Personal Data and Identifiers: Some Issues Regarding General Data Protection Regulations

Sung Chunhsien(✉) and Lu Shangchien

School of Civil and Commercial Law, Beijing Institute of Technology,
No.6 Jinfeng Road, Zhuhai, Guang Dong, China
scotsung@gmail.com

Abstract. This paper investigates the identifier in general data protection regulations in relation to personal information and privacy matters. The paper compares different legal systems regarding protection of personal information, taking the system of general data regulation (as adopted by the EU and by China) as the most influential protective regime. Because general data is defined as any information related to a person, the key term "identifier" becomes excessively broad in its application. The second half of this paper contributes to the discussion about identifiers and the de-identification process envisaged in the regulations, addressing certain criticisms of the identifier provisions.

Keywords: Data protection · Personal data identifiers · GDPR

1 Introduction

This paper has as its subject matter the regulation of the protection of personal data, i.e., the forms of information that are included in the personal data protection schemes used in the EU and in China. Generally, personal data protection regulations employ in-sequence protective schemes: measures relating to the data subject, to data processing (including data collectors and controllers), to security, and to aftermath (response and compensation); the concept of the "data subject" is the key to these other protective schemes. Therefore, entities involved in data processing are required to comply with the personal data protection regulations if the processed data falls under the scope of the data subject. In particular, the internet has created a difficult environment, because it is easy to identify a person via a search engine.

1.1 Protection in Different Systems

Personal data protection and privacy matters are worldwide issues, and international organizations including the United Nations, the Organization for Economic Co-operation and Development, and Asia-Pacific Economic Cooperation have created similar framework provisions for their member states to comply with. Protective schemes commonly use one of two modes. The first mode is protection in functionality (i.e., different regulations apply to different forms of personal data usage); the second

J.-L. Chen et al. (Eds.): WiCON 2018, LNICST 264, pp. 163–169, 2019.
https://doi.org/10.1007/978-3-030-06158-6_17

mode is general protection (i.e., an identical regulation applies to each individual). The former mode corresponds to the US pattern; the latter mode corresponds to the pattern used in the EU and in China.

These patterns have different protective purposes. The US functionality-directed mode focuses primarily on user privacy, which is a fundamental right according to the US constitution. However, the EU regards all personal information as general data and is more concerned about rights relating to personal information, regardless of the nature of the information. This major difference corresponds to a difference in the scope of the concept of data subject in each protective scheme. As a result of this difference, entities involved in data processing easily fall within the scope of EU protective jurisdiction.

1.2 Territorial Scope and Worldwide Influence of General Data Protection Schemes

The European General Data Protection Regulation (GDPR) came into force in May 2018 [1]. The implementation of the GDPR has had a worldwide influence because of its provisions under “territorial scope”, which give worldwide jurisdiction over any entity involved in using the personal data of European residents. In other words, regardless of the location of an entity, its access to European personal data would put the entity under the control of the GDPR. In particular, use of the internet, which is considered to be a space without territorial limitations, may lead to unforeseen consequences because of the difficulty of managing internet users [2].

Since EU Data Protection Directive 95/4630 (GDPR Predecessor) established “the most influential international policy instrument” [3] in the field of data protection, the EU personal data protection regime has inspired many data protection provisions in different regions. In the case of China, although the Cybersecurity Law came into force in June 2017, the authority began to apply personal data protection policies in 2000 with a decision of congress safeguarding internet security. The first explicit ruling was the 2005 amendment of the Criminal Law (Art. 253-1) regarding the infringement of citizens’ personal information. The territorial scope was later established in the aforementioned Cybersecurity Law.

1.3 The Substance of This Paper: Personal Data Subjects in the EU and China

On account of the worldwide implications of general data protection regulations, this paper focuses on the role of the data subject in regulation in the EU and China, the world’s two biggest personal data protection jurisdictions in terms of population and economic scale.

The concept of the personal data subject (i.e., the definition of personal data and the nature of that data) is the fundamental issue at stake. This paper therefore provides a comprehensive analysis of the data subject as conceptualized in the EU and China regulations.

Information related to personal data is classified into general information, identifiers, and sensitive data. This classification closely corresponds to different levels of security and their corresponding safeguarding measures.

2 Data Subjects: General Personal Information

2.1 Any Information

Personal data is a broad concept that can cover any form of information used to recognize a natural person. Accordingly, the term "any information" needs to be taken literally; in other words, any information related to a person that may have an impact on his or her privacy rights is the subject matter of data protection regulations. As a result, any element involved in verification of identity, including physical, biometrical, or factual information, falls within the category of general data. Under personal data regulations, identity verification also refers to the use of a combination of information to identify a person. Therefore, if a single piece of information can with the help of other information be used to identify a person, these pieces of information are subject to personal data regulations.

The identification of an individual using multiple pieces of information is known as "internet doxing" or re-identification.

2.2 General Data Related to a Person

Personal general data is a broad term, and it is difficult to give a clear definition. Generally speaking, it refers to any information related to a person, including the following:

- physiological features, such as appearance, eye color, height, weight, health status, and genetics (including medical history, genetic data, and information about sick leave)
- personal circumstances, such as social security or ID numbers, phone numbers, residential address, email addresses, location data, and economic status
- habits or behavior, such as character traits, religion, cultural factors, political opinions, and geotracking data
- biographical information, such as date of birth, workplace data, level of education, salary, tax information, and student ID number.

However, owing to policy considerations, not every type of information listed above is included in the regulations. The EU and China have also used different terminology for the various types of personal data in their respective regulations.

2.3 Data Protection Regulations

The EU GDPR defines personal data as "any information relating to an identified or identifiable natural person." The GDPR gives certain forms of information as examples, including but not limited to "a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." In the case of China, the Provisions on Protecting the Personal Information of Telecommunications and Internet Users (Internet Provision China) [4] define personal data as "information with which the identity of the user can be distinguished independently or in combination with other information." Internet Provision China gives certain forms of information as examples, including but not limited to "a user's name, date of birth, identity card number, address, telephone number, account number, passwords."

In connection with the China Internet Provision, it should be noted that the scope of personal data is a successor to the GB/Z 28828–2012 Guideline for personal information protection within information systems for public and commercial services.

2.4 Rights in Relation to Personal Information in China

Another important aspect of the China regulation is the latest reform of the General Provisions of the Civil Law, which came into effect on 1 October 2017. The reform specifies "right of personal information" as an independent measure within civil rights. Article 111 states that "Natural persons' personal information shall be protected by law. Any organizations and individuals who need to obtain personal information of others shall obtain the information according to law and shall ensure the safety of the information. It is not permitted to illegally collect, use, process, or transfer the personal information of others. It is illegal to buy and sell, supply, or publish the personal information of others." Although the latest reform and implementation of Article 111 in Civil Law have led to some updating of cybersecurity standards in the privacy domain, the term "personal information" as used in Article 111 is not clearly defined, and no further interpretation is provided. Thus, the issue of whether the forms of personal data listed in Internet Provision China are regarded as a civil right or are protected by civil law remains unclear.

3 Data Subjects: Identifiers

3.1 Identifiable Data Subjects

In most data protection schemes, several types of data are categorized as personal because they enable the singling out or identification of a natural person. Identification of this sort is not evaluated purely in terms of individual pieces of information; it covers any combination of pieces of information that may directly or indirectly identify a person. Since identification may take place on the basis of one or more pieces of information, once the person is singled out, these pieces of information, taken singly or jointly, are defined as "identifiers" [5].

Although the concept of identifier is at the core of the concept of the personal data subject, it is controversial. In the context of personal general data, the regulations only govern personal information that is identifiable. Therefore, an identified data subject is a person who can be clearly known, named, or recognized; directly identifiable examples include a person's full name or appearance; indirectly identifiable examples include a person's mobile phone number, email address, or any form of ID number.

Nevertheless, the consequences of information combination result in an ambiguity in the term "indirectly identify a person." To take a practical example, a list of first names does not enable the singling out of a person; however, the addition of further information, such as residential address, workplace data, or surname, allows a particular group of people to be extracted by means of the combination of information. Accordingly, when an individual can be recognized from the group on the basis of a combination of information, that combination of information is regarded as an identifier.

3.2 Online Identifiers

Consequently, information combination may give rise to a large number of potential identifiers, and given the possibilities of the internet, more online identifiers are likely to emerge. The EU GDPR clarifies online identifiers in its Recital 30, which covers information from a number of sources that may single out a person, such as devices, applications, cookies, radio frequency identification tags, and tools and protocols (including IP addresses).

It should be noted that traces of such sources may become identifiable by means of other information (either online or offline). For instance, the keywords entered into a search engine are temporally saved in cookies, which may indicate a tendency or behavior on the part of the user. This sort of information may, in combination with identifiers or other information, single out a person.

3.3 De-Identification

Anonymous information falls outside the scope of general data regulations. In other words, data processors who are not willing to be governed by the regulations must understand and make use of the principles of data protection concerning identifiers and de-identification.

In Article 4(5), the GDPR defines pseudonymization as follows:

> ‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

However, it is not easy to ensure treatment of data that makes personal information no longer identifiable; as long as the information can be combined with other information (e.g., in case of failure to keep the additional information separate), such a combination may suffice to narrow the range to a specific group or to single out a person.

In Recital 26 of the GDPR, even though personal data have undergone a process of pseudonymization, if a person can be singled out by the use of additional information, that processed information should be considered as an identifier. In other words, anonymous information must afford no possibility of identifying the data subject.

3.4 Appropriate Measures and Data Doxing

In most cases, general data protection regulations recommend that the process of de-identification should take the form of appropriate technical and organizational measures that are able to ensure ongoing confidentiality and integrity. However, the authorities provide no specific information about what counts as "appropriate measures." For example, in the case of the GDPR, apart from pseudonymization, the only measure suggested is data encryption. This suggestion offers very limited help to entities involved in data processing. Data encryption is commonly associated with concerns about data leaking or hacking; identifiers are more commonly associated with concerns about data "doxing" on the internet (i.e., combination of information).

Data doxing involves narrowing the scope so that a particular person can be singled out. Thus, in order to single out an individual, the conditions used to narrow down the possibilities are vital and unpredictable factors. For example, a surname is regarded as personal information; however, a surname on its own usually indicates nothing and is therefore not regarded as an identifier. However, once a surname is combined with workplace information, these two conditions taken together indicate a particular group of people. If the surname is rare or unique, or if the workplace is very small, these two conditions may be sufficient to single out a person, regardless of any additional information. Therefore, unpredictability is an outcome of the characteristics of personal data, not of the form of personal data.

4 Conclusions: Some Thoughts on Re-Identification

With reference to distinguishing identifiers from personal information, the GDPR offers the following guidance to data processors in its Recital 26:

> To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

Therefore, objective factors and technological considerations are crucial to distinguishing identifiers.

The previous section approached unpredictability in terms of whether internet information is capable of doxing. This section takes a technical approach: the best method of taking personal information outside the scope of identifiers is to incorporate de-identification (or pseudonymization) as a fundamental process. Regardless of the forms or catalogues of personal information, de-identification must take account of the "directivity" of the information. Four sorts of directivity are described as follows:

1. Personal information that directly indicates a single person, such as full name or appearance. The data processor should treat this sort of information using strict de-identification or pseudonymization approaches (i.e., completely anonymous treatment).
2. Personal information that indirectly indicates a single person, such as social security or ID numbers and mobile phone numbers. The data processor should treat this sort of information using pseudonymization approaches (i.e., blocking part of the information).
3. Personal information that can be used for cross-examination (doxing), which includes most information, such as part of a name, date of birth, home phone number, residential address, and workplace. The data processor should treat this sort

of information using pseudonymization or isolation approaches (i.e., keeping the types of information separate).

4. Personal information presented with pure numbers and without directivity, such as weight and height. This sort of information apparently falls outside the scope of identifier, because without any further identifiable information, mere numbers would not suffice to identify anything. It is nevertheless necessary to isolate this numerical data from identities.

Personal data protection is a critical issue in the context of the internet, as it has become more difficult to isolate one piece of information from another. Thus, the scope of the concept of identifier is broader than it first appears. In particular, as most forms of e-commerce involve personal information, the establishment of safeguards regarding identifiers is a difficult task that must be reconsidered.

References

1. Commission Regulation 2016/679, 2016 O. J. (L119) 1(EU)
2. Voss, W.: European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting, vol. 72, The Business Lawyer (2016–2017)
3. Bennett, C., Raab, C.: The Governance of Privacy: Policy Instruments in Global Perspective. MIT Press (2006)
4. Provisions on Protection of Personal Information of Telecommunications and Internet Users, MIIT Order No. 24 (2013)
5. Elliot, M., Mackey, E., O'Hara, K., Tudor, C.: The Anonymization Decision-Making Framework. UKAN, Manchester (2016)