Measures to regulate data element circulation

MANAGEMENT

By TANG KE
and XIONG QIAOQIN

It is important to build a multi-layered, traceable, safe, and effective data element market that has good regulation and standardized circulation. This will help maximize the benefits and efficiency of data allocation according to the rules, prices, and competition of the market. It will also help protect data privacy and data sovereignty.

The exhibition hall of Shanghai Data Exchange Corporation at the 2021 World Artificial Intelligence Conference in July, 2021 Photo: SDE

Data exchange in China

A thriving data element market calls for a platform that enables data circulation. Recent years have witnessed China setting up a few data trading centers and trying out the business. For instance, China's first data trading platform—the Global Big Data Exchange in Guiyang, Guizhou Province; Shanghai Data Exchange, whose core technologies include de-identified metadata regulation, independent listing control, and ID identification matching; and Qiantang Big Data that promotes the organic integration of the manufacturing industry and big data.

In general, China has only started exploring this industry on a small scale, while relevant rules and regulations are still lacking. Although there are over 20 local-government-authorized data trading facilities in China, more than half of them are inactive or even failed. Three out of more than ten private enterprises labeled "data trading" have been cancelled.

To make matters worse, illegal data trading and data leaking are common. The Investigation Report on the Protection of Chinese Netizens' Rights and Interests 2020 released by the Internet Society of China indicates that netizens in China lost 80 billion yuan in merely one year due to spam, information fraud, and personal information leaks. The per capita economic cost per person was 124 yuan in one year.

The reasons behind these phenomena lie in the features of data, which differ from other production factors due to data's replicability and nonexcludability. While a large number of data users benefit from replicating and reselling data, it is not easy to gather evidence of illegal access or trading.

It remains costly for stakeholders to reduce non-consensual replications and data leaks. For one thing, data stakeholders are more inclined towards blocking the access to data in a bid to avoid data from being resold or overused, which would lead to a shrinking market for legal data trading. In addition, driven by self-interest, people in need of data may seek illegal access to data and overuse it.

Since data is easy to replace yet hard to trace, online data trading platforms may resort to data intersection, tampering, or resale. This further exacerbates behaviors that harm customers' interests, such as using big-data analysis to set varied prices for the same product or service based on different consumers' consumption habits.

What's more, China's current data trading platforms are mostly de-centralized ones that function as intermediary agents or processing centers for data analysis. These platforms still lack techniques in tracing data exchange. Meanwhile, most of them enter the market based on membership, while leaving their massive data resources inactivated.

Room for improvement

Efforts should be made to better regulate data element circulation and transaction, cultivate China's data element market, eliminate the black market of data transaction, and unleash the potential of the digital economy.

First, a multi-level market trading system should be built based on data diversity. Data is not a standard commodity and takes many forms. Any data can be combined, split, or adjusted to be new data. Therefore, data circulation should be reflected in a multi-level market trading system with different trading rules applied according to specific types and scenarios. The markets can include cross-border, national, or regional data exchanges and trading platforms, among others. The trading subject can be classified into different categories such as raw data, de-sensitized data, data reports, calculation results, query services, customized services, and financial data products. Transactions can take various forms such as purchase, lease, auction, mortgage, etc.

In addition, data can be divided into three types: fully open, partially open, and never open according to different levels of privacy, security, and sensitivity. Among them, partially open data has different levels of de-sensitization, usage restrictions, and processing requirements considering the nature of data, data usage, regulatory difficulties, and privacy costs of linked data.

Second, blockchain technology should be applied with the emphasis on the traceability of data transactions. The replicability of data makes it no longer exclusive to the seller once it has been sold. Any dispute that arises between the parties would expose the seller to great risks. To avoid these pitfalls, technologies and systems are needed to record and trace the prior negotiation and validation of data transactions, the flow of transactions, and the reach of data. Traceability is therefore at the heart of data transactions, and the blockchain is the basic technology for achieving it. To start with, the blockchain naturally has a function of confirming the rights. The timestamp technology and consensus mechanism in the blockchain architecture can prevent transactions from being tampered with. In the event of a dispute, the data owner can prove legal ownership through the timestamp of the blockchain transaction, and can trace backwards to detect non-consensual or illegal behavior.

The blockchain can also help buyers to determine whether the data set meets the requirements before the transaction and leave behind a trail of evidence. All records of data transactions including data samples (or hash values) are fixed on the blockchain, preventing illegal data leakage and avoiding the possibility of cheating by participants in the transaction process.

In addition, the pledge and dispute resolution mechanisms for data transactions are embedded in the architecture of blockchain, which not only enables automatic transfer of funds for transactions using smart contracts, but also allows for online arbitration and even automatic judgment based on evidence on the chain.

Third, it is necessary to ensure data security by developing privacy protection technologies. Data transaction is prone to privacy breaches, and data security concerns are exacerbated by interception, tampering, and reselling practices on trading platforms. This calls for a variety of methods to strengthen privacy protection in the process of data generation, processing, use, and circulation.

Therefore, standards should be developed for the protection of data according to data content and application scenarios, in order to achieve targeted, graded, and differentiated protection of privacy and security. Meanwhile, cryptography-based privacy-protection technologies featuring multi-party secure computing, federated learning, privacy computing, and trusted hardware should be developed to strike a balance between data security and performance and to ensure that data is not leaked while facilitating legal access and utilization of data. The derivation of data protection obligations in data transactions should also be advocated. The scope of use and prohibition with regard to data can be specified in the transaction contract to prompt the data seller to fulfill his/her obligation to monitor the buyer.

Fourth, data sovereignty should be safeguarded, with cross-border data flows being treated prudently. The cross-border flow of data is conducive to promoting enterprise innovation and industrial upgrading. Access to and use of data also affects a country's economic prospects and international political status. The formulation of international rules for cross-border data flow involves geopolitics and tussles about national data sovereignty. Thus, participating in or leading the rule-making process and the regulation of cross-border data flows will help China to guarantee data sovereignty and carry out global governance.

At the same time, how to coordinate countries in terms of different data protection requirements is also relevant to the quality of cross-border data flows. Otherwise, it may pose a threat to users' rights and data security, and may also lead to international disputes due to differences in audit systems or technical differences. In addition, cross-border data flows pose greater challenges to data security in terms of prevention, forensics, enforcement, and remedies. Therefore, national data sovereignty should also be firmly upheld to ensure the orderly and prudent flow of cross-border data.

Fifth, the system for data regulation and governance should be strengthened based on the principle of contextual justice. Data regulation and governance is an integral part of data transactions and sharing, and is an effective guarantee for the rights and interests of all data subjects.

Thus, based on the principle of contextual justice, data transaction contracts should specify the scope of data use or prohibition. Meanwhile, data and relevant entities should be governed in a hierarchical and focused manner based on different types, objects, uses, and risk levels. Rules and solutions should be developed for specific contexts following risk assessment.

The use of artificial intelligence, the blockchain, smart contracts, and other technical algorithms boasts real-time reach, immutability, hierarchical authority control, and automatic execution. They should be leveraged to improve the efficiency of data regulation. What's more, the legal systems should be improved for stronger in-process and post oversight and better arbitration for the data element market, in order to build a trustworthy, traceable, safe, and orderly market with sound supervisory and supporting systems.

Tang Ke and Xiong Qiaoqin are from the School of Social Sciences at Tsinghua University.