To Whom Do Data Belong?

——Data Ownership and Protection in the Context of Web-Crawlers

Ding Xiaodong*

Renmin University of China

Abstract: Platform data has already become an important asset for web-based companies, but this sort of data frequently includes large amounts of personal information. Platform data can be seen as belonging to an individual, belonging to a platform, belonging to some combinations of the two, or can be seen as a form of Internet-based public data. Analysis of legal clauses and doctrines as well as analysis based in legitimacy and consequentialism both fail to completely delineate data ownership. One potential reason for this is that there are many types of platform data, and that each type is highly dependent on circumstances. The determination of rights in regard to platform data should be done in a way which revolves around a contextual regulatory framework, one in which the rules of reason is applied on a case-by-case basis and in which gradual changes are done in a bottom-up manner, and not one which seeks to establish a universal set of data regulations. In actual judgments, factors such as the nature of the platform and the nature of the data crawling behavior should be comprehensively considered while ensuring a balance of data circulation and data protection.

Keywords: data ownership, data protection, web-crawling, unfair competition

DOI: http://dx.doi.org/10.19873/j.cnki.2096-0212.2020.06.007

* Ding Xiaodong, Law School, Renmin University of China.

This paper is a periodic achievement of two projects of the National Social Science Fund of China, those being the major project:"Personal Data Protection and Data Rights Systems in the Age of Big Data" (Project No. 18ZDA146), and the general project: "Research on Personal Information Protection and Corporate Data Ownership in the Context of Big Data" (Project No. 18BFX198).

Correspondence concerning this article should be addressed to Ding Xiaodong, Law School, Renmin University of China, Beijing. E-mail: dingruc@163.com.

Data plays a crucial role in the development of web-based companies. The more users an Internet company and/or platform have, the more users it can attract, and the more advantageous a position the company will have in its competition with other web-based companies. This snowball-like growth effect means that web-based companies often treat data as a central competitive asset (Katz & Shapiro, 1985, p. 424; Lemley & McGowan, 1998, p. 479). Whichever Internet company can possess the most data and utilize such data in the most effective way will possess a competitive advantage.

Perhaps because of data's importance, there has recently been a slew of disputes involving it. Some examples of this are Huawei and Tencent's data dispute, the disagreement between Shunfeng and Cainiao, the case of Sina vs. Maimai, the case of Dianping vs. Baidu, the unfair competition dispute brought by Taobao against Meijing, the case of Craiglist vs. 3Taps in the United States, and the case of HiQ vs. LinkedIn. In all of these cases, the central question at hand is data, namely, when a web platform acquires data from another platform through technical means, is that behavior legal or reasonable? Or, in simpler terms: exactly to whom does a platform's data belong? ①

There have already been a large number of researches done on this question within legal academia, but researches have tended to look at questiones from the perspective of departmental law and, accordingly, many researches done have focused on the issue in this context. For instance, some scholars have analyzed the legality of data crawlers from the angle of the Anti-Unfair Competition Law of the People's Republic of China (Zhang, 2013, pp. 46-51; Ning & Wang, 2016, pp. 161-168; Fan, 2015, pp. 84-94; Yang & Qu, 2013, pp. 30-34; Yang, 2014, pp. 12-21; Fan, 2015, pp. 84-94; Liu, 2018, pp. 26-30), some scholars have analyzed corporate data property right protection from the perspective of civil property rights (Long, 2017, p. 75; Long, 2018, p. 50), and other scholars have analyzed corporate data protection from the perspective of intellectual property law (Xu, 2018, p. 56). Although this type of research has provided valuable insights into the question of data ownership, it has failed to examine the question of data ownership from a comprehensive perspective, especially in regard to platform data issues (Yao, 2019, pp. 114-125; Hu, 2017, pp. 1-14). ② Furthermore, although economic literature has increasingly recognized data ownership issues and provided helpful knowledge for their analysis (Fei, 2018, pp. 3-21; Du, 2018, pp. 19-25; Wang, 2015, pp. 131-135), data ownership issues are still not a purely economic question, and their legitimacy cannot be solely established on the foundation of pure efficiency analysis. For instance, from the sole perspective of efficiency, platform ownership of data is the most efficient because the centralized application of data on a large scale can efficiently resolve the externality and exchange costs produced by data. This sort of cursory analysis, however, does not consider the issue in the context of personal privacy and data circulation in larger-scale public domains. One extreme example of these shortcomings is that from the economic perspective, a platform could utilize private personal data to incentivize or even threaten individual

① In the ten questions raised by Ailbaba's Luohan Academy in June 2019, one of the questions "To whom do data belong? Who really benefits from them?" See Alibaba's Luohan Academy Raises 10 of the Most Important Questions for the Future of the World and Scholars' Response.

② The two scholars Yao Jia and Hu Ling considered the corporate data ownership issue from the perspective of data use and business models.

labor, with the result of increasing efficiency. This sort of system, however, evidently may not be rational. ① Therefore, although economic literature has provided beneficial analysis of platform data ownership questions, this sort of research can only form one part of a more comprehensive analysis.

Based on these considerations, I conducted comparatively comprehensive researches on platform data ownership to examine the issue of platform data ownership by looking at the typical technique of web crawling in data disputes. Web-crawling refers to a method by which information is automatically collected through a code or script in the worldwide web according to a certain set of rules. Throughout the development of the Internet, both web crawlers and anti-web crawlers have become extremely common and because of their ubiquitous use, I have analyzed the issue of data ownership and data protection from the perspective of web-crawling.

Let me be blunt, there is no clear way to determine the ownership of platform data. Platform data can include various types of data, including large amounts of personal data, in regard to which individuals possess relevant privacy rights. Platform data are also collected by corporations, and these corporations have related rights and interests in regard to this data. Platform data can also be part of the public domain, where neither an individual nor a corporation possess exclusive rights to it. Besides this, the nature of platform data is also highly dependent on context. Based on these characteristics, I believe that contextual protection should be applied to platform data, and that whether it be individual or corporate data, regulations should be determined in a bottom-up manner, on a case by case basis. In the consideration of each case, one must consider the nature of the platform, the nature of the data, and the nature of the web-crawlers in order to ensure the proper balance of data privacy protection, the protection of corporate data rights and interests, and data access.

Web Crawling and Data Disputes

Search engines such as Google, Baidu, Sougou, and Bing, the context in which web-crawling techniques were first applied, remain the most common context in which such technology has been utilized. As far as search engines themselves are concerned, the use of such search engines by web-crawling technology is a process from which all participants benefit. The search engine is able to realize its potential to efficiently collect and sort information while the web pages that are crawled are more widely disseminated through the search engine.

Despite this, parties who do not want their data to be crawled quickly emerged. Internet-based companies gradually developed two applied methods to counter web-crawling. The first method was the development of a gentlemen's agreement: The Robots Exclusion Standard (also Robots Exclusion Standard or Robots Exclusion Protocol), which gave rise to a set of documents called the robot.txt file

① In the field of economics, a Russian economist once gave an example of this. If land ownership is allocated to the landlord and the landlord guarantees a particular level of hunger among the peasants, then from the perspective of efficiency, this would be beneficial to the maximum use of the land, because peasants will work tirelessly, but this sort of system obviously lacks legitimacy. See (Russian) Chayanov: Peasant Economic Organization, Zhenghong Xiao, Guanze Chenyue, Zhongyang Bianyi Publisher 1996 Edition.

developed by website owners. This file was placed in the root directory of a given website's server and indicated which webpages in the directory were not allowed to be seized through crawling. Friendly crawlers would often first read the robot.txt file before collecting information from a website and subsequently refrain from downloading the web page which has been restricted in the file. The second method developed by web company employees was a technical approach to countering web-crawling, in which various technical changes were implemented in order to prevent web-crawlers from visiting the page. An example of this is that one can configure a website to require a visitor to input a password if the page is accessed too quickly, thereby excluding non-human page visits.

Another example of this is that a website can change its HTML tags from time to time, to make it impossible for web crawlers to match the structure of the webpage.

During this period in which web-based companies were taking part in both web-crawling and methods to counter them, legal disputes revolving around data began to emerge. In 2000, a company named Bidder's Edge carried out a web-crawling operation on eBay's website. eBay filed claims in the Northern California court system, alleging that Bidder's web crawling activities on eBay's site went against their Robots Exclusion Standard, in which their behaviors constituted a trespass against eBay and an instance of computer fraud and misuse, and finally violated the Anti-Unfair Competition Law of the People's Republic of China. In the end, the court found Bidder's Edge responsible for the charge of trespass, believing that the defendant had not received authorization to interfere with the plaintiff's ownership rights in regard to information stored on the computer, and that the behavior directly harmed the plaintiff.

In this case, one of the defenses that Bidder's Edge proposed was that because all of the information on eBay's website was publicly available, there had been no trespass. In response to this, the court's opinion was that eBay's server was private property, and that the public right to access it had to be granted by the company. eBay did not typically allow web crawlers to access the site, and in this case, eBay explicitly told Bidder's Edge that it was not allowed to use robots to crawl eBay's site. Therefore, there was trespass present in this case.

In regard to the right to access webpages, the reasoning behind the judgments made in Chinese cases has been relatively consistent with the reasoning in the eBay case. For instance, in Sina vs. Maimai and Dianping vs. Baidu, the courts held that web-crawling without permission and the collection of large amounts of data from the other party's website constituted illegal behaviors. In these cases, courts commonly referred to the regulations within the Anti-Unfair Competition Law of the People's Republic of China, finding that this sort of behavior "disrupts the order of market competition and harms the legal rights of other operators and consumers," and, therefore, is in violation of the second article of the Anti-Unfair Competition Law of the People's Republic of China, which states that "Businesses shall, in their production and distribution activities, adhere to the free will, equality, fairness, and good faith principles, and abide by laws and business ethics."

Of course, there have also been cases with different reasoning behind their judgments. In HiQ vs. LinkedIn, HiQ took measures to crawl LinkedIn's website, but a judge of the the United States

District Court for the Northern District of California held that this behavior was not in violation of the law because the data on LinkedIn's site is public data, and that even if the behavior violated the Robots Exclusion Standard set up by LinkedIn, it was still permitted by the law. The behavior was akin to pushing open a store's unlocked door in broad daylight and having a look inside, in no way could it be considered illegal trespass. Based on this, the court did not find HiQ's web-crawling behavior in violation of the law but in fact found LinkedIn's counter-web-crawling technology illegal and required LinkedIn to remove barriers to access it had instated against HiQ.

One thing that makes data disputes that arise from web-crawling even more complicated is the fact that website data often originates from individuals, meaning that there are data privacy issues associated with such behavior. One example of this is that in the aforementioned Sina vs. Maimai case, Sina accused Maimai not only of violating the company's Robots Exclusion Standard, but also asserted that Maimai's web-crawling had not received the authorization of users. In the case of HiQ vs. LinkedIn, LinkedIn also brought up the issue of data privacy protection and asserted that HiQ's crawling activities against LinkedIn would impact such protections. In regard to the question of whether collecting data from online platforms requires individual authorization, the courts have given different judgments. For instance, in HiQ vs. LinkedIn, the court held that crawling would in no way influence the protection of citizens' privacy, but in Sina vs Maimai the court clarified the necessity of user authorization outside of platform authorization.

In the web-crawling and data dispute between Jinri Toutiao and Weibo, Jinri Toutiao emphasized that users have rights to personal data. Weibo believed that its data were illegally crawled by Jinri Toutiao,[①] but Jinri Toutiao believed that this sort of data belonged to the user but not to Weibo, and that as long as a site has user authorization, it could legitimately carry out web-crawling. Jinri Toutiao believed that its web-crawling behavior was not in violation of the law because the app's front page had an option inviting user authorization, and it was only after users enabled this option authorizing Jinri Toutiao to seize the user's Weibo data that the crawling occurred. This function allowed users to automatically post the information posted on Weibo on Toutiao's "Mini-Toutiao" product for a set time period.

Four Views Concerning Data Ownership

We can now proceed to summarize views on data ownership. With regard to platforms with large amounts of personal data, we can largely categorize views on data ownership into four categories.

Individual Ownership of Data

This view of data ownership is that data belongs to the individual user. In the dispute

① In Weibo's view, "a certain third-party news platform directly seizes content from individual media accounts without the acknowledgment or authorization of Weibo".

between Jinri Toutiao and Weibo, Jinri Toutiao's position is a classic example of this point of view. Jinri Toutiao believed that Weibo had no right whatsoever to user data and that therefore, as long as web-crawling was authorized by the user, even if Toutiao violated Weibo's Robots Exclusion Standard, the behavior would not be in violation of the law. Weibo can, of course, choose to bring suits against users, especially against large, high-profile users for violating their agreements with the company as the company's user agreement clearly states that Weibo enjoys the exclusive right to user content, and many of these high-profile users have signed very clear contracts. This way, whenever users, especially high-profile users, post content on Weibo and authorize Jinri Toutiao to use that content, Weibo can sue those users and require courts to find this behavior in violation of the law. Even if courts were to find such behavior illegal, Jinri Toutiao can claim that its behavior is not illegal and that although the user behavior is illegal, it has nothing to do with Jinri Toutiao.

In reality, if individual ownership of user data was strengthened and the right to personal data was viewed as a right of publicity but not as a right to property, or if the right to ownership was seen as a legally established consumer right (Ding, 2018, pp. 45-50), then Weibo's user agreement would probably be rendered null and void from the outset. Once individual ownership of data is viewed as an untransferable right of publicity, it follows naturally that data collectors and users cannot limit the free exercise of this right to data. Just as private individuals cannot limit a citizen from freely using their own name through a contract (Hansmann & Kraakamn, 2002, pp. 368-387), corporations would be unable to require individuals to forfeit their rights to data through such contractual means.

The right to data portability recently established by the EU can be viewed as another manifestation of this sort of individual right to data. If the right to data portability established by the "General Data Protection Regulation" (GDPR) is recognized, it follows that platforms not only cannot limit personal data, but additionally must provide assistance to ensure the free circulation of personal data. The GDPR stipulates that "the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used, and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided." According to this data right, users would even be able to require Weibo to make itself more open to other platforms, so that the free circulation of their data could be achieved.

Platform Ownership of Data

The second view of data ownership is that data belongs to the platform. The most typical version of this view is encapsulated in the new user agreement Weibo put out in the aftermath of the breakout of its dispute with Jinri Toutiao. This agreement stipulated that "Regarding information uploaded on Weibo by users, including but not limited to text, pictures, video, audio, no matter if the Weibo content can be established as a protected object under copyright law, users agree that they cannot rescind the Weibo platform's exclusive right to post Weibo content, and that the Weibo content posted

by users can be exclusively displayed on the Weibo platform." ① This new user agreement in essence defined data as belonging to the platform and eliminated user's right to authorize the use of Weibo content.

As can likely be imagined, the idea that data belongs solely to the platform is not well received by most. After Weibo released this new user agreement, it was met with fierce resistance and criticism from users and media alike, resulting in Weibo clarifying the agreement and eventually amending it. This newer user agreement stipulated that users have copyright over content posted on Weibo, and that as a posting platform Weibo only enjoys a limited right of use over such content. Users can distribute content to which they have the complete right to other platforms according to their own wishes, with no need to receive Weibo's approval, examination, or agreement. Notwithstanding this however, the post-update user agreement still emphasized that individually authorizing, permitting, or assisting a third party in illegally seizing content already distributed on Weibo without Weibo's approval is still illegal. Therefore, the adjusted user agreement means that while Weibo does not have a relative right to data relative to users, it does enjoy a relative right to data relative to other platforms.

Combined Ownership of Data by Both Individuals and Platforms

The third view of data ownership is that it is owned by both individuals and platforms. This view is commonly expressed in the judgments of Chinese courts. For instance, in the case of Sina vs. Maimai the court held that the premise of data openness must be authorized by both the individual user and the platform. In addition to this, in order to emphasize the importance of personal data protection the court also proposed the "threefold authorization" model of "user authorization" + "platform authorization" + "user authorization" which entails that the data provider must receive user approval before collecting data, and that when the data provider authorizes a third party to use this sort of information, the third party platform must clearly notify the user of the goal, manner, and scope of the information's use, and once again receive the user's approval. The line of reasoning held within this judgment by the courts means that users and platforms both maintain a certain right to data, and that data are to a certain extent commonly owned by both the individual and the platform.

Of course, within the framework of combined ownership of data by individuals and platforms, the delineation of this right between both parties is still an issue. In the case of a web platform in competition with a given platform undertaking web-crawling, twofold and threefold authorization regulations are quite sensible and would be realistically implementable. In other contexts, however, requiring platforms and individuals to go through twofold or threefold authorization is likely to result in all sorts of challenges. For instance, when a user copies and pastes large amounts of data from one platform to another platform, this behavior obviously has not received the authorization of the

① This user agreement at the same time stipulated that "without the prior written permission of Weibo, the user must not authorize any third-party platform in any way to directly or indirectly use Weibo content, including but not limited to authorizing any third-party to post, copy, transfer, edit, quote, link, download, synchronize, or in any other way use a part or the whole of any Weibo content, nor should the user themselves do anything listed above."

platform, but does this sort of behaviors violate the commonly held data property right? Besides this, when the right to ownership of a platform's website is transferred to another party, does this property transfer need to obtain approval from users? In 2018, Renren Network was sold to Duoniu Media company, and the property in this sale included user data. Throughout the entire process of the sale however, Renren Network never went through a phase of obtaining user approval. Undoubtedly, requiring Renren Network to obtain the approval of all its users before the sale would not have been realistic.

Public Ownership of Data

The fourth view of data ownership is that data is publicly owned. This view holds that once a platform involves the Internet, the platform's data possesses a public nature and is not owned by any individual or corporation. In the case of HiQ vs. LinkedIn, HiQ hired professor Laurence Tribe from Harvard Law School as an advisor, with Tribe believing that the right to access data and information is akin to the right to free speech, which is protected by the First Amendment of the U.S. Constitution and with the essence of speech being its capacity to be circulated and shared, that right possesses a public character. Therefore, seizing data does not require the authorization by the platform or the individual.①

Professor Orin Kerr, an Internet law scholar, once typified the public nature of the Internet. In his view, the general principle of the Internet is its openness. Thus, this sort of openness allows anyone from around the world to distribute information and data, and data can be accessed by anyone without undergoing identity verification. When a computer owner decides to set up a web server on their device and allow documents to be accessible through the web, it is presumed that the documents can be accessed by everyone (Kerr, 2016, pp. 1143, 1163). Professor Kerr also drew an analogy, stating that connecting a server to the Internet is akin to putting a product up for sale at a public trade fair, anyone can access data on the web just as anyone can access the product at the trade fair (Kerr, 2016, p. 1163). It is only in special circumstances, such as when a webpage sets up a password, that a webpage changes from an open webpage to a closed webpage (Kerr, 2016, p. 1161).

The Chinese web commentator Fang Xingdong has expressed similar views. Fang believes that from the earliest incarnation of the Internet, ARPAnet, to the later TCP/IP agreements, in addition to a series of Internet regulation mechanisms and technical standards organizations, it has been "firmly established that the core values and technical regulations of the Internet are openness, accessibility, freedom, and equality" as well as "connection without discrimination, selection, or conditions," but currently the Chinese Internet industry has been increasingly undertaking a "high-wall" approach to data and site traffic. Therefore, Fang believes that whether it was Taobao refusing to allow Baidu to search stored page information, the "3B conflict" in which Baidu is attempting to refuse to allow 360

① Tribe points out that LinkedIn and Facebook are the modern equivalents of the "town square," and if one wants to ensure the speech is able to get its meaning across, then private social media platforms must be treated as public forums.

searches through its Robots Exclusion Standard, WeChat's repeated selective blocks of competitors like Didi, Taobao, Jinri Toutiao, and TikTok, or Baidu's large scale diversion of searches to its own site while not even displaying outside websites in search results, all go against the original spirt of the Internet.

Data Ownership: An Analysis of Legal Clauses and Doctrines

Which one of the four views on data ownership makes the most sense? In order to answer this question, we must first analyze the delineation of the individual right to data and the corporate right to data from the perspectives of legal clauses and tenets. This analysis will demonstrate that the delineation between the two is not at all clear.

The Individual's Data Rights

First, the scope of personal data and personal data rights are both uncertain, resulting in the scope of protected platform data to also be similarly uncertain. Originally, the laws of China and other countries provided for the protection of personal data. Companies, societies, and governments all had a common understanding of the priority of personal data protection. For instance, in the cases of data disputes between Tencent and Huawei, Cainiao and Shunfeng, and Jinri Toutiao and Weibo all parties saw personal data protection as of the utmost importance and emphasized the importance of obtaining user authorizations, but the problem lies in whether the different sorts of data produced by users on platforms can be considered as personal data. Should personal data receive the same level of protection in all different application contexts?

According to the prevailing definition of personal data or individual information, both are data that have been or can be discerned as individual (Schwartz & Solove, 2011, pp. 1814-1815). For instance, the Cybersecurity Law of the People's Republic of China stipulates that individual information refers to "all kinds of information, stored in electronic or other form, which individually or in combination with other information allows the identification of a natural person's individual identity." The EU's "General Data Protection Regulation" defines personal data as "any information concerning an identified or identifiable natural person." According to this prevailing definition however, different categories of data produced by users on web-platforms can possibly be considered personal data or not personal data. This is because the platform's user data have the potential to directly or, in combination with other information, identify an individual, or can essentially have no use in identifying the individual. Whether the individual can be identified is determined largely by the specific application context, the subject being identified, and the identification's degree of difficulty. To take user comment data as an example, this sort of data displayed anonymously on a platform would render it difficult to identify the associated individual for the average person, but if this data was viewed in combination with information such as the respective user's purchase history and tracking history, this data would possibly be able to be used to identify the individual person, and

to the user's close associates, perhaps only one comment would be enough to identify the individual.

Besides this, the boundaries of the individual right to data are also uncertain, and it is difficult for individuals to establish an exclusive right to their own data. A major contributor to thought on data privacy, Alan Westin, once defined data or information privacy more broadly as control over information (Westin, 1967, pp. 7). ① This framework was accepted by the legislatures of various countries and regions and forms the foundational reasoning behind these jurisdiction's data privacy laws (Ding, 2019, pp. 96-110), but the problem lies in the fact that there are enormous differences in how the law confers these data rights in different countries, regions, and contexts. The law may confer rights such as the right to access and data security rights (Cate, 1997, pp. 370-373), and also possibly newer rights described previously such as the right to be forgotten or the right to data portability (Ding, 2018, pp. 94-107). Whether it be among different countries or experts, consensus on this issue still has not been achieved.

The uncertain characteristics of personal data make it difficult to draw a boundary between individual and corporate data, or even render differentiations, which at first glance, seem relatively clear, subject to doubt. For instance, in research on corporate data, much of the research classifies data as either original or processed data, with original data commonly including personal data and processed data not being personal data, due to having undergone the processing and demarcation process. An example of this is data based on summed up personal data, which people commonly do not consider as personal data, and whose ownership is seen as belonging to the corporation. This sort of differentiation, however, still faces some challenges. If the right to delete data is conferred and an individual requires the thorough deletion of their personal data, or an individual clearly requires that any handling of their data be rescinded, then corporations will face controversy over processed or summarized data based on original data. ②

The Platform's Data Rights

In regard to the platform's right to data, contrasting the legal protections provided to corporate data by various countries leads one to realize that there are still many controversies in this area, and that there is no commonly recognized legal boundary in terms of a platform's data rights.

First, it is difficult to delineate the platform's data rights through existing legal protections of databases and intellectual property rights. In terms of substance, platform data are most akin to a database in that both are the agglomeration of a large amount of data. In the case of databases however, countries have vastly diverse views when it comes to their legal status. The database protections of the US touch upon the key elements involved in the original compilation of the data and

① Westin believes that privacy is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

② This is not to say that I support the EU's right to erasure or right to refuse handling in respect to data. In my view, although the individual has the right to require that companies delete data or to refuse the handling of that data by a company in certain contexts, such rights should not obstruct statistical use (or such similar uses) or personal data.

offer no protection to the actual data contained within. In the case of *Feist Publications, Inc., v. Rural Telephone Service Co.*, Feist copied the entire contents of a phonebook, which had been compiled by Rural Telephone Service Co, without their authorization. The Supreme Court of the United States held that this behavior was not in violation of copyright because copyright only protects the original compilation of data and not the unoriginal content held within. Europe, taking a different approach than that of the US, not only protects the original compilation portion but also provides protection to databases in the form of *sui generis* rights. According to the stipulations of these *sui generis* rights, when "creating a database requires enough manpower, technical, and financial resources" the database is legally protected. Once the database has been created, others cannot use or copy either the entirety or majority of the data contained within the database.

One of the reasons intellectual property rights and other related laws have different views on the rights associated with databases is that data have so many attributes. On one hand, database creators undoubtedly expend much labor through the process of collecting and arranging data. From the perspective of classical labor theory of property (Locke, 1982, pp. 20-25), databases ought to receive the protections guaranteed by property rights, or at least some approximation of those protections. Historically, in the US, lower courts have acknowledged the so-called "sweat of the brow" and "industrious collection" doctrines, which hold that if a large amount of labor is expended in collecting data when producing a database, it should then receive legal protection. On the other hand, data have a strong public character, and data themselves should not be considered legally recognized private or intellectual property solely because there has been labor expended towards its end. Because data, when compared with other movable and immovable property, is obviously neither exclusive nor competitive in character, it is hard for data to be exclusively owned by an individual, especially considering that public use of data does not result in their deterioration. The Supreme Court of the United States clearly rejected the "sweat of the brow" doctrine put forward by the lower courts and emphasized that copyright only applies to portions involving creativity and that data themselves should be maintained as publicly owned. The U.S. Supreme Court clarified that if the legal protections afforded to databases were extended to the data that they contained, this would "distort basic copyright principles."

Furthermore, Contract Law of the People's Republic of China provides no help in determining the boundaries of the platform's right to data. Whether Robots Exclusion Standard can be seen as contractually binding is a matter of much controversy among the different judiciaries and legal doctrines of various countries. Robots Exclusion Standard can be seen as an expression similar to a contract, in that they communicate the wishes of an involved party to others, but does the act of the web crawler simply reading the agreement constitute the establishment of a contract? In practice, various countries have come to different judgments in regard to this sort of unilateral notification contract. For instance, in the case of standard form contracts within software installation packages or so-called "shrink-wrap licenses," some courts have held that once a presumed consumer having been able to see this sort of notification chooses to continue downloading the software, it is at that moment

that the unilateral notification can be considered a contract. In the judgments of other courts however, it has been held that this sort of contract is null and void.

In substance, Robots Exclusion Standard is not unlike signs found hanging inside the premises of many small Chinese retailers that read "Competitors Need Not Enter" or such and such a person "Need Not Enter." Private law does not give a clear answer to the question of whether these sorts of notifications can be considered contractually binding. On one hand, this sort of notice is reasonable to a certain degree as it conforms to the party autonomy principle by clearly indicating the store's wishes. On the other hand, however, this sort of notice could be considered null and void *ab initio*. When this type of notice is targeted at a specific group, it can be deemed as null and void for being against public order and good morals, or could also be considered null and void for violating the anti-discrimination principle of public law (Ding, 2014, pp. 1080-1096). ① Besides this, even if this sort of notice was contractually binding, it does not necessarily mean that the person seeing it accepts the contract therein, they could simply see it as a friendly, non-binding reminder, meaning that entering the store would not be equal to the establishment of a contract.

Furthermore, from the perspective of tort and criminal law, there is no clear standard for whether violating a Robots Exclusion Standard can be considered as infringement of rights or as a trespass of a computer system. Tort liability is generally established by four major factors, those being duty, breach of duty, causation, and injury. In the case of data crawling, however, it is difficult to establish that injury has taken place. In the majority of cases, web-crawling between web platforms is sustained and occurs over long periods of time and neither overwhelms the site traffic of the crawled site nor causes the site's Internet speed to decrease. According to the concept of trespass in common law and the crime of illegal seizure of computer system data in the Criminal Law of the People's Republic of China, ② whether data crawling is considered an illegal trespass of a computer system is itself determined by how the law defines the nature of data crawling.

Professor Kerr once systematically analyzed the question of illegal trespass on the Internet by comparing offline and online circumstances. Professor Kerr likened the technical anti-crawling obstacles (such as Robots Exclusion Standard, authentication codes, passwords) set up by Internet-based companies to physical barriers (signs set up by storefronts, fences, closed doors, locked doors) in the offline world. Professor Kerr pointed out that in the case of both the physical barriers and technical barriers, the law offers no standard guide for whether such barriers may be crossed or the boundaries of illegal trespass. As professor Kerr said, "Like their physical-world cousins, computer trespass laws feature unilluminating text. They prohibit unauthorized access to computers just like physical trespass laws prohibit unlicensed entry to physical spaces." Whether in the physical world or

① Once this sort of discrimination touches upon a factor of identity, it may be in violation of various countries' anti-discrimination regulations and principles.

② Article 285 Clause 2 of the Criminal Law of the People's Republic of China describes the crime of illegal acquisition of computer information system data. This crime stipulates that "whoever, in violation of the state provisions, intrudes into a computer information system other than that prescribed in the preceding paragraph (state affairs, national defense construction, or sophisticated science and technology) or uses other technical means to obtain the data stored, processed or transmitted in the said computer information system or exercise illegal control over the said computer information system shall, if the circumstances are serious receive criminal punishment".

the online world, "the meaning of the law must draw from social understandings about access rights drawn from different signals within the relevant spaces. Courts must identify the rules of different spaces based on understandings of the relevant trespass norms."

Finally, the Anti-Unfair Competition Law of the People's Republic of China also faces similar problems in addressing this issue. Many recent Internet data disputes in China quote the second article of the Anti-Unfair Competition Law of the People's Republic of China which concerns stipulations on business ethics. ① For instance, in the cases of both Sina vs. Maimai and Dianping vs. Baidu, courts cited data crawling as violating business ethics and constituting unfair competition in the reasoning behind their judgments. It must be indicated, however, that the judgments of these courts are mainly established on the basis of these particular cases and an assessment of their specific circumstances. The Anti-Unfair Competition Law of the People's Republic of China itself gives no rigid framework rules for what constitutes business ethics. As stated by many experts, laws against unfair competition must rely on other legal provisions and business norms to determine what can be considered business ethics, and that there often exists a high degree of inherent uncertainty within this area of law (Feldman, 2006, p. 197).

Data Ownership: A Pragmatic Analysis of Consequences

If legal clauses and tenets cannot provide a clear answer regarding the issue of data ownership, then can an analysis of consequences based on pragmatism establish the right to data ownership? In combination with the four views of data ownership summarized previously, it can be observed that it is difficult to establish any of the four views as wholly correct or reasonable.

First, allocating the right to data solely to the individual is unrealistic and would produce extraordinarily high transaction and communication costs. If the individual has an absolute property right to data, then, that would mean that platforms or other individuals would have to receive personal approval to access this sort of data. Under this sort of system, normal web crawling behavior such as that done by search engines would be unimplementable, and even the reading of personal data by other individuals would be in violation of the law (Ding, 2018, pp. 194-206). Besides this, determining that individuals have the sole right to data would make certain rights the platform enjoys in respect to the data impracticable, and would render platforms unable to undertake certain normal business activities or even be guilty of infringing on the information of citizens. For example, platforms would be unable to enter into exclusive user agreements with users such as Internet celebrities; Renren's sale of its website would not only constitute an illegal commercial activity but could also constituted an infringement on the information of citizens. ①

① Article 2 of the Anti-Unfair Competition Law of the People's Republic of China stipulates: Businesses shall, in their production and distribution activities, adhere to the free will, equality, fairness, and good faith principles, and abide by laws and business ethics. For the purposes of this Law, "act of unfair competition" means that in its production or distribution activities, a business disrupts the order of market competition and causes damage to the lawful rights and interests of the other businesses or consumers, in violation of this Law. "Operator" in this Law refers to a natural person, a legal person or an unincorporated organization engaged in production and marketing of goods ("Goods" include services when used hereinafter) or provision of services.

Second, allocating the right to data solely to the platform would go against common sense. Platform data ownership would not only negatively impact the individual's copyright or other intellectual property rights but would also possibly make it impossible to protect citizens' data privacy. Even if the data are found on the Internet, that doesn't necessarily mean that they can be arbitrarily used by a third-party platform. The most well-known example of this sort of issue is the Facebook-Cambridge Analytica scandal, in which Cambridge Analytica collected the user information of three hundred thousand users through an app and received the information of fifty million users from this group of three hundred thousand users' friends lists with Facebook's authorization. Although all of this information is publicly available online, its public availability obviously has a specially designated target and context. Cambridge Analytica's collection of these information without user authorization, as well as its use of these information in a context completely separate from its intended one, constituted an invasion of user data privacy.

Furthermore, allocating ownership to both the individual and the platform would create issues preventing the circulation and sharing of data. As discussed previously, when a platform trades data or shares it with another party, it can often be difficult to obtain user approvals in such circumstances. In the same way, when a normal user who wishes to transfer his or her personal data and must obtain the authorization of the platform, the transfer becomes difficult because many platforms may not want to see users leave, this being the attitude expressed in Weibo's user agreement. In summary, combined platform and individual ownership of data would further increase the systemic costs of data circulation and sharing.

Finally, while recognizing data as a public good may promote data circulation and sharing, it renders protecting the individual right to data as well as the rational data interests of platforms an impossible task. In the case of individuals, the public and connected nature of the Internet does not mean that there are no privacy issues associated with publicly available personal data, nor does it mean that this sort of data can be seen purely as a public good. In certain contexts, personal data can bring with it a series of privacy issues or can even be a product of an individual's digital labor, combining the labor and output of an individual. On the other hand, platforms often invest large amounts of capital and labor into both developing the platform itself and the process of collecting data. A system with absolutely no protection for the legitimate data interests of corporations would essentially allow for what in economics is referred to as "free-ride" behavior, in which it is difficult to protect and promote investment and market competition (Tamaroff, 2011, p. 16).

No matter to which party data ownership is allocated, there will always be issues. The deep-seated

① Article 253A of the "Criminal Law of the People's Republic of China" stipulates: "(Where one is) in violation of the relevant state provisions, sells or illegally provides personal information on citizens, shall, if the circumstances are serious, be sentenced to a fixed-term imprisonment of not more than three years or criminal detention in addition to a fine or be sentenced to a fine only; if the circumstances are especially serious, (the offender) shall be sentenced to imprisonment of not less than three years but not more than seven years in addition to a fine." The Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues concerning the Application of Law in the Handling of Criminal Cases of Infringing on Citizens' Personal Information further provides that "whoever provides any citizen's personal information legally collected to any other person without the consent of the person whose information is collected shall fall within the scope of 'providing citizens' personal information' as prescribed in Article 253A of the Criminal Law." This interpretation also stipulates that illegal acquisition and sale of over five thousand articles of citizen's personal information shall constitute "serious circumstances" mentioned above.

reason for this lies in the fact that the nature of data is often highly reliant on its specific context. Data are not like a normal object. A normal object's basic character does not change depending on its context and is still protected by real or property rights regardless of context, even though it may manifest completely different characteristics in different contexts. One set of data could be classified into different categories of data depending on the context and the target audience. Take social media user data as an example, this sort of user data can undoubtedly be considered public data to the user's friends on the platform because the intention behind this sort of data is that it be disseminated among this group. In the case of platforms and third-party corporations, however, this sort of user data is protected by data privacy, because it includes a large amount of information that could be used to identify an individual. Besides this, in the case of third-party platform competitors, an accumulation of this data can be viewed as constituting a database, or as needing some sort of legal protection. Because this sort of data has an extremely high commercial value, platforms invest large amounts of capital and expend much labor in service of it.

Contextualizing Data Ownership

Contextual Protection of Data Rights and Interests

The issue of data ownership relies heavily on context. This means that protecting the rights and interests of individuals and corporations in regard to data requires a contextual protection approach. Determining the category and nature of data within specific contexts, and further determining the data rights and interests of related parties based on each party's reasonable expectations in that specific context is a better approach to resolving data ownership disputes.

Contextual protection of personal data has already been well received by many scholars within data privacy research. For instance, the celebrated scholar of privacy context theory, Professor Helen Nissenbaum, once indicated that the fundamental principle of data privacy protection lies in maintaining the contextual integrity of data (Nissenbaum, 2009, p. 127). In other words, the reasonable circulation of personal data and information in certain contexts must be achieved (Nissenbaum, 2009, p. 140). The reason Nissenbuam's theory is so influential is that its concept of "respect for context" became the guiding intellectual framework behind the "Consumer Privacy Bill of Rights" drafted during the Obama Administration. The most important reason for this is that this theory accords with the fundamental characteristics of personal data protection. Another example is the privacy categorization theory of Daniel J. Solove, an authority on data privacy. Solove borrowed from Wittgenstein's context principle to point out that privacy has no central or inherent characteristic and that therefore protecting privacy in practice lies in ensuring that individual rights and interests are not infringed in specific contexts (Solove, 2002, p. 1087). Besides this, Professor Ari Ezra Waldman once asserted that one cannot understand privacy in relation to personal information or data solely from the perspective of individual rights, as the heart of the issue of privacy lies in trust (Waldman,

2015, pp. 559, 560, 590), and therefore the boundaries of the right to privacy must be determined according to reasonable expectations within a specific context (Waldman, 2015, pp. 590-630).

In practice, personal data protection does in fact apply the contextual protection approach. In the US, there is little legislation addressing personal data protection at the federal level,[①] although the Federal Trade Commission (FTC) has gradually established relevant regulations through enforcement in specific cases (Solove, 2014, pp. 583, 585, 586). This method of protection is undoubtedly highly contextual, to the point where some scholars summarize this method as common law's protection model. In Europe, even though laws such as the GPDR have established many regulatory systems addressing data protection, these systems in reality overlap, and there are many areas where they contradict each other. Additionally, these regulations are often subject to the constraints of legal principles (Ding, 2018, pp. 39-53). Therefore, even if Europe adopted a unified legislative model, this legislation has yet to establish clear boundaries in regard to personal data protection. Personal data protection's future orientation in the EU will still largely be determined by the regulatory evolution based on specific contexts and cases.

Viewing personal data protection from a contextual perspective makes solving difficult issues relating to data ownership much easier. For example, take the cases of the Facebook-Cambridge Analytica scandal and the sale of Renren Network. Without a contextual approach, it is difficult to explain why Cambridge Analytica's use of certain user's personal data became a scandal and Renren Network's transfer of data control did not incite much controversy at all. When taking a contextual view of personal data protection, it becomes much easier to understand the differences between the two cases. In the case of Facebook and Cambridge Analytica, the use of certain personal data damaged the legitimate expectations of individuals, and this use occurred without the consent of those individuals, thereby violating user privacy. In contrast, Renren Network also did not obtain users' consent prior to its sale, but the sale of Renren Network in no way changed the context or expectations surrounding the use of that personal data. Thus, even if the sale changed who controls that data, it did not constitute a threat to data privacy. [②] As long as the purchasers of Renren Network take on the responsibility of protecting personal data and use platform data in the context of the individual's legitimate expectations, personal data privacy will still be reasonably protected (Balkin, 2016, p. 1183). [③]

The contextual protection approach is also more suited to the protection of corporate and platform data rights and interests. In judicial practice with respect to data crawling disputes, there have been differences in legal bases cited by Chinese and American authorities. Chinese courts more often utilized competition law to protect platform data, whereas courts in the US tended to appeal to

① The federal legislation of the US on data privacy has taken a sectoral approach, which is mainly concentrated in high data risk areas such as medicine, education, and fields involving minors.

② Of course, Renren Network would probably need to notify users of the change in ownership in order to guarantee that they are aware of their rights. Article 13 of the EU's "General Data Protection Regulation" stipulates that when personal data relating to a data subject is collected from the data subject, the controller must provide the data subject with "the identity and the contact details of the controller and, where applicable, of the controller's representative."

③ From the perspective of legal responsibilities, this sort of responsibility is more akin to a fiduciary responsibility than a contractual one.

trespass as existing in common law and other legislation. Both the US and China have similarities in the contextual nature of their protections. The protection mechanism of Chinese competition law emphasizes individual case judgments and reasoning through analogy, establishing guidelines based on the specifics of each case, and not seeking legal answers through a unified set of guidelines. Furthermore, the concept of trespass in the US law is also highly dependent on context. What can be considered trespass and what can be considered reasonable access and use is determined by the specific context and various different factors within the specific case.

In terms of legal theory, applying contextual protection to data ownership uses the rule of reason and not rule *per se* to make judgments on issues of data rights. In regard to legal disputes, traditional legal domains generally emphasize the importance of legal regulations and mainly establish the boundaries of different parties' rights through regulations and their exceptions. In the legal domains of competition law and anti-monopoly law, however, foreign laws tend to use the rule of reason to determine parties' rights within each case. This method is often used to determine the rights and obligations of various parties because of the context-specific nature of this category of disputes, which makes it difficult to rely on other less context reliant regulations. Because of the highly contextual nature of data, applying the rule of reason will undoubtedly be more beneficial to carrying out more reasonable protection of personal and corporate data. ①

Factors of Data Ownership Determination

In the judgments of actual cases, there are a series of factors that must be considered in determining the contextual ownership of platform data. Firstly, data privacy protection must be a factor for consideration of paramount importance. In circumstances where data privacy would cause substantial risk to the individual or damage the individual's legitimate expectations, the priority of data privacy protection relative to corporate rights and interests in regard to data should be maintained. This is because once personal data are not reasonably protected, not only will legal rights and interests of individuals be endangered, but corporations themselves will lose the confidence from users and consumers.②

Secondly, one must focus on facilitating data sharing and interconnectedness while respecting the premise of guaranteeing personal data privacy. The sharing and circulation of data in no way harm data, but actually make returns to scale from data's use, more likely, enabling "big data" to fulfill its potential and provide a stable foundation for the artificial intelligence industry. After all, the main characteristics of "big data" are "high-volume, high-velocity and/or high-variety information assets,"

① It needs to be remarked that the contextual determination of platform data ownership does not resolve the issue of data ownership entirely. As a legal model and concept, the question of ownership will be ever-present in any issue that involves property-like profit. As data's value becomes more apparent by the day, one can imagine that data ownership disputes will become ever more prominent and that parties involved in the use and collection of data will increasingly use the frameworks and concepts of ownership to advocate their positions. In this sense, I neither seek to nor am I able to completely resolve the question of data ownership itself, but rather reconstruct and consider approaches to deal with the question.

② Under the condition that the platform's use of personal data will not result in risk or damage to legitimate expectations, then it is not advantageous to put too many restrictions on the reasonable collection of personal data, and personal data rights of all kinds should be expanded.

and without the sharing and circulation of data the development of big data and artificial intelligence will be left without fertile ground in which to prosper.

Thirdly, factors such as the nature of the platform carrying out web-crawling operations and the platform being crawled as well as the nature of the crawling behavior itself should be considered when determining the boundaries of unfair competition and reasonable use. In terms of the party carrying out the crawling, when that party serves the public welfare or could be considered a public utility, that party should be allowed to engage in data crawling behaviors. For example, in the case of search engines, both the US and Europe have adopted a relatively open attitude towards search engine data. Even in cases where web-crawling affects information and data protected by copyright, it has generally been regarded that the web-crawling carried out by search engines is within the scope of reasonable use. Because search engines are akin to a public utility in nature, there is no doubt that the web-crawling they engage in is beneficial to the public dissemination and use of data. ①

In terms of the party that is the subject of web-crawling, one should consider both the dimensions and character of that platform's data. When the crawled party has a large amount of data that can be considered original data or fundamental data, reasonable use of such data by third parties should be permitted more often. This is because when an extremely large web platform amasses a large quantity of data, the possibility of the monopolization of this data emerges. If this sort of platform is granted data protections that are too strong, the result could be the emergence of some sort of "digital fragmentation" or "digital feudalism," in which it is difficult to ensure data sharing and inclusivity. In the case of databases, this sort of issue has already emerged. In the US and Europe, a few academic database giants have an essential monopoly over academic papers, and if one wants to read these papers, one must pay a large amount of fees (Lim, 2006). In China, the effective monopoly over academic content possessed by the Chinese National Knowledge Infrastructure (CNKI, a large Chinese periodical database) has attracted much criticism. If overall open availability and orderly circulation of data resources cannot be maintained, then small and medium sized enterprises will face problems of data barriers, and it will be difficult for effective competition to form in online spaces.

In terms of the nature of web-crawling, all other factors held the same, when a platform uses data obtained through crawling to engage in a similar commercial context with the platform that has been crawled, at this point, the crawling behavior should tend towards being deemed as an unfair competition. However, when the goal of the data crawling is to further process the data acquired or use them in a different context, the law should err on the side of recognizing the behavior as constituting reasonable use. This is because there is no creative use in the former behavior and no differentiated service has been provided to consumers. This sort of data crawling completely qualifies as "a free ride," in which it does nothing to facilitate a healthy competitive market environment. In contrast, although the second sort of data crawling behavior does have some "ride-hitching" elements,

① Of course, search engines' status as public utilities should entail that they take on more public responsibilities in many cases. For instance, they should provide the public with impartial information. Once search engine companies lose their public character and close themselves off, they should lose their right to reasonable data crawling and data use.

but considering the strong public nature of data as well as the creativity and differentiated service that arise as a result of the behavior. This instance should be considered as a reasonable use of data or should at the very least be more cautiously determined to be unfair competition.

Of course, when determining whether data crawling constitutes unfair competition or reasonable use, one must consider factors well beyond those described above. For instance, one must have a legitimate expectation that combines the common business practices of the given context and industry norms. These various and diverse factors involved in these determinations undoubtedly increase the difficulty while coming to judicial and legal judgments. However, from another perspective, the comprehensive analysis of these various factors will provide a more comprehensive and holistic judicial analysis of this issue and in the end, a more complete set of legal interpretations (Dworkin, 1986, pp. 176-275).

Conclusion

There are four major viewpoints in regard to the data ownership issues that arise as a result of data crawling: Platform data are owned solely by an individual, platform data belong solely to a platform itself, platform data belong to both the individual and the platform, and platform data are publicly owned. However, the research presented here indicate that analysis from both the perspective of legal clauses and the perspective of legal doctrines fails to support any of these four viewpoints. A consequentialist analysis also revealed that any of these allocations of data ownership is unreasonable.

The reason platform data's ownership is impossible to clearly delineate lies in the multiple characteristics of data, and their characteristics often rely on a specific context. In some contexts, platform data are the domain of the individual, and the protection provided by data privacy laws takes precedence, while in other contexts, platform data are akin to the nature of a database, and should be protected akin to the protection granted to a database, in still other contexts platform data have a public character and require that their circulation and shared use should be legally guaranteed.

The complex and highly contextual nature of data requires the establishment of contextual protections and assurances for data. Whether it be protection of individual rights to data or the legitimate protection of the data rights and interests of corporations, it is important to emphasize that reforms and legal progress occur in a bottom-up manner on a case-by-case basis and not to rely too heavily on a top-down regulatory framework. In terms of legal theories, this means that data rights should be determined by the rule of reason and not through some sort of universally applicable regulatory arrangements.

At the level of actual judgments, many different factors must be considered when determining the ownership of platform data. One must consider the priority protection of data privacy, the legitimate protection of the platform's rights and interests concerning data, and especially consider facilitating the circulation and sharing of data. One must consider avoiding "free-ride" behaviors in the area of data, as well as the public character of data. One must avoid unreasonable competition, data

monopolies, and data barriers. It is only if all of these elements are considered that the Internet can facilitate the reasonable circulation and protection of data, from which we will all benefit.

REFERENCES

"Baidu Baike's entry on the subject". Source: https://baike.baidu.com/item/ 顺丰菜鸟之争 /20830652?fr=Aladdin, accessed 9-11-2018.

"CNKI's Business, 970 Million RMB in Yearly Profits, a Monopoly Over Academic Information". Source: https://xw.qq.com/amphtml/20190214A10XCZ00, accessed 12-21-2018.

"Cyber Security Law of the People's Republic of China" Article 76.

"General Data Protection Regulation" Article 4-1.

"General Data Protection Regulation" Article 20.

"General Provisions of the Civil Law of the People's Republic of China" Article 8.

"hiQ Labs v. LinkedIn: Is scraping public data protected speech?". Source: http://jolt.law.harvard.edu/digest/hiq-labs-vlinkedin-is-scraping-public-data-protected-speech, accessed November 2, 2018.

"If Renren Network is sold, then where does all of its user data go?". Source: https://www.sohu.com/a/276750384_99955893, accessed 9-16-2018.

"The user agreement dispute: if you send a Weibo, then who really owns it?". Source: http://tech.ifeng.com/a/20170918/44688326_0.shtml, accessed 9-11-2018.

"To whom do data belong? Who really benefits from them?". Source: http://mini.eastday.com/mobile/190626235344267.html , accessed 6-29-2019.

"Who took user personal data in the Huawei-Tencent data dispute?". Source: http://tech.sina.com.cn/i/2017-08-15/docifyixtym4580079.shtml, accessed 9-3-2018.

"Whose user: Weibo wants to sue Jinri Toutiao for illegally seizing content". Source: http://finance.sina.com.cn/roll/2017-08-15/docifyixcaw4855222.shtml, accessed 9-1-2018.

"Without written permission users cannot authorize third-party use of Weibo content". Source: http://www.sohu.com/a/192382481_260616, accessed 9-11-2018.

Alan, W. (1967). *Privacy and Freedom*. New York: Atheneum, 7.

Ari, E. W. (2015). Privacy as trust: Sharing personal information in a networked world. *69 U. Miami L. Rev*, 559, 560, 590.

Ari, E. W. (2015). Privacy as trust: Sharing personal information in a networked world. *69 U. Miami L. Rev*, 590-630.

Beijing Intellectual Property Court Civil Judgment (2016) Jing No. 73 Min Zhong No. 588.

Bus. Elecs. Corp. v. Sharp Elecs. Corp., 485 U.S. 717, 726 (1988).

Craigslist INC. v. 3Taps, 942 F.Supp.2d 962 (2013).

Daniel, J. S. (2009). Conceptualizing privacy. *90 California Law Review*, 1087.

Daniel, J. S., & Woodrow, H. (2014). The FTC and the New Common Law of Privacy. *114 Colum. L. Rev,* 583, 585, 586.

Daryl, L. (2006). Regulating access to databases through antitrust law: A missing perspective in the database debate. *Stan. Tech. L. Rev*, 7.

David, F. T. (2011). Bottling the free flow of information: A comparative analysis of U.S. and EU database protection. *12 Wake Forest J. Bus. & Intell. Prop. L, 3*, 16.

Digital Labor: The Internet as playground and factory, Edited by Trebor Scholz, Routledge, 2012.

Ding, X. (2014). Exploring the legal standards of anti-discrimination and equal protection laws from the perspective of the disparate impact standard. *Peking University Law Journal*, 3, 1080-1096.

Ding, X. (2018). The fundamental principles of the right to be forgotten and its contextual regulation. *Tsinghua Journal of Law*, 6, 94-107.

Ding, X. (2018). The predicament of personal information private law protection and a way forward. *The Chinese Journal of Law*, 6, 194-206.

Ding, X. (2018). What are data rights? - Viewing data privacy protection from the perspective of Europe's General Data Protection Regulation. *Journal of East China University of Political Science and Law*, 39-53.

Ding, X. (2018). What are data rights?- Viewing data privacy protection from the perspective of Europe's General Data Protection Regulation. *Journal of East China University of Political Science and Law*, 4, 45-50.

Ding, X. (2019). On the origin of personal information legal protections and their fundamental principles: Analysis based on "Fair Information Application". *Modern Law Science*, 3, 96-110.

Ding, X. (2019). User portrayal, individualized recommendations, and personal information protection. *Global Law Review*, 5.

Ding, X. On the nature of the right to data accessibility, its impact and Chinese Application.

Directive 96/9/EC, of the European Parliament and of the Council of March 11, 1996 on the Legal Protection of Databases, 1996 O.J. (L 77) 20, chap 3.

Directive 96/9/EC., at chap 3, para 7.

Du, Z., & Cha, H. (2018). An actual consideration of the data property rights system. *Chongqing Social Sciences*, 8, 19-25.

Ebay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058(2000).

"Facebook-Cambridge Analytica: A Timeline of the Hijacking Scandal". Source: https://www.cnbc.com/2018/04/10/facebookcambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html, accessed December 3, 2018.

Fan, C. (2015). Industry norms and unfair competition. *The Jurist, 5*, 84-94.

Fang, X. "BAT is beginning to be the antithesis of the spirit of the Internet". Source: https://www.guancha.cn/FangXingDong/2019_01_30_488663. shtml, accessed 12-1-2018.

Fei, F. et al. (2018). Data's nature, its associated property rights, and data competition in the digital age. *Research on Financial and Economic*, 2, 3-21.

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 347 (1991).

Fred, H.C. (1997). *Privacy in the Information Age*. Brookings Institution Press, 370-373.

Gartner IT glossary Big data, http://www.gartner.com/it-glossary/big-data, accessed December 3, 2018.

Helen, N. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 127.

Helen, N. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 140.

Henry, H., & Reinier, K.(2002). Property, contract, and verification: The numerus clausus problem and the divisibility of rights. *31 J. Legal Stud*. 373, 368-387.

HiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783 (2017), paragraph 6.

Hu, L. (2017). Information/Data property rights in the context of business models. *Journal of Shanghai University (Social Sciences Edition)*, 6, 1-14.

Leon v. Pac. Tel. & Tel. Co., 91 F.2d 484 (9th Cir. 1937); Jeweler's Circular Publ'g Co. v. Keystone Publ'g Co., 281 F. 83 (2d Cir. 1922)

Liu, J. (2018). On the Legal Protection Approaches of Competition Law in regard to User Data. *Economic Theory and Application*, 3, 26-30.

Locke (1982). *Two Treatises of Government (Second Treatise)*. In Ye, Qifang & Qu, Junong (Trans.). The Commercial Press, 20-25.

Long, W. (2017). Research on constructing new property rights surrounding data and their systems. *Politics and Law Forum*, 4, 75.

Long, W. (2018). Reexamining the property rights path to corporate data protection. *Eastern Legal Studies*, 3, 50.

Mark, A. L., & David, M. (1998). Legal implications of network economic effects, *86 CAL. L. REV*, 479.

Michael, L. K., & Carl, S. (1985). Network externalities, competition, and compatibility, *75 AM. ECON. REV*, 424.

Ning, L., & Wang, D. (2016). The determination of "Robots Exclusion Protocol" and an analysis of their role in competition law. *Jiangxi Social Sciences*, 1, 161-168.

Orin, S. K. (2016). Norms of computer trespass. *116 Colum. L. Rev*, 1143, 1154, 1161, 1162, 1163.

Paul, S., & Dan, S. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *86 N.Y.U. L.Q. Rev.* 1814, 1815.

ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1449-50 (7th Cir. 1996); Specht v. Netscape Commc'ns Corp., 150 F. Supp. 2d 585, 593 94 (S.D.N.Y. 2001), qff'd, 306 F.3d 17 (2d Cir. 2002).

Ronald, D. (1986). *Law's Empire,* Belknap Press of Harvard University Press, 176-275.

The Authors Guild Inc., et al. v. Google, Inc., 755 F.3d 87 (2d Cir. 2014)

Unfair Competition case of DianPing against Baidu Shanghai Intellectual Property Court Civil Judgment (2016) Hu No. 73 Min Zhong No. 242.

Unfair Competition case of DianPing against Baidu. Shanghai Intellectual Property Court Civil Judgment (2016) Hu No. 73 Min Zhong HiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783 (2017).

Unfair Competition Case of Taobao (China) Software Co., Ltd. vs Anhui Meijing Information Technology Co., Ltd. Hangzhou Railway Transport Court Civil Judgment (2017) Zhe No. 8601 Min Chu No. 4034.

Unfair Competition Dispute between Beijing Baidu Netcom Science Technology Co., Ltd. and Shanghai Hantao Information Consulting Co., Ltd. Shanghai Intellectual Property Court Civil Judgment (2016) Hu No. 73 Min Zhong No. 242.

Unfair Competition Dispute between Beijing TaouTianxia Technology Development Co., Ltd., et al. and Beijing WeiMeng Technology Co., Ltd. Beijing Intellectual Property Court Civil Judgment (2016) Jing No. 73 Min Zhong No. 588.

Vault Corp. v. Quaid, Inc. 847 F.2d 255, 269 70 (5th Cir. 1988); Arizona Retail Sys. v. Software Link, Inc., 831 F. Supp. 759, 766 (D. Ariz. 1993).

Wang, Z. (2015). Research on Personal Data Trade Permission Mechanisms in the Age of Big Data. *Theory Monthly*, 6, 131-135.

Xu, S. (2018). The intellectual property route to corporate data protection and its breakthrough. *Eastern Legal Studies*, 5, 56.

Yang, H. & Qu, S. (2013). On the legal character of robots exclusion protocols. *National Colleges Law Journal*, 4, 30-34.

Yang, H. (2014). The impact of robots exclusion protocol on Internet-based competitive relationships. *Intellectual Property*, 1, 12-21.

Yao, J. (2019). Corporate data use standards. *Tsinghua Journal of Law*, 3, 114-125.

Yuval, F. (2006). The behavioral foundations of trade secrets: Tangibility, authorship, and legality. *3 J. Empirical Leg. Stud*, 197.

Zhang, P. (2013). The general clauses of law against unfair competition and their application-thoughts provoked by Search Engine Web-Crawling Protocols. *National Colleges Law Journal*, 3, 46-51.

(Translator: Ryan; Editor: Zeng Yueying)

This paper has been translated and reprinted from *ECUPL Journal*, No. 5, 2019, pp. 69–83.