Overview of Regulations on Cross Border Data Flow

Luyao Sun1,a

1School of Law, Beijing Normal University, Beijing, China
a lu_yao_sun@163.com

Abstract: The high importance and immense power of data, coupled with a lack of consensus on the definition of this term, have led to multiple challenges in this field. Data protection legislation can be traced back to the Enlightenment era, which explains why the EU has always regarded personal data protection as a fundamental human right. As the third generation of personal information protection legislation, GDPR not only expands the extraterritorial application effectiveness, but also strengthens the administrative sanctions function of DPA. The EU has considered commitments regarding cross-border data flows in recent trade agreement negotiations, indicating a trend towards repositioning its considerations between data protection and trade interests. The primary issue facing the regulation of cross-border data flow under the WTO framework is to clarify the applicability of GATS commitments to digital trade and cross-border data flow. At the FTA level, few trade agreements explicitly address the issue of data localization requirements. For the first time, CPTPP sought to explicitly restrict the use of data localization measures, with the exception of "legitimate public policy objectives", and this template was subsequently followed by FTA. In response to the regulation of cross-border data flow in China, scholars have put forward their own opinions from different perspectives, from the objects of learning and reference to specific suggestions.

Keywords: Data related definitions; New Trends in EU Regulation; GATS applicability; CPTPP template; Selection of reference objects.

1. Introduction

The cross-border flow of data is a hallmark of globalization in the 21st century. The flow of physical goods and funds was a characteristic of economic globalization in the 20th century, but today, these flows have either stabilized or decreased. Globalization in the 21st century is increasingly defined as the flow of data and information. [1] The cross-border flow of data constitutes the lifeline of service trade, which in turn supports manufacturing and commodity trade. In the past few years, cross-border data flow has been a hot topic. This article aims to identify the current research status and future trends related to the regulation of cross-border data flow through the review and induction of relevant literature, in order to better grasp the forefront of relevant academic research and provide guidance for further research in the future. The literature sources include Chinese and foreign language databases such as CNKI, Peking University's magic weapon, EBSCO, PROQUEST, WOS, as well as authoritative websites, search engine statistics, government official data, and other online resources in this field.

The article is divided into four parts. The first part "Overview of Cross border Data Flow" mainly introduces the basic issues related to data that are still in the discussion stage, such as the lack of unified definitions and diverse standards for data type classification. The second part "Regulations on Cross border Data Flow" mainly introduces the current status of legislative regulations from three levels. Firstly, as a benchmark for global data legislation, the EU's stance and legislative progress in the field of cross-border data flows have attracted attention from various countries. The EU has always attached great importance to personal data protection, and the implementation of GDPR still maintains a high level of data protection. However, in recent trade negotiations, there has been a slight shift in the balance between data protection and trade interests. Secondly, the applicability of GATS to digital trade disputes and measures to restrict cross-border data flow under the WTO framework tests the current structure and function of the WTO. Finally, FTA became an alternative choice for member states to address cross-border data flow issues when the effectiveness of WTO e-commerce negotiations was minimal, with the CPTPP template being followed and disseminated. The third part, "China's Response," mainly introduces the views and perspectives of Chinese scholars on the regulation of cross-border data flow, including both drawing on differences in learning objects and similarities in specific regulatory methods. The fourth part, "Summary and Prospects," proposes further research directions based on the summary of the entire text.

2. Overview of Cross-border Data Flow

2.1. History of data

(1) The early stages of data collection. For thousands of years, people have been collecting data. The oldest census can be traced back at least to ancient Egypt, Greece, and China from 2000 to 1000 BC, where they conducted statistics on population, livestock, and food. In the long history of data collection, with changes in social priorities, values, power structures, and government goals, the types of data collected and the ways in which data is used have also changed. Between 1400 and 1500 AD, the rulers and administrators of the Inca Empire prioritized strategically important sources of wealth and power (which were kept confidential to the public). They first collected property information required for taxation, as well as personnel information required for military recruitment and labor, and listed newly conquered peoples and territories.

(2) From the 18th century to the 20th century. The rise of Enlightenment thought in Europe in the 18th century emphasized objective scientific exploration. Under the influence of major intellectuals at that time, the concepts of rule of law and national responsibility gradually formed, and the social contract between individuals and the state gradually
merged, resulting in the emergence of the Declaration of Human Rights. Starting from the late 18th century, governments of emerging ethnic countries in Europe and North America established statistical agencies to publish official statistical data on the state of the country and provide information for public discussion. Modern geospatial data systems are based on older cartography. Since the advent of Snow Maps, the innovation of printing and computer technology, as well as the rise of remote sensing technology, have made geospatial data and its applications universal and ubiquitous.

(3) After the 20th century. With the digital revolution, the types and scope of data have undergone significant changes, and the amount of data collected has grown exponentially. In this new situation, private sector participants are playing an increasingly important role in data collection through platform based business models, in which data is passively collected as a by-product of business processes. Digital platforms also expand the opportunities for citizens to collect data, which typically occurs when the government is unable to collect data. For example, the platform "Utunzi" allows individuals and organizations to report and record violent behavior against LGBTQI individuals, as well as various platforms that allow users to report air pollution levels, deforestation, and other environmental data at specific locations to raise awareness and motivate action. The basic origin of data protection laws can be traced back to the Enlightenment era, and the principles of data protection can be traced back to the United States' Principles of Fair Information Practice formulated in the 1970s, which formed the basis of the 1980 OECD Guidelines on Privacy Protection and Cross border Movement of Personal Data (hereinafter referred to as the "Guidelines"). Similarly, the fundamental substantive rights and obligations in the EU GDPR are first reflected in its 1995 Directive 95/46/EC on the Protection of Individuals Involved in the Processing of Personal Data and the Free Movement of such Data (hereinafter referred to as the "Directive"), which can be traced back to the Guidelines.

2.2. Type of data

2.2.1. Academic perspectives

Scholar Burri pointed out that so far, there is no difference between different types of data, such as personal and non personal data, personal or corporate data, or machine to machine data. Scholar Sen divided data into the following categories: personal data refers to data involving individuals; Company data refers to data that flows between enterprises; Business data refers to digital content such as software and audiovisual content; Social data refers to behavior patterns against LGBTQI individuals, as well as various platforms that allow users to report air pollution levels, deforestation, and other environmental data at specific locations to raise awareness and motivate action. The basic origin of data protection laws can be traced back to the United States' Principles of Fair Information Practice formulated in the 1970s, which formed the basis of the 1980 OECD Guidelines on Privacy Protection and Cross border Movement of Personal Data (hereinafter referred to as the "Guidelines"). Similarly, the fundamental substantive rights and obligations in the EU GDPR are first reflected in its 1995 Directive 95/46/EC on the Protection of Individuals Involved in the Processing of Personal Data and the Free Movement of such Data (hereinafter referred to as the "Directive"), which can be traced back to the Guidelines.

2.2.2. Perspectives of international organizations

(1) The OECD points out that there is no "correct" method to classify data into different types. One possible approach that may be relevant to decision-making is to distinguish the parties involved in data flow and approximate the personal content of these data.

(2) WB divides data types from two dimensions. The first dimension is the original intention of data collection, which can be divided into private intention data and public intention data; The second dimension is the method of data collection, which can be divided into new collection methods and traditional collection methods.

(3) UNECE classifies big data as follows: ① Social networks (human resource information): This information is a record of human experience, previously recorded in books and art works, and later recorded in photos, audio, and videos. Human resource information is now almost completely digitized, ubiquitous from personal computers to social networks. Data is a loose structure and often uncontrolled. ② Traditional business systems (process mediation data): These processes record and monitor related business events, such as registering customers, producing products, accepting orders, etc. The process mediation data collected from this is highly structured, including transactions, reference tables, and relationships, as well as metadata that sets their context. In operating systems and BI systems, traditional business data accounts for the vast majority of IT management and processing. ③ Internet of Things (machine generated data): This stems from the significant increase in the number of sensors and machines used to measure and record events and situations in the physical world. Machine data is the product of these sensors, whether it is simple sensor records or complex computer logs, these data have good structure. With the popularization of sensors and the growth of data scale, machine data has gradually become an important component of the information stored and processed by many enterprises. It has a good structure and is suitable for computer processing, but its size and speed exceed traditional methods.

2.3. Data related definitions have not been unified yet

2.3.1. Cross border data flow

The cross-border flow of data was first proposed by the Computer Applications Working Group under the OECD Science and Technology Committee in the 1970s. The earliest authoritative definition came from the OECD Guidelines in 1980, and currently the vast majority of literature refers to "cross-border data flow" as "personal data flow". Article 4 of the GDPR provides a broad definition of personal data, which means "personal data" refers to any information related to a natural person ("data subject") who has identified or can be identified; An identifiable natural person refers to a natural person who can be directly or indirectly identified, especially through identifiers such as name, identification number, location data, online identifier, etc., to identify their cultural or social identity. Scholars Yakovleva and Irion point out that the uniqueness of personal data lies in its combination of human dignity with valuable economic assets for commercial activities. Scholar Burri pointed out that there is no consensus on the definition of data flow in free trade agreements, despite the frequent use of this term in reports and research. However, despite the use of different terms in treaty language, there seems to be a clear trend towards a
broad and comprehensive definition of data flow: ① when information (data) is included in the provision of services or products; ② When the data crosses boundaries, although the data flow is not entirely consistent with a commercial transaction, and the provision of a certain service may involve multiple data flows. Scholars such as Gorlevskaya pointed out that the economic value created by cross-border data flows exceeds traditional trade commodity flows, and it is a broad term that has been proposed by scientists and practitioners, which can lead to difficulties in understanding. This term is often considered a buzzword or meme. The definition of big data is also an immature term, with over 50 definitions. For example, according to scholars, data is an information asset with high capacity, speed, and/or diversity, requiring cost-effective and innovative forms of information processing to enhance insight, decision-making, and process automation; ③ Definition from NIST: Big data contains a large number of datasets, mainly reflected in aspects such as data volume, diversity, speed, and/or variability, and requires a scalable architecture for effective storage, operation, and analysis; ④ Definition from McKinsey: A dataset that exceeds the capabilities of typical database software tools to capture, store, manage, and analyze; ⑤ Definition from ISO: Big data can be classified as information technology, but its definition is still under research.[4]

2.3.2. E-commerce and Digital Trade

Scholars Yakovleva and Irion point out that digital trade is a broader concept than pure e-commerce, and there is no unified definition of e-commerce or digital trade. In e-commerce, the flow of personal data is understood as an assistance provided by services, while in digital trade, personal data itself can become the subject of transactions. The OECD points out that there is currently no unified definition of digital trade, but there is an increasing consensus that digital trade includes digital transactions delivered in digital or physical form, involving consumer, enterprise, and government trade in goods and services. That is to say, although all forms of digital trade are achieved through word technology, not all digital trade is delivered in digital form, such as purchasing books through online markets or booking apartment accommodations through matching applications. The Singapore Ministry of Trade and Industry pointed out that there is no unified definition for digital trade agreements, and different jurisdictions use different terminology to describe digital trade agreements. For example, Singapore uses DEA (Digital Economy Agreement) instead of DTA, and in its signed agreements, DEA is defined as an agreement to establish digital trade rules and digital economic cooperation between two or more economies. The specialized digital economy agreements currently signed by Singapore include: ① Singapore Chile New Zealand Digital Economy Partnership Agreement (DEPA); ② Singapore Australia Digital Economy Agreement (SADEA).[5]

3. Regulations on Cross-border Data Flow

3.1. EU: Global Data Legislation Benchmark

3.1.1. Personal Data Protection Concept

(1) A tradition of high-level protection. Due to historical, cultural, institutional and other reasons, the European Union places greater emphasis on personal privacy protection than other regions. Europe is the birthplace of personal information protection laws, and the legislative leadership in protecting personal privacy is not a local phenomenon. The European Union has always believed that the right to personal information is a fundamental human right that must be recognized and protected through legislation, which also constitutes the EU’s basic position on cross-border data flow regulations. The EU adheres to high standards of privacy rules in the online environment, clarifies that any commitment regarding cross-border data flows cannot take priority over privacy protection, and regards it as a fundamental right. The EU clearly regards personal data as a fundamental human right, while in most other countries and jurisdictions such as the United States, personal data is not protected as such, indicating different international perceptions of the right to protect personal data. (2) The repositioning of the EU in data protection. Recently, the European Union has taken a step towards binding data protection rules, and all parties have agreed to consider commitments regarding cross-border information flows in future negotiations. This type of provision was reflected in the trade modernization section of the 2018 EU Japan EPA and the EU Mexico Global Agreement. In the latter two agreements, each party undertakes to re-evaluate whether it is necessary to include provisions on the free flow of data in the treaty within three years after the agreement comes into effect. This marks a repositioning of the EU on data flow issues, and the EU hopes to link this commitment with GDPR’s high standard data protection in due course.

3.1.2. Third generation data legislation: GDPR

(1) GDPR strengthens its administrative sanctions function. Article 83.1 of the GDPR stipulates the standards for administrative sanctions, namely effectiveness, dissuasion, and moderation. Scholars Voss and Hugues point out that under previous data protection legislation, the maximum fines for company violations are relatively low, which may result in a lack of compliance among US technology giants and other companies. According to GDPR, this situation will change. Although there are other methods such as judicial compensation and orders to stop data processing, the most typical form of GDPR sanctions is effective and significant administrative fines. The initial recommendation of the European Commission on GDPR was to grant DPA the power to impose administrative sanctions and to divide administrative fines into three levels, with the first level being 250000 euros, or 0.5% of global annual revenue; The second level is 500000 euros, or 1% of global annual revenue; In the most severe cases, the highest level of third tier, which is 1 million euros, or 2% of global annual revenue, applies. In the first reading of 2014, the European Parliament requested "regular data protection audits" in addition to the aforementioned sanctions, and rejected the floating ratio law, raising the maximum level of administrative fines to 100 million euros, or 5% of the annual global turnover of enterprises. The sanction power in data protection law is considered a revolution, and granting DPA sanction power is part of the overall trend of European law, aimed at strengthening sanctions and serving more effective market regulation.[6]

(2) GDPR expands its extraterritorial effects. Scholar Voss pointed out that data privacy laws with extraterritorial applicability already exist and are being adopted in different
regions of the world, as is the case with GDPR in Europe. Firstly, according to Article 3.1 of GDPR, GDPR applies to the processing of personal data in the activities of a controller or processor, regardless of whether the processing activity is carried out within the EU, as long as the controller or processor is established within the EU. Secondly, according to Article 3.2 of GDPR, GDPR applies to certain companies, even if they do not have institutions in the EU, as long as their processing activities involve: ① providing goods or services to data subjects within the EU, regardless of whether the data subject needs to make payments; If their behavior occurs within the EU, their behavior will be monitored. Finally, according to Article 3.3 of GDPR, GDPR applies to the processing of personal data by the controller, even if the controller does not have an institution within the EU, but the controller has an institution in a place where the laws of the member state can be applied in accordance with the principles of public international law.[7]

3.2. Legal Adjustments under the WTO Framework: The Applicability of GATS's Commitment to Digital Trade and Cross-border Data Flow

Under the framework of the WTO, important agreements related to e-commerce include GATS, GATT, TRIPS agreements, etc. Due to the limitations of technological development during negotiations, there is a significant lag in the regulation of e-commerce in these agreements. Compared to today's e-commerce coverage, the level of liberalization under GATS is very limited, and two issues test its applicability to e-commerce: service classification and China's asymmetric internet access. The author believes that GATS does not distinguish between online and offline services, and its regulations apply to government measures related to e-commerce; The six market access barriers defined in Article 16 of GATS cover digital protection measures, and the comprehensive blocking of foreign websites caused by asymmetric internet speeds in China can be considered as "zero quota", constituting the "limit on the number of service providers" under Article 16.2 of GATS.

Scholar Razon discussed the main issues faced when applying GATS to emerging services from the perspective of "the application of blockchain and GATS" in his article, that is, how GATS applies to new services that do not exist when they pass through. The reason for this problem is due to inappropriate and outdated service classification under GATS. Nevertheless, the author believes that many new services can still be accommodated in the GATS classification table and the broad definition of CPC (1991), especially in the latest version of CPC (2015), which involves internet and digital services and can provide better guidance for member states. At the same time, the author points out that blockchain, which relies on cross-border data flow, is an emerging digital service. Applying GATS to blockchain means that many prohibitive regulations imposed by the state - from prohibiting mining to criminalizing the possession of cryptocurrency - may be seen as trade barriers.[8] Scholar Ferracane pointed out that so far, no WTO member has filed a lawsuit against another member for violating GATS on the grounds of data flow restrictions, and the debate on when to consider measures to manage data flow as a violation of GATS is still in its early stages. The text of the World Trade Organization does not mention the Internet, censorship, e-commerce, or data flow at all, but it seems that there is a consensus on the fact that digital trade can be included in the GATS clause. This is mainly based on the interpretation of WTO rulings and GATS Council documents as part of the e-commerce work plan. When examining online service trade restrictions, there are three particularly relevant cases: the US gambling case China Publishing and Audiovisual Products Case and China Electronic Payment Service Case.[9]

4. China's Response

4.1. Selection of reference objects

(1) The US European model. From the perspective that China currently does not have an independent regulatory system for cross-border data flow, scholars Huang Ning and Li Yang believe that learning from the regulatory methods of the European Union and the United States while promoting domestic regulatory practices is the best way. They also suggest: ① establishing a data protection agency and applying to join the CBPR system. In 2004, APEC passed the Privacy Framework, with 9 basic principles corresponding to the 8 principles of the Guidelines; The CBPR, built on the APEC Privacy Framework in 2012, was officially launched, introducing privacy enforcement agencies and accountability agencies. It is one of the few data protection initiatives that the United States has joined and supported Multinational companies with business dealings with the European Union use SCC or apply for BCR certification. The US European model has proven to be effective. The exception clauses for national security, the 2016 European Commission pointed out in its report that many new measures are ostensibly based on "national security" reasons, including in the field of cybersecurity, such as the National Security Law, the Anti Terrorism Law, the draft Cybersecurity Law, etc. Most of this legislation goes beyond basic national security concerns and includes a broad and unclear definition of national security, causing legal uncertainty. Forcing companies to hand over sensitive data to authorities and the widespread risk of imposing unnecessary restrictions on commercial activities is inconsistent with China's commitment to a "predictable and open investment environment" and may reduce China's trade, investment, and innovation.

(2) Non US European mode. Scholar Zhao Jun pointed out that China, the United States, the European Union, and other countries have different regulatory logic and have formulated different rules in sequence, and there are significant differences in the principles of cross-border data flow regulation among all parties. Scholar Hong Yanqing believes that in the regulation of cross-border data flow: ① The EU paradigm is: GDPR, as the third generation of personal information protection legislation in the EU, its core goal is to require data recipients to provide a "substantial equivalent" level of protection to GDPR, and sufficient determination, SCC, and BCR are important legal tools in cross-border data flow; ② The US paradigm is to implement a low-level protection CBPR system at the international level, allowing data to be aggregated back to the US, while at the domestic level, avoiding data access by specific country entities through systems such as foreign investment review. The EU adheres to its own data protection model as the norm and requires other countries to follow suit; The United States, based on political and economic interests, compresses the regulatory space for countries to independently choose the level of data protection, which is not in line with China's
stance and nature. At the same time, the author suggests that when signing the "the Belt and Road" cooperation document, the cross-border flow terms of RCEP data should be taken as the blueprint to leave enough space for the development interests of all countries.

4.2. Specific recommendation

(1) Data classification or classification management. Scholar Shi Jingxia pointed out that China can advocate for setting different flow standards for different types of data in WTO e-commerce negotiations, and construct a gradient cross-border system based on data security attributes. Scholars Gao Gaogao and Liu Weiqi suggest classifying and managing data, and for data related to national security and public safety, the standardized starting point is collection; For data related to the enterprise itself, when it involves public interests, it is recommended to review it when leaving the country, taking into account efficiency and technological limitations, and adopt a filing system; For data involving data subjects, clear notification and consent from the data subjects should be obtained during cross-border flows.

(2) The theory of reasonable boundaries for localized data storage. Scholar Hong Yanqing proposed the "Reasonable Boundaries Theory of Data Localization Storage", which is based on three foundations: ① the strictness model of data localization storage, which has four indicators, namely the implementation subject of localization storage, the thorough degree of localization storage, the data scope covered by localization storage, and the exemption conditions for localization storage; ② The goals achieved by localized data storage include data security, personal data protection, and national level data protection; ③ The appropriateness and necessity relationship between purpose and means. Based on the theory of reasonable boundaries for localized storage of data, the author proposes a design scheme for the security evaluation method of cross-border data flow: firstly, the evaluation process is involved. Organizations with cross-border data transmission needs first conduct self-assessment, and then the competent department makes an audit. Finally, the requesting organization forms specific arrangements for cross-border transmission according to the requirements of the competent department; ② Secondly, the substantive content of the evaluation is to only adopt a "light supervision" model for data related to data security. If the data involves personal data protection, it is required to ensure personal information self-determination and higher data security thresholds; For data related to national security, the public has the power to carry out "strong supervision", discuss each matter, and directly intervene in specific scenarios.

(3) In conducting cross-border data retrieval in special law enforcement areas, scholar Hong Yanqing pointed out that the US model is an extension of turning its own enterprises into territories, as evidenced by its quickly passed "Clarification of the Legitimate Use of Overseas Data Act" in 2018; The EU model relies on the entire single digital market, allowing multinational enterprises to voluntarily comply with its regulations when entering the EU market. In the face of the "expansionary" law enforcement cross-border data retrieval model in the United States and Europe, the author suggests: ① China may need to take more thorough "defense" measures; ② Adopting an "offensive" posture is actually more likely for all parties to reach a compromise; ③ Finally, in the hedging of different models between China, the United States, and Europe, the best strategy for NetEase companies can only be to localize and store data, or even cut the same product into domestic and foreign versions. At the same time, it is pointed out that only adopting defensive strategies is not the best way out. China should divert the pressure of unilateral response to a multilateral framework for resolution.

5. Conclusion

This article reviews and summarizes the relevant literature on the regulation of cross-border data flow. (1) From the perspective of basic issues related to data, research on the regulation of cross-border data flow is still in its early stages. There is no unified definition of big data, personal data, cross-border data flow, e-commerce and digital trade, and digital trade agreements; Regarding the types of data, there are various classification criteria provided by scholars and relevant authoritative organizations, resulting in different classification results. ② At the WTO level, the issue of service classification tests the applicability of GATS to emerging services such as cross-border data flow. Although consensus has been reached through WTO practice and interpretation of relevant documents that digital trade, cross-border data flow, etc. can be included in GATS clauses, there is still a lack of clear confirmation in WTO texts At the FTA level, only a few trade agreements explicitly address the issue of restrictions on cross-border data flows. The CPTPP first stipulated measures for free data flow and data localization, with the exception of "legitimate public policy objectives", becoming a template for subsequent FTAs to follow. (3) From the perspective of China's response, domestic scholars have demonstrated from different perspectives that the EU and the United States can draw inspiration from in regulating cross-border data flows. Some scholars also suggest adhering to China's true nature and using RCEP as the blueprint in trade negotiations.

On the one hand, the multilateral trading system under the WTO remains an ideal place for cooperation. The rise of online services and the development of e-commerce have put forward new requirements for WTO rules, making the existing rules need to be updated and existing negotiations more urgent. WTO member states can actively negotiate on cross-border data flows within a multilateral framework, striving to reach consensus on relevant basic issues as soon as possible. On the other hand, data classification or classification management is an effective measure to regulate cross-border data flow, which will deepen our understanding of the essence of this term. At present, it is necessary to conduct in-depth research on classification standards in order to achieve effective classification or grading of data science. On this basis, a thorough and not chaotic regulatory network should be constructed to achieve a balance between free data flow and data protection, and to coordinate domestic rules with international legislation.

References

[2] Classification of Types of Big Data - Classification of Types of Big Data - UNECE Statswiki[EB].


