ISSN 2706-6827 Vol. 5, Issue 4: 125-134, DOI: 10.25236/IJFS.2023.050418

Discussing Data Ownership Disputes and Data Protection from Data Crawling Practices

Zhaohui Yang

China Jiliang University, Hangzhou, China

Abstract: *As a new kind of oil, data is an important resource that enterprises compete for in the digital economy, and mastering data means mastering the market. However, the game of endless crawlers and anti-crawler technologies reflects the conflicting relationship between data crawling and resource protection. In China's judicial practice, the general provisions of Article 2 of the Anti-Unfair Competition Law are usually applied to the regulation of data crawling. The application of this provision should find a balance between the protection of data resources and the restriction of data monopoly, so as to resolve the conflict between data grabbers and data collectors, data sharing and the protection of data rights and interests; it also needs to be discussed in terms of the competitive relationship, the legitimacy of the act and whether the principle of necessary facilities can be defended. As the cornerstone of the legitimacy of data capture, we should discuss how to protect data and maximize its value by building a system of data rights attribution, so as to alleviate the inherent tension between incentivizing data production and reducing the risk of individual privacy infringement, form a reasonable division between individual users, platform enterprises and the government and the state regarding the content and boundaries of data ownership, and build a "common ground" between the public, online platforms and the government and the state in data governance. This will lead to the establishment of a "co-construction, co-management and sharing" pattern of government national data governance.*

Keywords: *anti-unfair competition law; crawlers; data corroboration; legitimacy*

1. Introduction

According to a report released by the OECD in 2019, Enhancing Access to and Sharing of Data helps maximize the social and economic value of data reuse, can enhance the value of data to data holders and secondary data users, and bring additional positive spillover benefits to the national economy and society as a whole. The EU believes that data flow contributes to data-driven growth and innovation, and has explicitly made "free circulation of personal data" a legislative goal in both the General Data Protection Regulation and Regulation on the Free Flow of Non-Personal Data. Article 7 of the newly promulgated Data Security Law of the People's Republic of China, promulgated in June 2021, also stipulates that "the orderly and free flow of data in accordance with the law shall be guaranteed".

However, as a neutral technology, data scraping technology also stands behind the value orientation of resource sharing, fair competition and information protection. Excessively restricting data scraping can lead to oligopolies. In the information age, the sharing of data resources is undoubtedly very important, which is more likely to affect technological updates, which is not conducive to maximizing the convenient use of the Internet.

At present, in China, there are no specific legal provisions or rules for data capture, whether it is the Anti-Monopoly Law or the Anti-Unfair Competition Law, and judges usually make their decisions based on the general provisions of their discretion. The court's regulation of data scraping is often based on Article 2 of the Anti-Unfair Competition Law, which determines the legitimacy of the act by examining whether the act violates business ethics, but this model has certain ambiguity. Too much reliance on business ethics risks limiting the development of more economically efficient behaviors. Therefore, in the process of applying the rules, it is necessary to clarify the factors for determining the illegality of data capture.

Therefore, in the process of applying the rules, it is necessary to clarify the factors for determining the illegality of data capture. For the application of this clause, a balance should be found between protecting data resources and restricting data monopoly, so as to resolve the conflicts between data scrapers and data collectors, data sharing and data rights protection; It is also necessary to discuss the controversial points such as competition relationship, legitimacy of conduct, and whether the principle of necessary facilities can be defended, so as to clarify the illegal boundaries and legitimacy of the use of data scraping from the perspective of the Anti-Unfair Competition Law. The purpose of data scraping is to obtain greater competitive advantage and improve the quality or type of products or services, so competitive relationship and behavioral legitimacy are two core elements, in addition, the scraper should also be given a certain opportunity to defend.

2. What are crawlers and anti-crawlers

2.1. Basic technical principles of crawlers

Web crawlers are also known as web spiders and web robots. It is a program or script that automatically crawls the World Wide Web for information according to certain rules. [1]In other words, a piece of code can be written to collect specific information on a particular web page, thus collecting a large amount of information on that web page quickly and without the cost of human retrieval. [2]For example, various search engines actually borrow the technology of web crawlers to create web pages related to search terms, thus helping users to quickly obtain information related to the search term.[3]

Specifically, each piece of information on the World Wide Web has a uniform and unique address URL (Uniform Resource Locator), i.e. a web address. First, a queue of URLs to be crawled is created, starting with one or more URLs of the initial web page, each URL is continuously extracted in order, its corresponding web page is accessed and parsed, and all URLs in the crawled read page are then stored in the queue to be crawled for cyclic crawling until all URLs in the queue are crawled or when certain stopping conditions of the system are met. "

2.2. Anti-crawlers - how to deal with crawling techniques

While crawler technology brings convenience to users, there are also consequences that can cause information leakage on the target website and server crashes due to the large number of visits to the website. Therefore, many websites have established appropriate defensive measures to deal with crawler technology, common types include robots protocols and anti-crawler mechanisms.

The first is that the Robots Agreement, also known as the Crawler Agreement, Crawler Rules, and Robots Agreement, is a code of ethics that is common to the international Internet community and is designed to protect website data and sensitive information and ensure that users' personal information and privacy is not violated. The "rules" define the scope of the search engine's crawl, including whether the site is expected to be crawled and what content is not allowed to be crawled, and the web crawler can automatically crawl or not crawl the content accordingly. If you think of a website as a room in a hotel, robots.txt is the "Do Not Disturb" or "Welcome to Cleaning" sign that the owner hangs at the door of the room. This file tells visiting search engines which rooms they can enter and visit, and which are closed to them. But the robots protocol is a convention rather than a norm, and not all websites follow it.

The second is the anti-crawler mechanism, which includes a range of means to circumvent crawler technology. This includes a range of technical means to restrict access, or allow access only to real registered users, by identifying crawlers through UA, setting the frequency of IP access, identifying crawlers through concurrency, filtering statistics on the time window of requests, identifying legitimate crawlers, etc.

However, restricting crawler technology only by means of private remedies is often not enough. On the one hand, excessive restrictions on crawler technology may result in a monopoly of resource information, which is not conducive to the long-term development of information networks; on the other hand, relying solely on websites for private remedies may not be conducive to the protection of big data, which may lead to damage to the interests of individuals, society and the state.

3. Consideration of data scraping illegality factors - in the light of anti-unfair competition law

The law is the last line of defence for social justice, and private remedies such as sophisticated firewall designs or robots protocols do not seem to be effective in avoiding data crawlers alone. [4]Some web crawlers directly infringe on the rights and interests of others, some web crawlers are directly suspected of committing crimes, and more web crawlers are in the grey area of the law. "Anti-crawling" has become a "never-ending battle" in the world of the Internet, the purpose of which is to obtain data. It is at this point that the boundaries of rights are set by legal regulation. Data crawling first entered the legal landscape in the field of competition law, as in the 1999 case of eBay v. Bidder's Edge (eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1060-63 (N.D. Cal. 2000)). Cal2000)), among others.

3.1. Whether or not it constitutes a competitive relationship

Firstly, the identification criteria of the operator. With regard to the identification of the operator in the act of data crawling, on the one hand, the traditional view of the theoretical community adopts the "subject qualification theory", which holds that the operator under the scope of the anti-unfair competition law refers to the market competition with the same business relationship, and obtains the business qualification of the market subject in accordance with the law; another view is The other point of view is the "behavior standard theory", which holds that whether or not the operator has the behavior of market operation activities as the basis for judging the identity of the operator. [5]According to Article 2 of China's Anti-Unfair Competition Law, "Operators referred to in this Law are natural persons, legal persons and unincorporated organizations engaged in the production or operation of goods or the provision of services (hereinafter referred to as goods including services)." And according to the Interpretation of the Supreme People's Court on Several Issues Concerning the Application of the Anti-Unfair Competition Law of the People's Republic of China (Draft for Public Comments), market entities that have a possible relationship with operators in production and operation activities such as competing for trading opportunities and harming competitive advantages, the People's Court may determine that they are "other operators" as stipulated in Article 2 of the Anti-Unfair Competition Law ". It is easy to see from the above that the identification of business subjects in China lies in the way they behave rather than in their qualifications. It would be detrimental to the protection of market subjects to limit the identity of operators to the acquisition of qualifications only. Therefore, the determination of whether a subject is an operator should be based on whether the subject has committed acts that may compete for trading opportunities, harm the competitive advantage and other relationships.

The second is the determination of the competitive relationship. According to Article 2 of China's Anti-Unfair Competition Law, "The act of unfair competition referred to in this Law refers to the act of an operator who, in the course of production and business activities, violates the provisions of this Law, disrupts the order of market competition and harms the legitimate rights and interests of other operators or consumers." Traditionally, competitive behaviour may be limited to the same or similar goods and industries, but in the internet, the use of big data is gradually breaking down the barriers between industries, for example: the analysis of user information can be used to analyse their online shopping preferences as well as to determine their consumption levels and thus recommend offline consumption places at different price levels. At this point, although the data collector and the data crawler are in different industries, the user information they obtain is the same, and the goal is to compete for the number of users and the completeness of the information, so the competitive relationship in big data crawling should be understood in a broad sense: any relationship that may compete for trading opportunities or compromise competitive advantage should be considered as competitive.

3.2. Whether the conduct was justified

The question of whether the act of data capture is justified is central to defining the boundaries of the act. According to the Anti-Unfair Competition Law, "In their production and operation activities, operators shall follow the principles of voluntariness, equality, fairness and honesty, and abide by the law and business ethics." It can be seen that voluntariness, fairness, equality and good faith are the basis of whether the act of grasping is justified. The discussion of justification can be divided into the following aspects: whether the data is public and whether the data is processed.

First, whether the data is public. First of all, non-public information such as personal privacy, commercial secrets and national security need not be mentioned, and the crawling of such information must not be justified. However, there has been controversy over whether the crawling of public data is justified, such as the public data of some design user information, although it is authorised by the user for public display, but will the crawling of such information to other platforms violate the user's right to know and privacy? For public data, the author believes that two aspects should be considered: for public data designed for user information, it should be authorized by the user and the platform collecting the data, otherwise such information crawling is not justified; conversely, for public data crawling that does not involve user information, it should be considered justified.

Secondly, whether the data has been processed. By classifying the source of the data, it can be divided into raw data and processed data. Raw data is data that has not been processed or handled. Processed data, on the other hand, is the opposite, so data that has been deeply mined, processed or handled is processed data. Unlike raw data, processed data often has a higher commercial value and involves more cost and effort. In judicial practice, processed data is usually considered to be the "fruits of labour" and has the property of intangible property, so the act of capturing such data can be considered as "using the fruits of labour of others without permission". Therefore, the act of processing data is not legitimate.

3.3. Defences based on the essential facilities doctrine

The essential facilities doctrine first began in the 1912 US Supreme Court case on the railway terminal decision. Under this provision, if a dominant firm in an upstream market controls an indispensable and irreducible essential facility (including infrastructure, technology and natural conditions) for downstream production, it is obliged to allow downstream manufacturers appropriate commercial terms to use that facility in order to avoid anti-competitive consequences. [6]This principle creates a mandatory obligation for operators to ensure fair competition. Can this principle be applied to the Internet? Can data be an indispensable and irreducible necessity?

In November 2020, the State Administration for Market Supervision and Administration issued the Anti-monopoly Guidelines on the Platform Economy (Draft for Comments) (hereinafter referred to as the Draft for Comments), which provides in Article 14 that the analysis of whether a refusal to deal constitutes a refusal to deal may take into account Article 14 provides that the analysis of whether data constitutes an essential facility may take into account the refusal to deal on reasonable terms with the counterparty to the transaction by the operator who "controls the essential facilities in the platform economy" and sets out the rules for determining whether the data constitutes an essential facility: "Generally, it is necessary to take into account whether the data is indispensable for participation in market competition, whether there are other access channels, the technical feasibility of opening up the data, and the possible impact of opening up the data on the operator in possession of the data, among other factors." The Exposure Draft first addressed important issues in data monopolies, but in the Anti-Monopoly Committee of the State Council's Anti-Monopoly Guidelines on the Platform Economy, issued in February 2021, the paragraph in the Exposure Draft determining whether data constitutes an essential facility was deleted, while "the possession of data by the platform" was included in the The Commission's Guidelines on Antitrust in the Platform Economy remove the paragraph on whether data constitutes an essential facility from the Exposure Draft and include "the possession of data by the platform" in the determination of whether the platform constitutes an essential facility. [5]It is clear from this that we also have reservations about whether data is an essential facility. This does not mean, however, that there is no value in discussing whether data is an essential facility.

In order to avoid the monopolisation of the data market by large companies, which is not conducive to fair competition for later entrants, the principle of necessity can be applied in a case-specific manner, so as to ensure that the act of data capture is given a defence from the perspective of fair competition and the maintenance of a healthy market development. It also imposes an obligation on the head companies that have a large number of users to disclose a certain amount of data in order to avoid a monopoly situation.

Data scraping is not naturally illegal, it is simply a tool in the complex information age and the key is how to use it wisely. The key to the conflict between data capture is not only between the data collector and the data captor but also between the two and the public interest. In order to resolve the conflict of data capture, it is necessary to strike a balance between data protection and data sharing, and to use legal means to limit the unlimited expansion of rights on the one hand, and to protect the rights from infringement on the other, so as to achieve win-win cooperation among multiple parties.

4. Data Ownership Disputes and Opinions

4.1. What rights do the data belong to?

The importance of data and the fact that current domestic legislation has few, if any, specific rules on data ownership have led to a plethora of controversial issues regarding data in reality in recent years, with judicial practice facing huge challenges. The core of these cases, such as Sina v. Pulse and Popular Dianping v. Baidu, is the boundary of the legality and reasonableness of one platform's access to another platform's data through technical means. However, in the specific judicial proceedings, the courts did not directly and substantively recognise the ownership of the data, but sought relief in an indirect manner by flexibly applying the expanded interpretation of the underwriting provisions of China's anti-unfair competition law in individual cases. The plaintiff's interest in the data is a purely economic interest protected by law (as an intellectual work product), and does not enjoy a new type of enterprise data property right independent of personality rights, property rights, claims and intellectual property rights.

Firstly, whether or not to identify rights to data. There are various voices in the academic community as to what rights data belongs to, ranging from intellectual property rights, to property rights, or to some new type of right. However, it is clear that data does not meet the requirements of either intellectual property or property rights.

On the one hand, data cannot be classified as either a copyright, a patent or a trade secret in terms of intellectual property rights. In terms of patent law, data does not have the "triple" requirements of novelty, inventiveness and usefulness, but is more of an integration of a large amount of user information; in terms of copyright law, data does not have the originality or originality that it requires; in terms of trade secrets, data is circulating and should be allowed to circulate and be shared in order to maximise its value. From the perspective of trade secrets, data is circulating and should be circulated and shared in order to maximize its value, which is clearly contrary to the "not known to the public" of trade secrets, and "technical information, business information and other commercial information that the right holder has taken corresponding confidentiality measures". On the other hand, if the data were to be protected in rem. Data is clearly not a property, as it is a binary code that exists on computers and networks, with low reproduction and circulation costs, non-exclusivity or sharing, intangibility, and hardly has the disposability and exclusivity of a property. [7]For example, in the case of NFT digital collections, the purchaser acquires the credentials of the digital collection (consisting of a unique string of codes) rather than the ownership of the physical object of the collection.

The difficulty with the attribution of data rights is mainly due to the fact that data applications are variable and may show different properties in different scenarios, making it difficult to attribute some unique right.[8]

4.2. To whom should the configuration of data rights be attributed?

After discussing what data rights belong to, it is inevitable to discuss to whom the data rights should be attributed. The question of how to allocate data rights between users and operators, and operators and state authorities, covers the protection of personal information, fair competition between platforms and the protection of consumer rights. At present, there are four main theories in academic circles, with four models of ownership: user-owned, platform-owned, user-platform shared, and state-owned.

The first is the doctrine of user ownership. This doctrine favours the "absolute protection" of data - as personal information about the user - and is a strict protection model of affirmative, independent legislation. The doctrine holds that the ownership of individual user data should be vested in the user. This is because of the personal and property nature of data, the fact that data originates from the user - the individual - and the fact that the right to data extends to personal information, which, as a right to privacy or personality, should be protected effectively and efficiently, which also facilitates users to feed more personal information back to the platform.

This model of protection can adequately protect the human dignity and freedom of the individual user and the user's right to control data. [9]However, if this right is extended indefinitely, it is likely to be detrimental to the collection, processing and use of personal data by the platform, which may be contrary to all the circulation of data and its reproducibility in large quantities, and the need for the user's knowledge and consent to every use of data, which may increase the costs for the company and be detrimental to the development of the industry and the globalisation of the Internet.[10]

A worthy subject for this model is the European Union, with its 2016 General Data Protection Regulation (GDPR), the strictest ever, which empowers subjects of personal information with a range of rights such as access, query, deletion, withdrawal, data portability and other rights, and there are strict restrictions on the flow of data across borders.[11]

Second, the platform all said. Platform as the collection and integration of user information, for the development of the platform injected a lot of human and material costs, a single user information is not much value, it is because the platform will be a huge amount of user data integration and multi-platform data interaction, the user group classification, labeling, to make full use of the value of data. And some of China's judicial decisions, such as the Guomi v. Yuanguang case in which the court held that "the platform has property rights and interests in the data products and services formed by the fruits of its intellectual work," but the attribution of rights to the platform, while ignoring the protection of users' personal information, is obviously not conducive to the protection of users' data rights and interests, and has a This may lead to the plundering of user information between platforms, which will ultimately harm the overall welfare of society.[12]

Third, the user and the platform share said. Through the user authorization, the platform processing and use mode to share the ownership of data, the platform based on the user's authorization, the data integration and processing, to achieve the transformation of data to information, prior data asset. [13]This is a further refinement of the user-ownership doctrine and corrects the disadvantages of the single-ownership doctrine for either the user or the platform. However, the main problem facing this doctrine is whether the user will agree to authorisation and all data must be used based on the user's authorisation, which greatly hinders the flow of data and increases the cost of using the data for the platform; or, once the platform has acquired the user's data, prevents the user from authorising it for other platforms, which would be detrimental to fair competition.

Fourthly, there is the argument of state ownership. Data has great commercial value, it is not only the integration of users' personal information, it can also be seen as a collection of public will, if the data is completely placed in the market choice, it may lead to oligopoly or low usage. [14]The "state ownership" model emphasises the state's obligation to protect data-related rights and interests in order to adapt to the changing economic model, to regulate the balance between personal information protection and platform use, and to improve the efficiency of data use and maximise the public interest. This model emphasises that it is difficult to provide comprehensive protection for personal information under private law, and that public authorities should be used to reasonably allocate data rights. This is contrary to the idea of market-based data circulation and marketisation.

5. The essence of dispute resolution

5.1. Balancing the use and limitations of data capture techniques

Data crawling is not naturally illegal, rather excessive restrictions on data crawling may be detrimental to resource sharing and cause data monopoly. There are three subjects in the act of data crawling, namely the data holder, the data crawler and the public. In order to clarify the conflict between data crawling protection and restriction, the one-to-many relationship of the subjects should be stripped out and analysed one by one.

5.1.1. The conflict between fair competition and data protection

China currently evaluates the legality of information scraping from the perspective of anti-unfair competition law, due to the fact that late entrants to the relevant market are often limited by the difficulty of accessing data and find it difficult to develop and innovate. The acquisition and use of data, in turn, affects the success or failure of business operations to a certain extent, making data scraping a widely used technology. In many data scraping unfair competition cases, the root cause of the dispute is the conflict between data protection and competition rights.

Firstly, the protection of data is justified. The data collector often has to invest a lot of cost and effort to develop and collect the corresponding information, and needs sufficient customer base to provide information, and each piece of data has a certain value; while the data collector is equivalent to standing on the shoulders of giants, using a small cost to access the data collector's high-value data, which will obviously infringe on the rights of the data This clearly infringes on the rights of the data collector and thus creates a conflict of laws.

Secondly, excessive protection of data may not be conducive to fair competition, thus creating a monopoly. This is because the subjects that enter the market first usually have a large and active user base, and excessive protection of the data of these subjects may result in the later entrants being unable to compete with them, creating a winner-take-all situation.

5.1.2. The conflict between resource sharing and data protection

Data is the new oil and the free flow of data embedded in data crawling is of great value in the age of the digital economy. The importance of open data is already widely recognised worldwide.

In a report published by the OECD in 2019, it is stated that Enhancing Access to and Sharing of Data (EASD) helps to maximise the social and economic value of data reuse and can increase the value of data to data holders as well as secondary data users, bringing national economies and society as a whole additional positive spillover benefits to the national economy and society as a whole. Depending on the scope of the data and the degree of openness of the data, data access and sharing can generate social and economic benefits roughly equivalent to 0.1% to 1.5% of gross domestic product (GDP) for public sector data, and 1% to 2.5% of GDP when also including data from private subjects (up to 4% of GDP in some studies).[15]The EU believes that the flow of data contributes to data-driven growth and innovation, and in the General Data Protection Regulation (GDPR) and the Regulation on the Free Flow of Non-Personal Data Both the General Data Protection Regulation (GDPR) and the Regulation on the Free Flow of Non-Personal Data explicitly include the "free flow of personal data" as a legislative objective.[10] Article 7 of the latest Data Security Law of the People's Republic of China, enacted in June 2021, also stipulates that "the free flow of data shall be guaranteed in an orderly manner in accordance with the law."[5]

In the information age, the sharing of data resources is undoubtedly very important. Excessive restrictions on data disclosure will not only create an oligopoly, but may also affect technological updates, which is not conducive to maximising the convenience of the Internet. The conflict between resource sharing and information protection is more important than the conflict between data grabbers and data collectors, which is about balancing information disclosure, information transparency with commercial confidentiality and personal privacy. A large number of applications on the market require personal information (including but not limited to phone numbers, ID information, and geographic location), and furthermore, information such as fingerprints and facial recognition can be obtained through user authorisation. The range of information is constantly expanding and refined, but the extent to which users use their information is unknown. For example, Apple's security was questioned when the Apple cloud was hacked from Apple phones and the private photos of many celebrities were leaked, thus affecting the commercial value of Apple.

However, the conflict between data sharing and information protection is not irreconcilable and can only be harmoniously coexisted if the relationship between the two is balanced and protected by legal compulsion.

5.1.3. Promoting a win-win situation for all

The essence of conflict resolution is to maximise the use of data resources and maximise their value, but at the same time to control their boundaries, as the unrestricted spread of rights will inevitably lead to the infringement of the rights of others. The rights of all parties should be restricted accordingly, so that a win-win situation can be achieved. The interests of society as a whole are public, social and long-term, and are not simply the superimposition of the interests of individual subjects, nor is the maximisation of the interests of an individual or a party the maximisation of the interests of society as a whole. The protection of the rights of innovation and competition in data protection is a fundamental public policy that exists and is recognised worldwide, and it is a difficult task to see and deal with the relationship between the two and to construct a balance.

Our courts usually do not discuss what rights the data belongs to in relation to data capture, but usually find that the integration of such data is the "fruit of the labour" of the data collector, thus avoiding a discussion of the complex issue of data attributes, while the unauthorised use of the "fruit of another's labour The unauthorised use of the "fruits of another's labour" may constitute unfair competition or an infringement of intellectual property rights. The act of capturing information is likely to infringe both IPR and fair competition, but in copyright litigation, the plaintiff has a relatively heavy burden of proof and needs to obtain the user's authorisation one by one, with detailed claims specific to each article, before proceeding with the litigation; whereas in unfair competition litigation, where data exploitation is the core of the dispute, whether or not the user's authorisation has been obtained is not an important issue, and the court The court placed more emphasis on the overall business model of the enterprise and recognised that the processing of user data by the enterprise was a fruit of labour, allowing the plaintiff to initiate litigation directly against the defendant in relation to its business model without having to prove the user authorisation obtained one by one, achieving a broader strike effect and reducing the burden of proof on the plaintiff.

Data protection and competition protection are not only in conflict, but also share a common value - to promote technological innovation and enhance consumer welfare (Consumer Surplus). In order to reconcile the conflict between the two, we should start from this breakthrough point, taking into account the costs of data holders and the competitive rights of data grabbers, increase the cooperation between enterprises (B2B), increase the sharing of resources between enterprises, so as to refine the way and scope of data use, and further clarify through the "consensual + statutory" approach The illegal side of data capture.[5]

In judicial practice, through the case of Sina v. Pulse taking the data crawling act of unfair competition, the court established that "in the Open API development cooperation model, the third party should adhere to the triple authorization principle of 'user authorization' + 'platform authorization' + 'user authorization' when acquiring user information through the Open API. platform authorization' + 'user authorization' triple authorization principle", that is, not only the data of the crawled party needs user authorization to be stored, but also the data crawling of the third party has to obtain dual authorization from the crawled room (data collection party) and the user to promote data protection and Competitive rights work in tandem.

5.2. Reflections on the boundaries of data tenure empowerment

5.2.1. Right to personal data

Personal data, a type of data that is associated with the attributes of the user's characteristics as well as the processing through the internet incorporating the user's own attributes. The extent and scope to which such data can be collected by platforms, whether users enjoy the right to know and whether they can refuse to have their data collected and processed are then the scope of discussion. Specifically, individual users enjoy two specific rights and interests: on the one hand, the rights and interests of data personality, including but not limited to: the right to informed consent, the right to data confidentiality, the right to data correction and the right to data deletion; on the other hand, the rights and interests of data property: that is, natural persons, legal persons and unincorporated organisations enjoy the possession, use, benefit and disposal of the data data products and services resulting from their lawful processing of data, and can legally occupy themselves Use, gain and dispose of. [16]In judicial practice, China's law adopts a decentralised legislative model for personal data protection, for which a preliminary structure should be established for personal data, specifically divided into personality and property rights and interests, to prevent users' personal privacy from being infringed, but also to ensure the maximum use of data resources.[7]

The first type of data is directly related to the user's personal characteristics, i.e. the data collection and reading of biological characteristics such as the body, behaviour and physiological structure of a natural person, and then the relevant personalised classification and labelling of the user. These include biometric information such as iris, fingerprints, facial recognition features, identity-based information such as name, date of birth, ID card number, etc., and personal derivative information such as home address, telephone number, spouse status, etc. These are all individual user information and are subject to expansion or deletion according to changing times. Such personal data, regardless of who holds it, belongs to the individual, and the platform strictly abides by the "inform and consent" rule when collecting such data.

The second category is the data involving personal privacy of the respective text, image and audio content created and published by individual users in the online platform.[7]Such as call records, browsing records, search keywords, etc. For this kind of data, although its origin is personal, but is generated with the help of the platform, generally can be contracted by the individual when the platform, the two sides to agree, the individual has the ownership, the platform enjoys the right to use the rules, but the platform shall not share this kind of data privately to a third party, the platform use process also shall not infringe the rights and interests of the user. This includes, among other things, data relating to personal privacy that is self-published and made public by the user, which may be regarded as the user's implied consent that the data may be used by third parties without compensation, provided that the use is not detrimental to the interests of the right holder and that the principle of safe harbour and red flags are strictly observed.

The third category is data generated by individuals in their interaction with various online platforms that is not about personal privacy, such as visits to web sites, as well as likes and comments on relevant users' movements, on the basis of which the platform can recommend other content of the same kind that may be of interest to the user, which can be used directly without seeking the user's consent, provided that it does not infringe the rights of the right holder.

5.2.2. Platform Data Ownership

Data is public and its greatest value comes from the fact that it can be shared, reproduced indefinitely and circulated quickly, so only data that is in constant circulation has core value. Therefore, data information should not be absolutely controlled, but rather protected while allowing the platform to use and process it to a certain extent. Therefore, the effective circulation of data should be promoted on the basis of protecting the rights and interests of individual users.[17]

Specifically, the first category is the various types of data generated by the platform's own business management operations. Such as tax registration, personnel management and other data, which are made public on social media based on various needs, such as company annual reports, publicity, etc. Specifically, it can be further divided into: 1. Company commercial secrets. Such as product planning, system specifications, corporate data, customer information, etc. Such information is owned by the platform and cannot be crawled and used by third parties without authorisation; 2. Data that is compulsorily required by law and regulation to be made public by the platform. Such as tax registration, financial reports, etc. Such data rights belong to the platform, but third parties can use them without authorization, but they must not infringe on the interests of the right holder; 3. It is data that the platform makes public to the public. Such data may be regarded as implied consent and may be used by third parties, provided that it does not infringe upon the rights of the right holder.

The second category is data generated when the platform interacts with other platforms or customers to provide products and services or with government management, such as service information and project requirements, etc. Although this type of data is partly controlled by the platform, its ownership does not necessarily belong to the platform, and the corresponding rights need to be specified according to specific scenarios.

5.2.3. Government Data Ownership

Governmental state data rights are based on the intended autonomy of the user or platform, while part of the rights to use, process and benefit from data are ceded in order to facilitate the maximisation of public welfare.

This kind of data can be divided into two categories: one is the data on education, medical care, social security, transportation, justice, weather, etc., which are collected and processed by government departments on their own initiative in order to perform public management and service functions such as policy making, urban planning, economic regulation, resource management, tax collection and management, etc., so it is more reasonable for the ownership to be attributed to the state in order to facilitate the government to carry out its management functions This type of data has both public and personal attributes, and therefore all parties have ownership and use rights. The scope of use and ownership.

6. Conclusion

Data is the core resource of the digital economy and information society. The protection of data is a matter of urgency, but in China, there are no specific legal provisions on data ownership, and the law against unfair competition does not fully protect the rights and interests of data, so the competition between the interests of all parties reflected in the data ownership will be more acute and obvious. At present, all aspects of economic and social life are gradually permeated by data and its derivative data sets, leading to conflicting and symbiotic systems of legal rules.

The existing laws and regulations on data ownership and other data rule systems lag far behind the development of data practices. A single model of data protection that is too simple or a complex model that is too diverse is not conducive to the enhancement of social welfare and the optimal allocation of social resources. On the one hand, the analytical construction of data ownership can clarify the boundaries of the scope of data products and services and safeguard the legitimate rights and interests of rights holders. This model construction is conducive to the formation of a stable cooperation relationship between relevant stakeholders in the data industry, promoting the benign development of the industry and preventing data from being improperly restricted or circulated arbitrarily. On the other hand, as data resources have become an important element in the market economy, the data thorn phenomenon has forced data rights holders to adopt a strategy of duplication of competition rather than mutual cooperation in order to enhance their own interests, which may result in the inability to fully exploit the excellence of data and hinder competition in the domestic market and with overseas markets.

Therefore to avoid both of these tragedies perhaps lies in discussing the construction of a data ownership framework that focuses on maintaining a balance of data interests between individual users, platform companies, and the government and state within a rule of law framework. Of course, only when data ownership is considered in conjunction with data collection, data sharing, data flow and data security will there be a viable basis for relevant policy proposals.

References

[1] Big Data: What Is Web Scraping and How to Use It, Towards Data Sci [EB/OL]. (2018-02-02) [2021-07-08].https://towardsdatascience.com/big-data-what-is-web-scraping-and-how-to-use-it-74e7e8b58fd6.

[2] Su Qing. The Evolution of Web Crawlers and Their Legitimacy Limits [J]. Comparative Law Research, 2021(3):89-104.(In Chinese)

[3] Javed M. Ho Do Internet Search Engines Work? Scientific American [EB/OL]. (2004-11-29) [2021-07-09].https://www.scientificamerican.com/article/how-do-internet-search-en/.

[4] Su Qing. The Evolution of Web Crawlers and Their Legitimacy Limitations [J]. Comparative Law Research, 2021(3):89-104.(In Chinese)

[5] Liu Jifeng, Zhang Ya. Determination of the illegality of data grabbing from the perspective of the Anti-Unfair Competition Law [J]. Journal of Northwestern Polytechnical University (Social Science Edition), 2021(04):102-111.(In Chinese)

[6] Wang Jian, Wu Zongze. On Data as a Necessary Facility in Anti-Monopoly Law [J]. Law Research, 2021(2): 102.(In Chinese)

[7] Peng Hui. The Logical Structure and Empowerment Boundary of Data Ownership——Based on the Perspective of "Tragedy of the Commons" and "Tragedy of Anti-Commons" [J]. Comparative Law Research, 2022.(In Chinese)

[8] Mei Xiaying, Luo Ying. The Legal Attributes of Data and Their Civil Law Positioning (English) [J]. Social Sciences in China, 2019, 1.(In Chinese)

[9] Zhang Jinping. The Evolution of EU Personal Data Rights and Its Enlightenment [J]. Law and Business Research, 2019, 5.(In Chinese)

[10] Long Weiqiu. Research on the construction and system of new data property rights [J]. Tribune of Political Science and Law, 2017, 35(4).(In Chinese)

[11] Yan Xiaoli. Reform and Enlightenment of EU Data Protection System [J]. Cyberspace Security, 2017 (2): 22-26.(In Chinese)

[12] Cheng Xiao. Personal Information Protection from the Perspective of Civil Code Compilation [J]. Chinese Law, 2019, 4: 26-43.(In Chinese)

[13] Ding Xiaodong. Who does the data belong to? [J]. Platform data ownership and data protection from the perspective of web crawlers, 2019, 22(205): 69-83.(In Chinese)

[14] Wang Xixin. National Protection Obligation and Development of Personal Information [J]. Chinese Law, 2021(In Chinese)

[15] OECD. Enhancing Access to and Sharing of Data--ReconcilingRisks and Benefits for Data Reuse across Societies [EB/OL]. (2019-11-26) [2021-07-08]. https://www.oecd.org/ going- digital/ enhancing-access-to-and-sharing-of-data.pdf.

[16] Zhou Hanhua. Exploring the way of incentive-compatible personal data governance - the legislative direction of China's personal information protection law [J]. Legal Research, 2018, 40(2): 3-23.(In Chinese)

[17] Mei Xiaying. The Limitation of Private Law and the Construction of Public Order in Data Protection between Sharing and Control [J]. Chinese and Foreign Law, 2019, 31(4): 845-870.(In Chinese)