CAICT 中国信通院

Big Data White Paper

(2019)

CHINA ACADEMY OF INFORMATION AND COMMUNICATIONS TECHNOLOGY

December 2019

	Strengthening Online Information Protection", Among others.
Judicial Interpretations	The “Interpretation on Several Issues Concerning the Application of Law in Handling Criminal Cases Involving Infringement of Citizens' Personal Information”, the “Provisions of the Supreme People`s Court on Several Issues Concerning the Application of Law in the Trial of Cases of Civil Disputes over the Use of Information Networks to Infringe Personal Rights and Interests”, among others.
Departmental Rules	The "Provisions on the Protection of Personal Information of Telecommunications and Internet Users", the "Notice of the People's Bank of China Regarding the Effective Protection of Personal Financial Information by Financial Institutions", among others.
Administrative Regulations	Regulation on the Administration of Credit Investigation Industry, among others.
General Laws	There are also relevant provisions concerning the protection of personal information in general laws such as the General Principles of Civil Law, Criminal Law Amendment (IX), the Tort Liability Law, the Law on Protection of Consumer Rights and Interests, and the Anti-Terrorism Law.

Source: CAICT

Since 2019, the legislative process on data security has significantly accelerated. The Cyberspace Administration of China (CAC) has issued the drafts of four administrative measures on data security for comments. Among them, the "Provisions on Online Protection of Children’s Personal Information" has been officially promulgated and will be implemented on October 1. The formulation of a series of administrative regulations has attracted strong public attention to data security.

Table 6 - Legislative Process of Data Security Related Laws and Regulations in 2019

Time	Main Contents
May 24	CAC released Cybersecurity Review Measures (Draft for Comments)
May 28	CAC released the Data Security Administrative Measures (Draft for Comments)
May 31	CAC released the "Provisions on Online Protection of Children’s Personal Information" (Draft for Comments) (It was officially promulgated on August 23 and will be implemented on October 1.)
June 13	CAC released the Security Assessment Measures for Cross-Border Transfer of Personal Information (Draft for Comments)

Source: CAICT

However, it is no doubt that from the perspective of the legal and regulatory system, China's data security laws and regulations are still incomplete and there are many problems such as the lack of comprehensive and unified laws, the lack of interpretations of legal details, and insufficient coordination between protection and development. In 2018, there are two draft laws with “mature conditions and planned to be submitted for review during the term of office” in the legislative plan of the Standing Committee of the 13th National People's Congress - the Personal Information Protection Law and the Data Security Law. The era of comprehensive legislations on personal information and data protection is coming.

2. Data Security Technologies Facilitate the Landing of Big Data Compliance Requirements.

The concept of data security derives from the concept of traditional information security. In traditional information security, data is the content while information system is the carrier. Data security is the focus of the entire information security. The main content of information security is to ensure the confidentiality, integrity and availability of data through security technologies. From the perspective of the data life cycle, data security technologies cover sensitive data identification and detection, data classification, staging and labeling, and data quality monitoring at the data collection stage; data encryption and data backup and disaster recovery at the data storage stage; data masking[21], Secure Multi-Party Computation (SMC) [22], federated learning[23] at the data storage stage; full-copy data destruction at the data deletion phase; users, roles and permissions management, data transmission verification and encryption, and data activity monitoring and auditing throughout the data life cycle.

At present, China's data security laws and regulations focus on the protection of personal information, and the compliance of the big data industry will certainly take this as its core. Among the current data

[21] Refers to the process of masking sensitive data through masking rules to achieve protection for sensitive data.
[22] Refers to a technology to compute a public function with each party's private input such that in the end only the evaluation result is known and the private inputs are not exposed.
[23] Refers to the technologies used by multiple institutions to perform data use and machine learning modeling while meeting user privacy protection requirements.

security technologies, there are many technologies aimed at the protection of sensitive data in processing and use, such as data masking, SMC, federated learning, and so on.

The Data Security Administrative Measures (Draft for Comments) clearly requires that the provision and storage of personal information should be anonymized, and data masking technology is an effective way to achieve data anonymization. The application of static data masking (SDM) technology can ensure that the data is released to the public without involving sensitive information and that it is possible to perform normal mining and analysis without changing the characteristics of the sensitive data set in the development and testing environment. The application of dynamic data masking (DDM) technology can ensure the ability to return data requests in real time while eliminating the risk of leaking sensitive data.

Technologies such as SMC and federated learning can ensure that the computing tasks are completed and the correct computing results are obtained without the actual data of any party being exposed to other parties in the collaborative computing. The application of these technologies, on the one hand, can effectively protect sensitive data and personal privacy data from being leaked and on the other hand can complete data analysis, data mining, machine learning, and other tasks that were originally required to be performed.

The above technologies are currently the most mainstream data security protection technologies, and they are also the most conducive to the landing of big data security compliance. Each of these technologies has their own technical implementation approaches, application scenarios, technical advantages, and existing problems. The specific comparison is as follows:

Table 7 - Comparison of Main Privacy and Data Protection Technologies

Technology Type	**Data Masking**	**Secure Multi-Party Computation (SMC)**	**Federated Learning**

Implementation	Encryption, masking, k-anonymity, l-diversity, etc.	Homomorphic encryption, secret sharing, oblivious transfer, garbled circuit	同态加密、混淆电路、可信计算环境等 Homomorphic encryption, garbled circuits, trusted computing environments, etc.
Classification	Static data masking, dynamic data masking	Secure two-party computation, secure N-party computation, etc.	Horizontal federated learning, vertical federated learning and federated transfer learning
Application Scenarios	External data services, data development, data mining, etc.	Numerical computing, aggregation operations, SQL queries, machine learning, etc.	Machine learning
Technical Advantages	Wide application scenarios, high computing efficiency, diverse implementation methods, and partial data availability	High degree of privacy protection, no loss of data availability, and wide application scenarios	High degree of privacy protection and no loss of data availability
Problems	Sensitive data identification, data masking degree, and data availability need to be balanced.	Loss of computational performance, customized application scenarios	Loss of computational performance, limited available scenarios

Source: CAICT

There are multiple technical implementation methods for the above technologies. Different implementation methods may achieve different levels of protection for private data. Different application scenarios also have different requirements for the protection and availability of private data. As the main technology to help the landing of big data security compliance, users should choose the appropriate privacy protection technology and appropriate implementation method according to the specific application scenario in actual application. However, due to the existence of a diversity of implementation methods and the difference in

productized function points, it is very difficult for a technical user to choose the appropriate protection technologies. Standardizing the corresponding privacy protection technologies can effectively deal with this situation[24].

With the continuous development of the big data industry in the future, laws and regulations related to personal information and data security will continue to be introduced. In terms of corporate compliance, the application of standardized data security technologies is a very effective means for compliance. With the improvement of public awareness on data security and the continuous advances of technologies, data security technologies will gradually see a trend of standardization. Developing relevant product technical standards with reference to relevant laws and regulations, applying data security technology products that meet the corresponding technical standards and ensuring the legal compliance of the use of sensitive data and personal privacy data will become a major trend of compliance landing in the big data industry in the future.

3. The Framework of Data Security Standards and Specifications Continues to Improve.

Compared with laws and regulations and standards for data security technologies, standards and specifications also play an irreplaceable role in big data security protection.

The "Information Security Technology - Personal Information Security Specification" (the “Specification”) is an important recommended standard in the field of personal information protection. The Specification combines the internationally accepted personal information and privacy protection concepts, and proposes the seven principles: "consistency between rights and responsibilities, clear purpose, opt-in consent, minimization, transparency, security, and subject engagement". It provided the enterprises with more detailed guideline on how to improve the internal personal information protection system and rules of practical operations. On June 25, 2019, the revised standard (draft for comments) was formally released.

---

[24] Please refer to: "White Paper on Key Technologies of Data Circulation", CAICT, 2017.

A series of national standards focusing on data security have been released in recent years, including the "Security Capability Requirements for Big Data Services" (GB/T 35274-2017), "Big Data Security Management Guide" (G/T 37973-2019), "Data Security Capability Maturity Model" (GB/T 37988-2019), " Security Requirements for Data Trading Services" (GB/T 37932-2019), among others. These standards have played an important guiding role in the field of data security in China.

The series of "Trusted Data Services" specifications, which was introduced by TC601 (the Big Data Standards Technology Promotion Committee of the China Communications Standards Association) has extended the protection of personal information to comprehensive compliance of corporate data. According to the different roles and identities of data providers and data distribution platforms, these specifications have provided recommendations for corporate data compliance in terms of management processes and content. They have also enumerated management requirements and recommendations on platform management, management of circulation participants, circulated products management, and circulation process management when the data circulation platform provides data circulation services, as well as the requirements on service capabilities and service quality in such fields as data product management and data product supply management. The series of specifications was released in June 2019.

VI. Big Data Development Outlook

The Fourth Plenary Session of the 19th CPC Central Committee proposed that data, just like other production factors such as labor, capital, land, knowledge, technology, and management expertise, could participate in the distribution. This shows that data is becoming more and more important in the operation of the economy, and data is having a fundamental, comprehensive and revolutionary impact on economic development, social life and national governance.

On the technology side, we are still in the early stages of the "big data explosion". The further development of 5G and the Industrial Internet will create a "deluge of data", which will bring greater challenges to the storage, analysis, and management of big data and drive big data technologies up to a new level. The integration of hardware and software, and the convergence of data and intelligence will drive big data technologies to expand towards the direction of heterogeneous multi-mode, ultra-large capacity, and ultra-low latency.

In terms of applications, big data industry applications are extending from the consumer end to the production side, and from the descriptive and diagnostic applications to the predictive and prescriptive applications. At present, the Internet industry has entered the "DT era". In the next few years, with the completion of local government big data platforms and large enterprise data middle platforms (Zhong Tai), the application of big data in the fields of government affairs, people's livelihood and the real economy will be uplifted to a new level.

From the perspective of governance, with the continuous improvement of the national legal system for data security, data governance in various industries will also advance further. The chaos in data collection, use, and sharing will be curbed, the security management of data will become the bottom line of compliance by all trades and industries. The compliance of data circulation and applications will be greatly improved, and a healthy and sustainable big data development environment will gradually take shape.

However, China's big data development also faces many challenges.

For example, the home-grown development of big data technologies and products is not strong; the level of data openness and sharing is still low, the data flow across departments and industries is still impeded, and valuable public information resources and commercial data are not fully circulated; data security management is still weak, and personal information protection faces new threats and new risks. This requires big data practitioners to put more effort into the research of big data theory, technology R&D, industry applications, security protection and other aspects.

New era, new opportunities. We have also seen that the development of big data is much more closely integrated with the next-generation information technologies such as 5G, AI, and blockchain. What's worth mentioning is the blockchain technology. On the one hand, blockchain can, to a certain extent, solve the "congenital diseases" of the big data industry such as difficulty in determining data rights, serious data silos, and data monopoly; on the other hand, big data technologies such as privacy computing technology can in turn promote the improvement of blockchain technology. With the support of the new generation of information technologies, China's digital economy is developing in the direction of better mutual trust, better sharing, and better balance, and the "production relationship" of data is being further reshaped.

2020 is approaching. The "Thirteenth Five-Year Plan" will come to a successful conclusion and the "14th Five-Year Plan" is waving hands towards us. We look forward to standing at a new historical starting point. The new generation of information technologies represented by big data will make greater contributions to building China into a strong manufacturing power, a strong cyber power and a strong digital power.

China Academy of Information and Communications (CAICT)

Address: No. 52 Huayuan North Road, Haidian District, Beijing

Postal Code: 100191

Phone: 13683007576

Fax: 010-62304980

Website: www.caict.ac.cn