

THE ECONOMIC CHARACTERISTICS OF DATA COMPETITION IN CHINA'S DIGITAL ECONOMY



BY VANESSA YANHUA ZHANG & JOHN JIONG GONG¹



¹ Vanessa Yanhua Zhang, Managing Director at Global Economics Group (Beijing and New York) and Senior Researcher at MRLC of Renmin University (Beijing, China). John Jiong Gong, Professor of Economics at the University of International Business and Economics (Beijing) and Director at Global Economics Group (Beijing, China). We would like to thank Amanda Jing Yang for research assistance and Wenlian Ding, Sam Sadden for helpful comments. None of the institutions above necessarily shares the views expressed in this article and we retain sole responsibility for any errors.

I. INTRODUCTION

The importance of data in the digital era cannot be more emphasized. There have been ample discussions of the role of data, or more popularly, big data, in future economies. Data is oil, data is gold, as gurus have been saying lately. As technologies of generating, storing, processing and using data continue to move forward, their economic and legal implications start to emerge and challenge our traditional ways of dealing with data related issues. In particular, issues pertaining to data ownership and property rights, privacy protection and allocation of value from data among stakeholders have become increasingly controversial, as manifested in more litigation and regulatory oversight decisions.

The dispute over data related issues clearly has competition policy and antitrust implications. In China, we have already seen several cases centered around data-related competition. In this paper, we discuss three recent and representative cases of data competition and provide our own analysis under a framework that takes into account what we describe as the seven main characteristics of data competition: multi-homing, non-rivalry, private value, usage difference, unpredictability, externality and reciprocity. Data is a unique good, unlike many physical goods.

An analysis of these cases provides insight into how courts in China tend to treat data-related competition issues. It may also shed some light on the direction of regulation and potential legislation over data competition in China.

II. SEVEN CHARACTERISTICS OF DATA COMPETITION

The digital economy endows data specific features and characteristics which might not be common in the industrial era. Compared with the traditional physical resources, data resources, in our opinion, have the following seven characteristics from an economics perspective.

A. Multi-Homing

In the traditional industrial economy, information is usually controlled by one entity. In a digital economy, however, the structure of the value chains has evolved from linear mode to multilateral networks. Data generation accordingly has become more complicated, manifesting in a rich set of business models. Often times, data is not always generated and hosted by one party, but rather by multiple parties, which we describe as “multi-homing.”² Take a take-out app “Meituan” as an example, end users, the app developer, delivery companies, and the restaurants are all involved in the generation of data of some sort, with each party having part or all of the information pertaining to a transaction. Additionally, when more and more users use digital products and services, ranging from various information-oriented websites to mobile apps to search, social network, entertain and communicate with each other, users’ data information is generated and hosted by several parties involved in these activities. Therefore, unlike the traditional physical resources or information prior to the digital economy, data generating becomes a multi-homing phenomenon.

B. Non-Rivalry

In economics, a good is said to be rivalrous if its consumption by one consumer prevents simultaneous consumption by other consumers, or if consumption by one party reduces utility/ability to another. A good is considered non-rivalrous if, for any level of production, the cost of providing it to a marginal individual or firm is almost zero. In a way, data usage is like watching satellite television or an exhibition of art work; it does not entail additional cost for the data owner when multiple parties try to use it. That means it strides on the boundary between private goods and public goods, as one data operator’s using data information virtually allows another operator to use the same data information simultaneously without incurring additional costs. Therefore from a social welfare perspective, data should be physically shared by multiple parties at multiple times so as to generate greater economic and social benefits.

² “Multi-homing” comes from the “two-sided market” or “multi-sided platform” theory, meaning that users on one side of the platform often participate on another platform. For more information of the two-sided market theory, see Rochet & Tirole, “*Platform Competition in Two-Sided Markets*,” 1 J. EUR. ECON. ASS’N 990 (2003). For more articles about “multi-homing,” see “*Innovation, Competition and Platform Economy: Collected Paper of Jean Tirole, the Winner of Nobel Prize in Economics Sciences*,” ed. by Kou & Zhang, Law Press China (2017).

C. Private Value

Although data is generated in a multi-homing environment, its nature and the associated value could be different things to different parties. In other words, data has private value characteristics. Recall the previous example of a take-out app “Meituan.” The ordering app arguably has the most comprehensive information of the transaction, covering both the supply and demand side. It not only possesses the information about the end user who places an order and the selected restaurant, but also knows more about the dynamics in a broader market context, such as the end user’s search preferences and the restaurant’s products and prices. Therefore, the ordering app can optimize matching results and promote targeted advertisements for both the supply and demand side. The delivery companies, however, obtain static sales information after the order is generated, and the amount of information is less rich than what was stored on the app platform. Thus, delivery companies seldom take the initiative to promote targeted advertising for both market sides.

D. Usage Difference

Even for the same data information, different users apply different methods to collect, develop and use it in order to obtain different analytical results and cover different service categories. Still use the previous take-out order example. The ordering app platform connects the end users and the restaurants to find the most suitable transaction pattern for matching each other and completing the transaction. Therefore, this economic activity occurs with the goal of completing sales. But for the delivery company, its objective of collecting and using this piece of information is to optimize delivery routes and the logistics arrangements so that delivery will be in a timely manner. This same data has distinct values for different users, and there are noticeable differences in the way it is used by multiple parties.

E. Unpredictability

The use of the data information relies not only on the prior experiences of the user, but also on future development of technologies, which might be difficult to predict. The future usage of the data may not be completely identified and defined with existing methods. For example, in 2009, Google successfully predicted when and where the H1N1 influenza outbreak would occur in the United States after analyzing the search information of billions of Internet users, which was never imagined only a few years ago.³ As another example, cell phone location information is widely used in a variety of location-based services which were not envisioned when location-based services were first introduced. The development and utilization of big data depends not only on current real-time data but also on the accumulation of historical data in human society. When data information is collected and processed, the user may not be fully aware of how the data would be used in any specific way in the future, but it is still necessary to collect and process the data today.

F. Externality

After the data information is developed and utilized, its economic and social effect often exceeds the scope of the users who originally generated the data, and extends to affect a larger group of users or even to a wide scope of social members. In other words, it has a strong externality effect. Considering the take-out app as an example, after analyzing the search preferences of the end consumers, the app platform may recommend other restaurants visited by other consumers of the restaurant they searched. Similarly, for a certain type of restaurant, the app platform may also recommend the restaurant to other consumers with similar preferences according to the search preferences of the end consumers. Through data optimization and analytics, the app platform improves the accuracy of the marketing conversion rate and increases the possibility of successful matching.

G. Reciprocity

When users disclose their ID on the Internet, data flow will generally be bilateral rather than one way. Search engines, for example, have evolved from web-based service to location-based service. A decade ago, when we used Google or Baidu to search one restaurant near a specific location, it would pop up a list of restaurant names on both organic search results and paid-search results. Nowadays, when users preset the location information and personal data in the default settings that search engines often ask the users to choose, search results become bilateral rather than the “one-way” and “on-call” model. Search engines change their role from passively receiving search orders from the users to

3 Mayer-Schönberger & Cukier, “*Big Data: A Revolution That Will Transform How We Live, Work, and Think*,” Houghton Mifflin Harcourt, 2013.

proactively posting search results to the users, especially on the location-based apps and websites. Personal data has been used to optimize the best consumer experience and provide efficient and effective choices to the consumers, sometime discovering the demand that consumers might not recognize themselves. Sometimes, however, such optimization might also bring hassles to the consumers who might be overloaded with information.

III. CASES OF DATA COMPETITION IN CHINA

The seven economic characteristics of data help us better understand the nature and nuance of competition in the digital economy in China. Recently, with the development of the Internet industry and mobile communication, disputes and conflicts among Internet companies and their ancillary service providers have dramatically increased. The courts and industry regulators have to deal with the complicated issues even before the theories are available. In this section, we will discuss three representative cases and shed some light on the characteristics we described above.

A. *Sina v. Maimai*

Sina Weibo founded by Weimeng Innovation Technology Inc. in Beijing in 2010 is a social networking service (“SNS”) provider which is famous for its “Weibo” (Chinese “Twitter”). Maimai owned by Taoyou Technology which was founded in 2013 is a competing SNS service. In August 2014, Sina Weibo found that many non-Maimai users appeared on Maimai’s portal with their Weibo personal information such as photo, name, profession and education background, etc. Sina Weibo sued Maimai for misappropriating information about Weibo’s users. Although users could log onto Maimai using their Weibo account, the non-Maimai users’ information was apparently imported from Weibo’s portal without Weibo’s or the users’ permission. Meanwhile, during Maimai registration, users were required to upload their contact lists in their mobile phones, through which Maimai displayed the names and images of cell phone contacts with Weibo accounts and exposed information about the user’s profession and educational background without their acknowledgement. In April 2016, Beijing Haidian District People’s Court ruled that Maimai’s conduct constitutes unfair competition and that Maimai must cease such conduct immediately and compensate Sina Weibo CNY 2 million (U.S. \$309,000) for economic damages.⁴ Maimai appealed to the Beijing Intellectual Property Court. On December 30, 2016, the Beijing Intellectual Property Court dismissed the appeal and upheld the original ruling.⁵

The most critical dispute in the case is the use of user data. Through user registration, Weibo obtained large numbers of highly sensitive professional and educational information. The Beijing IP Court (“the Court”) found that Maimai’s actions constituted unfair competition because it obtained the information of non-Maimai users through their Weibo accounts without Weibo’s approval, which Maimai would otherwise have had to pay for itself. Such conduct also reduced Weibo’s competitive advantage. In the judgement, the Court recommended the management obligations that the network operators should perform in the acquisition of user data. In addition, during data processing, the Internet Service Provider (“ISP”) should pay special attention to protect the user’s right to know and the “Triple Authorization” principle, i.e. “User authorization” – “Platform Authorization” – “User Authorization,” was proposed by the Court. The Court suggested that in collecting user information ISP should follow the principles of legality, propriety and necessity. From the perspective of consumer protection, “propriety” means that data collection requires the consent of user. “Necessity” means minimum sufficiency; or in other words, a platform should not collect unnecessary information. Last but not the least, the data collection process that Weibo used is considered as legitimate since Weibo had obtained users’ permission and assembled the data via its business operation. Meanwhile, such information is not sensitive user information and not qualified as user’s privacy. Therefore, the Court clarified that a platform could, upon users’ consent, claim collection and usage rights on commercially valuable information of its users through its business operation.

⁴ First Instance Civil Judgment of Beijing Haidian District People’s Court, (2015) Hai Min (Zhi) Chu Zi No. 12602, April 26, 2016, available at: <http://www.bjcourt.gov.cn/cpws/paperView.htm?id=100526100768&n=1>.

⁵ Second Instance Civil Judgment of Beijing Intellectual Property Court, (2016) Jing 73 Min Zhong No. 588, December 30, 2016, available at: <http://www.bjcourt.gov.cn/cpws/paperView.htm?id=100526087782&n=1>.

B. Hantao (Dianping.com) v. Baidu

Hantao operates Dianping.com, a local information platform providing customer reviews of local services such as restaurants, hotels, hair salons, etc. and covering restaurant reservations, take-out delivery, group discounts and other O2O (Online to Offline) services. Dianping works mostly like Yelp + OpenTable + Groupon. Baidu is the most popular search engine in China. Besides search engine services, Baidu also offers key features such as online map Baidu Map, Baidu Zhidao (Baidu Knows or Baidu Q&A) and Nuomi (take-out delivery). Baidu was accused of having scraped customer reviews from Dianping.com, and, without the permission of Hantao, displayed these reviews on Baidu Map and Baidu Zhidao. On April 28, 2014, Hantao sued Baidu for unfair competition at the Shanghai First Intermediate People's Court. The court of first instance ruled in favor of Hantao on May 26, 2016 by requesting Baidu to cease the unfair competitive conduct of using customer reviewers on Dianping.com and to compensate Dianping.com RMB 3,230,000 (approximately U.S. \$492,739).⁶ Baidu appealed the decision before the Shanghai Intellectual Property Court ("the Court") but the Court dismissed the appeal and upheld the rulings of the court of first instance on August 30, 2017.⁷

The main issue in this case is whether the use of information collected by other entities is legitimate under the context of Internet competition.⁸ The Court acknowledged Hantao's (Dianping.com) legitimate business interests in the customer review information and held that Baidu in fact competed with Dianping.com in providing the review service on its Baidu Map, which resulted in an infringement of Hantao's interests. The Court stated that the amount of information used by Baidu via integrating customer reviews from Dianping.com into its online mapping feature was beyond what was necessary. Such conduct, confirmed by the Court, would eventually reduce the firms' incentive to invest in data collection, thereby harming consumer welfare in the long run. Taking these factors into account, the Court held that Baidu's conduct violated generally recognized business ethics and therefore dismissed Baidu's appeal.

This case raises an important issue of ownership and usage rights of data information. In the past, most "scraping" cases in the Internet world brought under Article 2 of the Anti-Unfair Competition Law ("AUCL") focused on the appropriation of content "owned" or "created" by competitors. In this case, the "scraped" content was not "owned" by the plaintiff Dianping.com, but was produced by the users of Dianping.com. This position seems largely consistent with the Trial Guidelines on Network Related Intellectual Property Right Cases, issued by the Beijing High People's Court in April 2016.⁹ According to these guidelines, the courts in Beijing may hold the unauthorized use of information from a website to be unfair competition under Article 2 of the AUCL if: (1) the information can increase business opportunities and competitive advantages for the plaintiff; and (2) the use of the information provides users with an effective alternative to the website where the information is taken from. To rule on the "scraping" issue, a key factor in the Court's analysis was that Dianping.com had made significant efforts in the collection and use of the original user's information, even if the owners of the reviews were individual consumers. Although the Court's finding may have been fact- and case-specific, the judgment contains some upbeat language on the positive effects and efficiency of producing, collecting and using consumer-related information by Internet players. This stands in contrast to some of the developments in Europe where the handling of consumer-related information is at times viewed skeptically – not only from a privacy, but also increasingly from an antitrust, perspective.

C. Dispute between Cainiao and SF Express

SF Express is a leading express logistics service provider in China. Established in 1993, SF Express provides integrated logistics solutions. Besides high-quality logistics services, SF Express has expanded its service to other segments of the industry chain including manufacture, supply, sales and distribution, and makes use of big data analysis and cloud computing technologies to provide logistics services by integrating storage management, sales forecast, big data analysis and settlement management, etc. SF Express had its IPO on the Shenzhen Stock Exchange in February 2017. Hive Box was established in April 2015 by SF Express, STO Express, ZTO Express, Yunda Express and GLP, which provides express platforms to send and receive express products through its delivery lockers. Hive Box aims to provide solutions for "the last

6 Civil Judgment of Shanghai Pudong New Area People's Court, (2015) Pu Min San (Zhi) Chu Zi No. 528, May 26, 2016, available at: http://www.pkulaw.cn/Case/payz_125774070.html?match=Exact.

7 Civil Judgment of Shanghai Intellectual Property Court, (2016) Hu 73 Min Zhong No. 242, August 30, 2017, available at: http://www.sohu.com/a/169734455_221481.

8 Jingbo Fan, Dispute between Hantao and Baidu regarding Unfair Competition ("汉涛公司诉百度公司等不正当竞争纠纷案"), Case Comments of ChinaCourt.org, November 16, 2017. <http://www.chinacourt.org/article/detail/2017/11/id/3076212.shtml#6645367-t sina-1-84383-8146f288d6e862c69dbee60f4449c31d>.

9 "Trial Guidelines on Network Related Intellectual Property Right Cases," Beijing High People's Court, effective as of April 13, 2016, available at: http://china.findlaw.cn/fagui/p_1/385990.html.

mile” in the logistics industry. By the end of 2016, Hive Box has had more than 35,150 cabinets in more than 75 major cities in China, and had established strategic cooperation with more than 100 well-known property companies in the country. Cainiao is an express delivery platform established by Alibaba and Yintai Group in 2013. Cainiao aims to establish a data-driven and social-collaborated platform for logistics and supply chains. It wants to establish an open, sharing and social platform for logistics infrastructures, so that any delivery to any places in China may be accomplished within 24 hours. By November 2015, more than 70 percent of the express packages, more than 1,000 logistics and storage companies, and more than 1.7 million logistics staff in China were running through the Cainiao network. In 2016, Hive Box started the cooperation with Cainiao. Cainiao provided the consumers’ cellphone numbers to Hive Box, and Hive Box provided Cainiao the corresponding status of the express products at the cabinets. However, in March and April 2017 when both parties negotiated to renew the cooperation agreement, Cainiao requested to use Cainiao Guo Guo (“Cainiao Parcel,” an app providing express delivery status information) for all the cabinets of Hive Box, and was rejected by Hive Box due to user privacy issues.

On June 1, 2017, SF Express claimed that Alibaba’s Cainiao unilaterally disconnected Hive Box’s data interface. While, Cainiao, however, argued that SF Express would suspend its data interface. State Post Bureau (“SPB”) got involved on June 1, 2017 and it was concerned about the dispute. In the evening of June 2, SPB gathered executives from both companies to communicate about the interface issues, and both parties agreed to fully recover the business cooperation and data exchange on June 3.

Both SF Express and Cainiao emphasized that they were fighting for the information security of the users. The truth is they were competing for big data, which is deemed to be the main resource in the future. The logistics information is the core competitiveness of SF Express, which it cannot disclose to another third-party platform without restrictions. Cainiao has a lot of ecological logistics data, while SF Express has the closed-loop logistics data. However, in the big data era, the key issue is not to access some of the data but to control the data. How to control and make good use of the big data to build the ecological logistics system and what kind of data can or cannot be shared to the others are important issues that both parties will strive to overcome.

IV. CONCLUSION AND POLICY IMPLICATIONS

Data usage rights and control rights are the core issue of the disputes and conflicts among competitors. In China, although companies may have their competitive niches, there are more and more overlaps of their business operations when Internet companies expand their territories. Companies have blurred their traditional boundaries and competition, no matter explicit or implicit, may lead to head-to-head data “wars” when both companies eventually seek out core data to maintain their competitive advantages. Given that China is the country with the largest number of Internet users, policy makers should take into account the competitive nature of these matters when dealing with disputes involving data competition.

First, market definition would be an important step to help zooming in on the competitive nature of the conduct at issue. When analyzing unfair competition or anti-monopoly conduct, one might start with the conduct at issue. Although dynamic competition and platform competition make the market boundaries blurry, people could still catch cross-point where both companies compete to increase their comparative advantages. In the *Hantao (Dianping.com) v. Baidu* case, it’s not so obvious how a search engine competes with an online review/reservation platform. Users’ review data would help the court focus on the competition of review information on both Baidu Map and Dianping.com which, in general, both provide location-based consumer review services.

Second, data usage rights and control rights take more weight in the evaluation of legitimacy of competitive conduct. Although data ownership is important for the source of legitimacy, most of the time, if not all the time, data is owned by users or directly generated by users’ behavior in various economic activities, such as social networking, e-commerce, searching, reviewing, etc. Individual users have no intention or capability to analyze or conduct research and development on their own data although it might seem important to them. Those entities which collect the data have more incentive or capability of making further valuable use of the data for either the entities’ own interests, or for increasing all users’ welfare. Therefore, discussion of ownership might be the starting point but might not be the end point. The authority should further investigate the economic effect and/or efficiency of data usage rights and control rights which take the most important role. They may need to solve questions such as whether competition for the data usage and control rights would reduce other firms’ incentives to invest in data collection activities, whether such competition would bring efficiency gains, whether a firm’s data collection conduct is free riding by nature, or whether interoperability and sharing of data is social welfare enhancing. Answers to all these questions should be based upon diligent theoretical and empirical analysis of the cases and should be case-by-case.

Third, when dispute and conflicts cannot be solved through government intervention, authorities might have to carefully consider whether a consent decree would be a good solution. In China, most indigenous innovations in traditional industries are top down. That said, the government then takes the lead and implements those innovations via its national innovation policy. Internet industry is an exception where the majority of innovations take place through a bottom up mechanism. Data accumulated from the business activities of the participants of the Internet ecosystem is important, but should not be considered as an essential facility. As we discussed above, data might be hosted by different hosting carriers. There might be certain hosting carriers who have more data than others, which might give the authority the temptation to issue a consent decree to request the data advantaged entities to share the data information with other data disadvantaged entities. Before doing that, the authority needs to carefully evaluate whether data belongs to the common good and whether government regulation should be involved in resource allocation or what other mechanisms might be a better solution than a consent decree.¹⁰

¹⁰ Tirole, *Economics for the Common Good*, Princeton University Press (2017).

