PERSONAL DATA PROTECTION COMMISSION SINGAPORE

PRACTICAL GUIDANCE IN RELATION TO DATA COLLABORATION ARRANGEMENT INVOLVING COMMON DATA INTERMEDIARY

1. The Personal Data Protection Commission (PDPC) received several requests for guidance on data collaboration arrangements involving a common data intermediary ("DI") that processes personal data on behalf and for the purposes of groups of data contributors ("DCs"). This document consolidates the PDPC's guidance and provides additional clarification on various PDPA-related issues relating to data collaboration arrangements.

Data Collaboration Model

2. In general, the guidance sought by the organisations relates to three broad stages in data collaboration models:

    a. **Feasibility Testing Stage**: Before undertaking data collaboration projects, the DCs may wish to conduct feasibility tests utilising a common DI. This may involve DCs disclosing tokenised NRIC numbers to the DI to determine if meaningful research can be performed. For example, the DI would compare a list of tokenised NRIC numbers of individuals collected from the DCs and determine whether there is sufficient overlaps amongst the DCs to proceed with the projects.

    b. **Data Processing Stage**: The two main modes of data processing are: i) the common DI combines the raw personal data contributed by the DCs and performs minimal or no further data processing. The resultant combined dataset is then shared with the DCs; or ii) the common DI creates derived data and insights from the merged datasets contributed by the DCs. The derived data is then shared back with the DCs to be either used as a new attribute relating to an individual or in its aggregate form.

    c. **Marketing Stage:** DCs plan to use the insights and models generated to inform their marketing strategies, and may include sharing leads with each other to test out these marketing strategies.

PDPC's guidance

Feasibility Testing stage

3. The PDPC notes that feasibility tests usually involve the processing of tokenised NRIC numbers by the common DI on behalf of and for the purposes of the DCs and it will be conducted as a **preparatory step to the eventual fusing and**

anonymisation of datasets, solely for the purpose of ascertaining the pool of common individuals amongst the datasets provided by the DCs. Tokenised NRIC numbers are used as this is a convenient form of universal or common unique identifier that is fairly accurate and unchanging. As tokenised NRIC numbers alone, without other information about an individual, are less likely to identify specific individuals, the risk of identification of individuals is assessed to be low. As such, they can be used in this preparatory step. Given this, **consent is not required for the data collaboration partners' disclosure of tokenised NRIC numbers** to the common DI. Care should nevertheless be taken to adopt responsible practices in the creation of tokens and when transmitting lists of tokens. For example, using a confidential salt when creating the token, selecting an appropriate encryption or hashing algorithm, and password-protecting the lists of tokens during transmission.

Data Processing Stage

4. PDPC would like to first clarify on the roles and responsibilities of DCs and DI in data collaboration arrangements. Where an organisation with possession or control of personal data – referred to as the DC – engages another organisation to process the datasets on its behalf and for its purposes, the organisation processing data is considered a DI. Where a DI is involved in the anonymisation of personal data, both DCs and DI have different roles to play in ensuring compliance with the PDPA Obligations.

    a. DCs are responsible for:

        i. Specifying to the DI its business requirements and management decisions (including approving the technical and process measures that the DI recommends) in relation to the outsourcing;

        ii. Putting in place legal and process safeguards to ensure that DI is protecting the data adequately when it is processing the data.

    b. DIs are responsible for:

        i. Making recommendations on the technical means of anonymising and/or fusing the data in order to achieve DCs' requirements;

        ii. Making recommendations on technical and process measures to protect data that it is processing on behalf of DCs, including addressing re-identification risks and securing data in transit;

        iii. Implementing the technical, process and legal measures and safeguards that have been approved by DCs.

5. The DI should also ensure that it only processes the personal data for the DCs' purposes in order to stay within the reduced set of obligations that DIs are subject to. In the event that the DI uses or discloses personal data in a manner that goes beyond the processing required by the DCs (e.g., using or disclosing data obtained from the Data Contributors for other purposes), it will be required to comply with all Data Protection Provisions under the PDPA. The DI should formally establish this relationship and arrangements with the DCs, such as through a written contract or agreement.

6. Under the mode outlined at paragraph 2(b)(i) above, the resultant dataset returned to the DCs is a combination of the raw personal data contributed by the DCs. As such, **PDPC will consider the DCs to have direct access to the personal data provided by the other DC in the merged dataset. Consequently, DCs will be considered to have either collected or disclosed personal data of their respective customers[1], for which either an exception to consent applies or customer consent is required**. The same position will apply in relation to personal data of non-customers that are shared by the DCs with the common DI. As the common DI is processing the personal data on behalf of and for the purposes of the DCs, it can only do what the DCs themselves are permitted to do. In this case, the common DI may merge the datasets without consent if the DCs are able to rely on an applicable exception to consent to do so. However, if DCs are unable to rely on an applicable exception, consent will be required and the common DI may only merge the datasets if the DCs have obtained consent for that purpose. As the Consent Obligation under the PDPA is applicable only to DCs, the responsibility to assess the applicability of exceptions to consent lies with the DCs. DIs that process data on behalf and for the purposes of the DCs will be entitled to accept and rely on the representations and warranties from the DCs that either an exception is applicable or customer consent has been obtained. Nevertheless, the DI should, as part of performing due diligence, ensure that the question of applicable exception or customer consent is addressed as part of its discussion with the DCs.

7. **Modes of Processing.** Under the mode outlined at paragraph 2(b)(ii) above, the derived data in the form of i) aggregated insights; ii) attributes of customer user profiles; or iii) attributes relating to an individual are generated based on personal data contributed by the DCs. Each of the DCs has no direct access to the raw personal data contributed by the other party, as the processing is done by the common DI. As these insights are derived by analysing or aggregating customers based on business rules and drawing inferences relating to common features or preferences of either aggregated group[2] (e.g. profile creation) or individual

[1] For example, additional attributes of common customer obtained from the dataset from the other DC.
[2] Aggregation is one of the techniques to anonymise personal data. Refer to PDPC's Guide to Basic Anonymisation Techniques for more details.

customers, the new or additional attributes cannot be used to recreate the personal data contributed by each DC. As such, **PDPC will consider DCs to have used personal data of their own customers to derive aggregated insights or new attributes, for which separate consent is not required**. However, this guidance is conditioned on implementing sufficient accountability measures to prevent one DC from gaining access to the raw personal data from the other DCs. The accountability measures should minimally include:

a. **Legal safeguards/agreement**: A contractual agreement between the DCs and common DI, clearly specifying the common DI's obligations and responsibilities in respect of its processing and generation of aggregated insights on behalf of and for the purposes of DCs. DCs should also ensure that the common DI does not use or disclose the personal data in a manner that goes beyond the processing required by the DCs.

b. **Assess and ensure that there is low risk of DCs gaining access to raw personal data contributed by the other DC**:

(i) **When sharing personal data between DC and DI**: Technical and procedural safeguards should be implemented to prevent re-identification of individuals. Examples of technical/procedural safeguards include pseudonymisation, hashing of personal data such that no personal identifiers will be shared in the clear; and/or having a clear separation of roles such as having separate teams within the common DI (e.g., one team that defines the common salt and another team to fuse/merge the data).

(ii) **When common DI is processing personal data and generating derived insights**: In cases where the processing is minimal and/or based on simple and well-known mathematical calculations that could be easily reverse-engineered to obtain the original dataset, the resultant dataset may not be considered as derived insights and the PDPC's guidance at para 5 will apply.

(iii) **When sharing derived data/insights between DI and DC**: The customer profiles to which the derived data/insights relate to should not be granular such that it enables DCs to deduce the original dataset contributed by other DCs. For instance, derived insights or specific attributes about a customer profile consisting of a larger pool of individuals will likely be considered to be aggregated insights and it is less likely that DCs will be able to deduce the raw personal data contributed by the other DC.

8. In addition to the above accountability measures, DCs should also take the following measures when considering data collaboration arrangements:

   a. **Assess and ensure that the purpose of using the derived insights is reasonable:** DCs should determine the actions and decisions they are planning to make with the derived data/insights that is shared by the common DI (e.g., matching derived data against CRM for marketing purposes), and the risks of impact to individuals if they carry out such actions and decisions. The purpose will be considered reasonable if it is what a reasonable person would consider appropriate in the circumstances.

   b. **Ensure accuracy of personal data**: DCs should put in place measures and processes to ensure that the raw personal data shared with the common DI is materially accurate before further processing takes place. Where the aggregated or derived data being shared with the DCs are based on pre-agreed categories and profiles of individuals (determined by DC), DIs should ensure that the selection criteria are applied accurately when labelling or assigning specific individuals to each category/profile.

Marketing Stage

9. After generating insights as described at paragraph 2(c), DCs may want to share with each other contact information of individuals identified as leads so that they may seek consent from the individuals for direct marketing purposes. In this situation, as the DCs will share personal data (i.e., contact information of individuals) with each other, DCs will be considered to have collected (or disclosed) personal data, and consent will be required, unless any exception applies.

END OF DOCUMENT