawa law and technology journal, 2007, 4(1/2): 135-161.

[40] EUROPEAN COMMISSION. Guidelines on open access to scientific publications and research data in Horizon 2020 (version 3.2)[EB/OL]. [2020-08-06]. https://ec.europa.eu/research/participants/data/ref/h2020/grants_{manual}/hi/oa_{pilot}/h2020-hi-oa-pilot-guide_{en}.pdf.

[41] MONDSCHEIN C F, MONDA C. The EU' s General Data Protection Regulation (GDPR) in a research context[M]// KUBBEN P, DUMONTIER M, DEKKER A. Fundamentals of clinical data science. Cham: Springer Nature Switzerland AG, 2019: 55-71.

[42] PURTOVA N. The law of everything: broad concept of personal data and overstretched scope of EU data protection law[J]. Law, innovation and technology, 2018, 10(1): 40-81.

[43] SHEN Xiaoyu, WU Yaohan. Looking at GDPR data privacy protection from Google' s fine[J]. Legal person, 2019(4): 98-100.

[44] RADBOUD UNIVERSITY. FAQ GDPR in research[EB/OL]. [2020-08-01]. https://www.ru.nl/rdm/gdpr-research/faq-gdpr-research/.

[45] What is a data subject?[EB/OL]. [2020-08-06]. https://eugdprcompliant.com/what-is-data-subject/.

[46] SHENG Xiaoping, WU Hong. Analysis of different stakeholders'motivations in scientific data open sharing activities[J]. Library and information service, 2019, 63(17): 40-50.

[47] Civil Code of the People' s Republic of China[EB/OL]. [2020-08-07]. http://www.npc.gov.cn/npc/c30834/202006/75ba6e483b8344591abd07917e1d25cc8.shtml.

[48] DIAO Shengxian, HE Qi. On legal countermeasures for personal information leakage in China—comparative analysis with GDPR[J]. Science technology and law, 2019(3): 49-57.

[49] HERT P, PAPAKONSTANTINOU V. The data protection regime in China: in-depth analysis[R]. Brussels: European Union, 2015.

[50] KATULIC T, KATULIC A. GDPR and the reuse of personal data in scientific research[C]// SKALA K, KORICIC M, GRBAC T G, et al. 2018 41st international convention on information and communication technology, electronics and microelectronics (MIPRO). Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics-MIPRO, 2018: 1311-1316.

[51] XIAO Dongmei, TAN Lige. EU data protection impact assessment system and its enlightenment[J]. Journal of library science in China, 2018, 44(5): 76-86.

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