O'REILLY®

Private and Open Data in Asia

A Regional Guide

Franklin Lu

HOLE #8
PAR 3
5
WINNING SCORE
SAN JOSE
25 FINSBURY CIRCUS
London, UK EC2M 7EE
LONDON
floor trader 788
TRADING LICENSE #34096789
€ 9322.50
acct JN44K03C
NEW YORK
SINGAPORE

Private and Open Data in Asia: A Regional Guide

Franklin Lu

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

Private and Open Data in Asia: A Regional Guide
by Franklin Lu

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

Editor: Tim McGovern
Production Editor: Nicole Shelby
Copyeditor: Jasmine Kwityn

Interior Designer: David Futato
Cover Designer: Randy Comer
Illustrator: Rebecca Demarest

October 2015: First Edition

Revision History for the First Edition
2015-10-12: First Release

While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-491-93588-0

[LSI]

Table of Contents

Overview: Why Asia?

The rise of big data—high volume, high velocity, and high variety data—in recent years coincides with the economic and political rise of Asia. As Asia continues to expand economically, it becomes an important market for big data. Business models relying on the collection, manipulation, enhancement, sale, or use of data—and it is rapidly becoming apparent that all businesses benefit from being more data driven—must pursue the treasure trove that is the East. Asia already dominates the world in terms of Internet access (nearly half of the world's entire population of Internet users, around 45%, reside in Asia). South Korea and Japan are highly developed countries, with high Internet penetration rates (roughly the same as the United States and Europe, sitting at 80%+). More importantly, China, India, and Indonesia have enormous populations, but relatively low Internet penetration (46%, 24%, 16%, respectively). While these three countries already have massive Internet-using populations that will provide both data and the market for data, they will also continue to grow as their national Internet ecosystems mature. And with economic prosperity, Internet penetration will increase, and so too will the usage of smartphones, social media, and ecommerce. In addition, with the rise of smartphones, many of these countries have skipped the personal computer age, going directly to mobile. Not only are Asian Internet users multiplying, they are also attached to technology in a way that allows for big data to flourish, accessing the Internet through apps and hardware that more easily allow for the collection of more metadata than browsers. Collectively, five countries—China, Japan, Korea, India, Indonesia—make up the bulk of the East Asian Internet-using population.

Contextualizing all this personal data is open data: big data sets open to the public for use. Open data from fields such as healthcare, education, agriculture, transportation, energy, and finance offer opportunities to build businesses and services. Open data's availability varies from country to country. Getting to this data can be difficult based on cultural barriers, government restrictions, privacy policies, and/or the lack of databases (or their inaccessibility, whether they're locked up in filing cabinets or "locked" in PDFs or unreadable legacy file formats).

The decision to enter the Asian market, as a data-driven business or a data-focused one, is fraught with questions—while the business of big data is lucrative is Asia, is it more lucrative than business in the United States? Do the benefits outweigh the costs, namely a new market to adapt to, a new culture to understand, and a new government to work with (or around)? This question is complex and not easily answered, however, all companies seeking to do business in these countries should know the surrounding legal environment as a first step. What are data privacy laws like? What businesses already exist? What open data initiatives are there? This report will offer an overview of the current state of big data and open data in these large, Internet-using, Asian countries.

China

The largest and most prominent of Asian countries is by far China. With its massive economic influence, strong central government, and huge Internet-using population, China represents a unique but massive market for big data–related business. While big data flourishes, however, open data struggles.

China currently lacks any legislation that specifically addresses the issue of data privacy and data protection. However, the General Principles of Civil Law and the Tort Liability Law are general laws that may be interpreted to include data privacy rights as part of an individual's right to privacy. The extent to which data privacy is protected under these general laws is up to interpretation. There is evidence that China is seeking to tighten its policy on the matter of data privacy with, for example, the arrest and deportation of Peter Humphrey, who mined data for GlaxoSmithKline. In cases such as these, China's government has demonstrated that it will interpret current laws to include data privacy breaches as infringements. As China continues its explosive growth, especially in the realm of ecommerce and social media, the need for data privacy guidance will only increase. In 2013, China issued "Information Technology Security—Guidelines for Personal Information Protection Within Information Systems for Public and Commercial Services." The Guidelines define the state's expectations for data privacy and protection. In both content and legal standing, they are similar to the US Fair Information Practice Principles. They are not legally binding, but they do set the tone for the preferred practices for businesses dealing with personal information in China. Individuals from whom data is collected are to be informed of the retention period of the data, the purpose of the data collection, the method of data collection, and the scope of the data security. Data is to be processed in a manner consistent with the announced purpose and method, and is to be deleted after the retention period is up. The "Guidelines" emphasize the fact that China is in fact moving forward in terms of its data privacy and protection laws. Although they lack the full force of law, the Guidelines set the tone for future legislation coming out of China.

Beijing's official legislation regarding data privacy is only part of the landscape for big data in China. Three large companies dominate big data currently in the world's fastest growing market. Baidu, Alibaba, and Tencent, collectively known as BAT, are familiar to those already involved in business in China but a brief introduction for the foreign audience is in order: BAT comprises the three biggest players in China's Internet industry. Baidu is a search engine first and foremost, and therefore collects data based on user searches. Alibaba, an ecommerce giant, has access to valuable market data—the purchasing habits and preferences of consumers. Finally, Tencent is primarily know for being the creator of WeChat, the largest messaging app in the world (measured by monthly active users). It comes as no surprise that all three companies are attempting to put their wealth of data to use. Baidu has already begun delving into deep learning and data-crunching technologies. The search giant has used big data to do everything from modeling disease patterns to predicting the winner of the World Cup. Baidu leads the charge for the big data revolution in China, investing in R&D with numerous big data and deep learning labs, located in both the United States and China. Similarly, Alibaba has also utilized big data to streamline its ecommerce in terms of helping sellers understand the targeted buyers, and customizing consumer recommendations. Alibaba also maintains a cloud computing subsidiary, Aliyun. Aliyun is noteworthy for having issued a Data Protection Pact, which guarantees that Alibaba will protect consumer and business data privacy.

Although Beijing's official legislation is not necessarily strict regarding data privacy, companies such as Alibaba are taking the initiative to guarantee customers that their data is secure. Tencent lags behind the others in terms of technology—the company is not quite as invested as Baidu is in the realm of deep learning—yet it still employs big data, for example, in targeting customers with advertisements.

China's data privacy policies and the companies that dominate the Chinese Internet industry may not appear too different from those of the United States. However, several stark contrasts exist. Primarily, the Chinese industry operates under the shelter of the Great Firewall, and under the shadow of the Chinese government. Google, for example, has had a difficult time in China—from the fight over censorship to security breaches. It is not surprising, therefore, that Baidu takes 80% of the Internet traffic in China, with Alibaba and Tencent occupying the roles that Amazon and Facebook occupy elsewhere. BAT seeks to expand into one another's territories (for example, Tencent partnering with China's second largest ecommerce website, JD.com), as well as expanding into newer fields where big data can be used in different ways (for example, in finance or health care), allowing more business opportunity.

In many ways, the political economy of China encourages disruption-based models: large, internationally successful businesses might have a hard time porting over into China due to government oversight and involvement and different culture, but smaller, more flexible companies might be able to establish niche positions and disrupt major players before becoming bogged down in the current system.

Finally, it might go without saying, but *culture matters.*When targeted with ads within WeChat, where wealthier users supposedly received a BMW ad, while a "lower class" ad for Coca-Cola was shown to other users, those receiving Coca-Cola ads complained and expressed the desire to receive BMW ads. This incident is amusing, but also illustrative of the ways that the Chinese people accept that targeted advertisements exist, based on the data that they shared with WeChat, but view ads as status markers rather than simply annoyances to be ignored. A majority of Americans, on the other hand, express a disapproval for them.

Despite China's fascination with big data, the quest for open data remains at large. China's government has never been about transparency, and big businesses dominate the data marketplace. A few cities, including Shanghai and Beijing, have individual sites where open data sets are available. However, the sets are by no means extensive, and their launches were hardly publicized. Even for these cities, whether or not the data should be completely free and available is still debated. Nationally, there is no open data initiative to speak of. As Joel Gurin of the Open Data Institute has said, "Unlike the U.S. and other countries where national governments have taken the lead by establishing clear open data policies, it is citizens, non-profits, and urban government leaders driving the movement for more data in China." The creation of Open Data China is the most visible start to this movement.

China is definitely a country to watch. Its explosive economic growth coupled with the experimentation of open data on a municipal level, which could turn into national open data initiatives, may turn China into an open data goldmine in the coming years. Indeed, the potential of the Web to transform politics from the ground up on an administrative level is being revealed there.

Japan

Since 2010, Japan's Internet penetration has been hovering at around 80%, roughly the same level as the United States. While this means a large portion of a wealthy population has access to the Internet, it also means that Japan's room for growth is limited. Statistically, Japan's growth in terms of Internet users is under 10% annually, which is dwarfed by most other East Asian countries. Japan's economic growth has slowed down, as with the other developed countries of the Eurozone, and so too has its growth in Internet usage. Nevertheless, the population that does have Internet access is extremely large, and therefore, Japan is a market that cannot be ignored.

Japan's data privacy legislation is generally stricter than similar legislation in the United States—although perhaps it's better described as more precise and well defined. Japan's data privacy law comes in the form of the Protection of Personal Information Act. The law does not regulate data privacy directly so much as it empowers various ministries within the government to regulate different aspects of data privacy. Industries may fall under the jurisdiction of one or several ministries, and therefore business may be required to comply with multiple regulations and guidelines. Businesses dealing with personal information databases will be made to follow the specific guidelines within their respective industries. Personal information itself is defined broadly to include almost any information regarding an individual that may be used to identify, and a database is defined officially as any data set containing information from over 5,000 individuals. To maintain a database, a business will have to specify the purpose of collecting the data and remain within the scope of that purpose. The manner of obtaining is to be fair, data is to be kept adequately secure, and consent should be obtained before sharing the information.

Japan's legal landscape regarding data privacy is different from that of the United States in that it favors a more opt-in approach than an opt-out approach. Consent must be obtained from individuals in Japan; in the United States, consent is taken to be implied if it is not denied. In terms of sectoral laws, there are further requirements depending on industry—some industries define certain procedures, including the appointment of data privacy officials and the requirement of internal inspections on data security practices; other industries maintain strict standards for data privacy that must be met, the method being left up to the business. Japan's data privacy law also provides for specific and strict penalties—violations are met with fines and even imprisonment up to six months. Japan's data privacy law does not distinguish between moving data inside and outside of Japan, which means that the law is relevant to businesses that are not primarily located in Japan. In practice and looking ahead to the future, the landscape is shifting to a more big data–friendly environment: it seems likely that Japan will attempt to revise its data privacy laws to accommodate for the increasingly large role that big data is playing.

Currently, big data already finds a home in many Japanese industries. Interestingly, the Japanese government itself is a huge player in the realm of big data. Japan's central government has attempted to employ data on population movement, tax revenue information, and more in an attempt to aid local municipalities in economic revitalization. The use of big data as a tool to facilitate economic and political policy by the Japanese government also means that Japan has adopted an Open Data Initiative. The initiative is an attempt to make public certain data such that it can be used for secondary purposes—for profit or for public improvement, among other purposes. The Initiative attempts to create, first, transparency and confidence in the government. Second, the Initiative seeks to increase collaboration and participation from both public and private sectors. Third, the ultimate result is that the constant flow of data will facilitate economic growth and efficient government.

In fact, Japan and open data have a longer history than just government involvement. Even before the government's movement toward open data, Japanese people have found uses for it. Most notably, open data facilitated the recovery from the 2011 earthquake—car

GPS data was used to find drivable roads, electricity shortage data was made available to encourage energy saving, and websites (*http://sinsai.info*, for example) were created to allow users to share relevant information.

Japan is not the fastest-growing country in Asia. Japan is also not the country with the most room for growth. However, Japan is the largest developed Asian country in terms of Internet users, which means that Japan is a viable place to engage in the big data market. Primarily, Japan's advantage over a less developed nation such as China is a government that maintains strict data privacy laws (which are likely to be altered) and seeks to promote the flow of open data. While China may be a country to look out for in the coming years, Japan is a great place to look now.

Korea

While Japan generally takes the cake for most developed East Asian country, South Korea has Asia's best Internet-related infrastructure. Boasting the highest connection speeds, highest available WiFi locations, and an 85% Internet penetration, it comes as no surprise that Korea is interested in the conversation surrounding big data. South Korea's mobile phone penetration is also extremely high (thanks in no small part to major national player Samsung).

South Korea's general privacy law comes in the form of the Personal Information Protection Act (PIPA). Enacted in 2011, PIPA defines personal information as any information that can itself, or in combination with other information, identify an individual. Sensitive data, or any information regarding and individual's ideologies, beliefs, political views, health, sexual life, and other personal information is also further regulated. PIPA regulates all aspects of this information, including how it is collected, recorded, stored, processed, searched, corrected, and destroyed. Furthermore, PIPA's jurisdiction extends even overseas, when a violation of PIPA would affect a Korean citizen or company. Therefore, it is crucial that companies doing business in Korea are familiar with the letter of the law. PIPA requires that data controllers obtain consent (although this consent may be obtained online, it cannot be tacit or implied consent). Data controllers are also responsible for maintaining acceptable security for data, and in the case of companies, a high-ranking employee is to be appointed to a data protection management position. Furthermore, Korea has industry-specific laws—for example, the IT Network Act places further restrictions on the collection, movement, and manipulation of data in the telecommunications industry. Both the general laws and the industry specific laws have teeth—they both possess enforcement agencies. Under PIPA, the Minister of Public Administration and Security (MOPAS) has authority, but separately, under the IT Network Act, the Korean Communications Commission (KCC) has legal authority. Korea already has some of the strictest data protection laws in the entirety of Asia. In addition, Korea has demonstrated a willingness to further tighten up its legal grip on the issue of data privacy. As Korea, just as any other nation, faces security breaches, Korea has continued to adapt its legal policy to fit the need for security. Moving forward, it may be expected that data privacy regulation in Korea continues to be strict.

Korea is, simply looking at the statistics, incredibly well connected. The aforementioned statistics regarding the overwhelming number of mobile users has some interesting, not-so-surprising side effects. For example, while ecommerce is dominated by PC shopping (70%–80% for most countries), South Korea sees a 50/50 split between online shoppers using a PC and online shoppers using a smartphone—Koreans are comfortable making purchases on mobile devices. The prevalence of mobile devices also means mobile-oriented social media. KakaoTalk is a messaging application that is on roughly 93% of smartphones in Korea; it is also aggressively expanding into a multipurpose platform for games, calls, video sharing, and more (following WeChat's model). The messaging service's wild popularity both illustrates and strengthens Korea's love for the smartphone.

Knowing how Internet connected Korea is, it comes as no surprise that open data is flourishing there. The OECD recently rated Korea as the top country in the world for open government data. The qualifications for a high ranking involved 19 variables, but boiled down to three main factors: availability, accessibility, and government support. Korea stands out as an Asian country that not only makes data available, but also seeks to help the private sector create businesses using said data. South Korea makes data sets available through a web portal, opendata.kr, which offers increasing amounts of data from various sectors of Korean government. Another popular page, *http://data.seoul.go.kr*, offers data specific to the Seoul municipality. The page offers real-time updates on public transportation, business hours and locations, and more, offering information that could be pertinent to businesses in machine-readable format. The Korea Energy Agency aims to use a competition model on open data to improve efficiency, in another example.

The overall climate in South Korea is one that prioritizes data security on the private side, and data openness on the part of government.

India

As the second most populous country in the world—and home to the second most Internet users as well—India plays a large factor in the discussion for big data going forward. Yet, despite its gigantic population, India is very much like China in that the size of its Internet-using population can only grow. While China has roughly half its population hooked up to the Web, less than a quarter of Indians have Internet access. This is, however, not cause for despair—just a few years ago, in 2010, India's Internet penetration failed to break even 10% of its population. Like China, India's economic growth means the growth of infrastructure, and with more of India's population receiving Internet access, India will be a place to watch closely.

Article 21 of India's Constitution serves as the primary legal basis for the data privacy law in India: "no person shall be deprived of his life or personal liberty except according to procedure established by law." India's Supreme Court recognizes the right to privacy as part of the right to life and personal liberty. This manifests itself in the Information Technology Act of 2000, which requires that corporations dealing with personal data implement security practices. Personal data is defined as any of the following: passwords; financial information; physical, mental, and psychological health condition; sexual orientation; medical records and history; biometric information. However, while many countries have similar acts, the specifics of the IT ACT and the IT Rules (2011) require that service vendors obtain *written* consent before they obtain personal data. Furthermore, movement and use of data to any party outside of India requires the receiving party to have at least the same level of security, as well as either a contractual agreement that requires the movement of the data or consent from the original data subject. This policy is far stricter than any data privacy policy in the United States, and companies outsourcing jobs to India or bringing business to India will have to be aware of the difference in policy. Some companies—notably Google—dislike the laws in India, finding them too harsh on data companies. Stricter data protection policy also encourages companies that already have stricter policies implemented to make the move into India. This specifically applies to EU companies moving into India; the European Commission already maintains strict data policies.

In many ways, the most important relationship US and EU companies have with India has been through outsourcing. The number of workers in India employed by US and EU companies, particularly in the IT services industry, is numerous, and these workers often receive significantly lower salaries as compared to their counterparts in the United States. The jobs outsourced to India are often lower-level IT jobs. With the rise of big data, the need for higher-level jobs —advanced data analytics, which requires significant training—is placing strain on not just the job market, but markets around the world. India, however, is aware of this demand. The Indian government is attempting to align itself with the data privacy laws of Europe; by aligning itself to EU standards, India will have the legal capacity to store and analyze sensitive and IP-protected data from Europe. Furthermore, online programs, schools, and existing companies are training qualified professionals. A huge advantage doing data-related business in India is the increasing pool of qualified, talented, and cheap labor.

India also has its own significant open data initiative. The culture of open data is teeming in India, with groups such as DataMeet drilling into the gold mine of big data. The government of India announced the National Data Sharing and Accessibility Policy (NDSAP) in 2012. The policy is an attempt to make transparent and accessible government data, all posted to *http://data.gov.in*, an open source portal. While NDSAP is a suggested policy, however, it is not mandatory. As such, not all sectors and branches of Indian government participate, and not all the data is easily readable or consistent. India's lack of infrastructure in many places makes it difficult to collect, store, or process data.

India's main focus for the past few years has been on providing a friendly climate for outsourcing, but its near future will see an increasingly important consumer market for data. Policies will conform to European standards in order to have access to European records where necessary, but it's unclear whether India's policies for its own citizens will follow this model or succumb to political exigencies and encourage governmental intrusion along the model of China. Although the most speech-restrictive elements of the Information Technology Act were struck down by India's Supreme Court, the law still provides the government with warrantless access to databases.

Indonesia

Indonesia is at the bottom of the list of these five Asian countries in terms of Internet users. Although home to more people than Korea and Japan combinded, Indonesia also has a far lower Internet penetration than both these countries (around 20%). Nevertheless, Indonesia's economic power makes it an important player, rivaling India in terms of growth.

Article 28G of the Indonesian Constitution reads as follows: "Every person shall have the right to protection of his/herself, family, honour, dignity, and property, and shall have the right to feel secure against and receive protection from the threat of fear to do or not do something that is a human right." Included in the interpretation of Article 28G is the right to privacy, which may be further interpreted to include the right to data privacy. While the Constitution loosely guarantees a right to data privacy, data privacy law in Indonesia is not very strict. The Law on Information and Electronic Transaction, or IET, describes the proper guidelines for electronic transactions. Specifically, Article 26 of the IET requires that consent be given for the collection of personal information electronically. However, the details of the enforcement and the scope of the IET remain to be further described by the Indonesian government. With Government Regulation No. 82, enacted in 2012, Indonesia attempted to outline a more definite stance with regards to data privacy. Regulation 82 defines personal data as any data specific to an individual that is to be stored, treated, maintained, and kept confidential. In many ways, Regulation 82 is more sophisticated than comparable laws in other countries: it covers service-level agreements between companies and requires providers to provide source code to their customers or to third-party escrow agents. It further enumerates proper practices for data controllers, which include maintaining confidentiality, obtaining consent before collecting data, proper use of the data, notifying the owners of the data in case of a breach, and upholding proper security practices. Nevertheless, the enforcement of law remains ill defined.

Indonesia, unlike China, has a perhaps surprising affinity for open data. For example, Indonesia is the only other East Asian country besides South Korea that is a part of the Open Government Partnership. As a developing country, Indonesia has a strong focus on open data for the sake of government and corporate transparency. The Swandiri Institute, for example, uses (and advocates for) open data for environmental sustainability and anti-corruption efforts. Jakarta itself is a hub for open data, with events such as HackJakarta drawing together a strong community of open data enthusiasts.

Regulation 82 shows Indonesia's second-mover advantage, clarifying relationships that other legal frameworks have avoided (or left to contract law), but the developing state of Indonesia's economy leaves much up in the air—whether power structures will favor government, established companies, or citizens, in particular. The swift adoption of mobile phones and growing middle class population may portend unforeseen models for data sharing and provisioning; in any case, the Indonesian market for data services is a wide open playing field.