S&R

Data+

It's Personal: A Roadmap for Data Mapping in Digital India

Introduction

India's newly published Digital Personal Data Protection Act, 2023 (the "**DPDP Act**") is poised to take effect soon. Accordingly, "**data fiduciaries**" (*i.e.*, those entities which determine the purpose and means of processing personal data, including in conjunction with other entities) should start designing a long-term compliance strategy to meet obligations under the new law.

Although the DPDP Act is not yet in force, it appears that the Central Government (the "**Government**") may provide about six months for entities to align themselves to the new regime. While different dates may be appointed for different provisions, several such provisions require specific rules which are yet to be prescribed.

Thinking Ahead

However, while waiting for the Government to (i) frame rules, and (ii) notify discrete provisions, organizations could use this transitional phase to prepare for future compliance requirements.

Immediate Steps: Data Mapping

Before anything else, data fiduciaries could draw up a compliance roadmap, the starting point of which should include a comprehensive data mapping exercise. This exercise, in turn, could involve the following steps:

Step 1: Understanding the scope and definition of personal data under the DPDP Act

Organizational databases are likely to contain vast volumes of information, not all of which may be considered 'personal'. The DPDP Act defines personal data as any data about an individual who is identifiable by or in relation to such data.

What may be considered personal data under the DPDP Act?

Personal data may include any kind of information about an individual person. It could involve objective information, such as the person's name, mobile phone number, locational data (using functions on a cellular device, for instance), an internet protocol (IP) address, a cookie ID or details about their marital or medical status. A person's economic information (*e.g.*, details about a bank loan, annual income or billing history) may be considered personal data too.

Generally speaking, information that is only about a business or corporate entity is not considered personal data because it would not identify a specific individual. However, an individual's personal information may be so intertwined with data about a business or a company that such information may be able to identify and/or relate to that individual (*e.g.*, where the business is owned and managed by a sole proprietor).

The meaning of personal data may also involve subjective information, including opinions, estimates or assessments. Such subjective information may be especially relevant in certain sectors, such as insurance (*e.g.*, assessments made about an individual's life expectancy) or banking (*e.g.*, evaluations about the reliability or creditworthiness of a borrower). Further, information or an opinion about an individual from their leisure activities, such as their retail or other preferences based on online purchases or web browsing history, may also qualify as personal data.

In summary, personal data includes information about people irrespective of their position or capacity – whether they are consumers, patients, employees, customers or clients. However, that information must be *about* an identifiable individual. Information is 'about' an individual when there is a connection between the two. This is ultimately a question of fact and may depend on the context and circumstances of each case. For example, information will be 'about' someone where the person is the subject of an opinion. Similarly, information will also be 'about' someone when it reveals or conveys something about them.

Decisions on whether data is personal or not should be made on a case-by-case basis, with reference to the circumstances and context of each situation. Some information may not be personal data when considered on its own. However, when combined with other information, it may become so. Information can therefore be dynamic, and the character of data may change over time.

Identifiability

Since identifiability is a central feature of personal information, entities often consider subjecting data to de-identification processes. Pseudonymization and anonymization are some such techniques. Pseudonymization refers to a process of disguising identities which reduces the risk of harm if and when a breach occurs. Anonymization relates to a process pursuant to which all identifying elements are eliminated from a personal dataset.

However, de-identified, encrypted or pseudonymized personal data (discussed below) – if susceptible to use for the purpose of re-identifying an individual – may continue to retain its status as personal data. Nevertheless, when such data is rendered anonymous in a way where the individual concerned is no longer identifiable, it may cease to remain personal. For data to be truly anonymized, however, the procedure involved must be irreversible.

Anonymization

Anonymization requires the use of mathematical/technical methods to distort data adequately and irreversibly for the purpose of ensuring that (re)identification is not possible. In this respect, anonymization is distinct from plain vanilla de-identification techniques that merely involve the masking of identifiers from datasets (although such methods do tend to make identification more difficult and/or costly). However, given the pace and nature of technological change today, it is difficult to identify precise standards. After all, despite the removal of identifiers, de-identified data does involve a higher risk of re-identification. Current research suggests that in certain situations, it may be possible to identify specific individuals even from apparently anonymized datasets.

The DPDP Act does not explicitly refer to, or exclude, anonymized data from its ambit. However, as long as it can be shown that it does not identify a specific individual, whether on its own or in conjunction with other information – such data is likely to remain exempt from the application of the DPDP Act. Previous versions of this law had defined anonymization as the irreversible process of transforming or converting personal data to a form in which the data principal cannot be identified. Since anonymization is a standard practice in data aggregation processes, data fiduciaries could consider employing it as a technique, especially if aggregated insights are enough for their business aims (*e.g.*, to examine a general trend or demographic).

Pseudonymization

Although the DPDP Act (unlike the EU's GDPR) does not mention or recommend pseudonymization, this process may prove useful, including in respect of satisfying core obligations under the law. After all, the new regime does require data fiduciaries to implement appropriate technical measures to ensure effective compliance. Further, entities are required to protect the personal data in their possession or control by taking reasonable security safeguards to prevent a breach.

Tests for determining personal data

For data to 'relate' to an individual, either of three key elements needs to exist, which are associated with (i) content, (ii) purpose, or (iii) result.

Content

The element of content is present where information about a particular individual may be available with an entity, regardless of the purpose or impact of such information on the person concerned. For example, results from a blood test may unequivocally relate to a patient, or the information contained in a bank's database under a particular name will obviously relate to such customer.

Purpose

The element of purpose may be invoked when certain valuable information which a data fiduciary or a third party intends to use relates to a certain individual. This can happen when the data is likely to be used with the aim to evaluate, collate, commercially exploit and/or influence the status, preferences, profile or behavior of an individual.

For instance, online identifiers (such as IP addresses and cookies) obtained through the handheld devices used by, or associated with, certain individuals may leave behind traces which, when combined with uniquely identifying elements or other information received by servers, are capable of creating their profiles.

Result

Despite the absence of content or purpose, data may nevertheless relate to an individual when its use is likely to have an impact on their rights and/or interests. It need not be a major impact, as long as an individual is likely to be treated differently from others as a result of such data processing.

Checking for personal data within mixed datasets

In most situations, a dataset is likely to be 'mixed' – comprising both personal and non-personal data. Examples of mixed datasets include those where customer or client information is clubbed together with transaction details, including those involving payments made through credit and debit cards. In addition, some organizations use customer relationship management ("**CRM**") services provided by third parties that require an individual's data to be made available in the CRM environment. Data held in the CRM service may include information necessary to manage interactions with the customer, such as their email address, phone number, as well as the products and services they have purchased. This dataset can therefore include both personal and non-personal customer data.

In some situations, the information conveyed by a piece of data may concern something else, and not a specific individual as such. Even in these cases, the information may relate to an individual indirectly. For example, a digitized automobile service register may contain information about a car, its technical problems and its service check details (all non-personal data). However, this information may be associated with a separate record of vehicular registrations and/or number plates, which, in turn, may be linked to a list of individuals. Since the service provider that holds these records may be able to establish a connection between the vehicle and its owner, even the non-personal vehicular information may relate to the owner and thus represent their personal data vis-à-vis the service provider (*i.e*., the data fiduciary).

Step 2: Determining what personal data is collected and/or used by the organization

This step should also include ascertaining the different *types* and/or *categories* of personal data which is being/has been collected, processed and stored.

Step 3: Finding out where the data is stored, including which third-party systems house it and where those servers are located

Step 4: Mapping where the data goes from the point of collection and across the organization, including through internal departments and externally to vendors, processors or other third parties

Steps 3 and 4 should also include ascertaining *who* (*i.e*., which individual, department and/or other entity) has stored, accessed and/or otherwise processed such data, and *where*.

Step 5: Determining whose data it is, how long such data is retained and in what formats.

Step 5 could include ascertaining if the data is structured or not. Each individual whose personal data is being/has been processed by the organization needs to be specifically identified from organizational databases. Later, once the DPDP Act takes effect, such individuals need to be contacted for the purpose of giving statutory notices. However, in case of proprietary processes or systems that store unstructured data, certain manual interventions and discovery modes may be necessary.

Conclusion

The best way to conduct the data inventory and mapping will depend on (i) an organization's size and complexity, (ii) the amount of time allotted to the exercise, and (iii) the sophistication of the participants involved. Even an initial discovery process may provide valuable information about data life cycles within the organization.

The data inventory, along with its supporting procedures, should ultimately aim to produce a state where the organization is able to identify both the location of the data, as well as its storage details, *at the level of each individual person* – such that, if and when data principals, in exercise of their statutory rights under the DPDP Act, wish to access their data, it can be provided to them within a reasonable time.

For some organizations, the personal information contained in legacy databases may be tracked via native tools available through standard enterprise software products. Manual and informal processes can also be used. More sophisticated approaches may include the use of commercially available products and platforms which have been especially developed for this purpose (*i.e.*, for data inventory and mapping). Various technology companies, including startups, provide such tools and technical solutions.

During the inventory process, it may be useful to consider if a piece of personal information is really necessary to be retained by the organization, and why. Other than on the basis of a data principal's explicit consent, the DPDP Act allows personal information to be processed for 'certain legitimate uses'. Ascertaining whether such a basis exists (or not) may expedite compliance with some of the key obligations under the DPDP Act, including in the future.

While technical solutions may be sourced from third parties to help with data mapping, organizations may prefer to use in-house platforms. Irrespective, a careful inventory of both data and allied processing practices is an important first step to prepare for future compliances.

*This insight has been authored by **Deborshi Barat** (Counsel); he can be reached at dbarat@snrlaw.in for any questions. This insight is intended only as a general discussion of issues and is not intended for any solicitation of work. It should not be regarded as legal advice and no legal or business decision should be based on its content.*

S&R ASSOCIATES
ADVOCATES

NEW DELHI
64 Okhla Industrial Estate
Phase III
New Delhi 110 020
Tel: +91 11 4069 8000

MUMBAI
One World Center, 1403 Tower 2 B
841 Senapati Bapat Marg, Lower Parel
Mumbai 400 013
Tel: +91 22 4302 8000