**PRACTICAL GUIDANCE SOUGHT BY ORGANISATIONS INVOLVED IN A DATA COLLABORATION ARRANGEMENT**

1. Guidance from the Personal Data Protection Commission (the "Commission") was sought in relation to a data collaboration arrangement involving several organisations and a public agency (collectively referred to in this Guidance as "data collaboration partners"). The purpose of the data collaboration is to address social well-being issues through the use of anonymised or synthetic datasets. The anonymised or synthetic datasets will be prepared by a dedicated team from the public agency, using pseudonymised data[1] provided by the data collaboration partners.

2. The data collaboration arrangement will involve the following steps:

   (i) **Defining population of interest:** There will be one team ("Team A") from the public agency that defines the parameters for a population of interest.

   (ii) **Generating pseudonymised data using a common salt:** Team A will then define a common salt, which will be used with an agreed irreversible encryption algorithm to generate a one-way hash.

   (iii) **Sharing of population of interest and common salt with data collaboration partners:** Team A will share (a) the parameters for the defined population of interest; and (b) the common salt with its data collaboration partners.

   (iv) **Extracting the relevant data and generating pseudonymised datasets**: Based on the parameters and the common salt shared by Team A, the data collaboration partners will select a sample dataset from their database to ensure sufficient overlap with the population of interest, and then apply the one-way hash to generate pseudonymised datasets relating to the selected individuals, consisting of hashed identifiers and the corresponding data points that are relevant for the data collaboration.

   (v) **Sharing of pseudonymised datasets to a dedicated team for fusing and anonymising of datasets:** The data collaboration partners, including Team A, will then share the pseudonymised datasets with a separate team from the public agency ("Team B") that is in charge of fusing and anonymising the datasets on behalf of the data collaboration partners. There will be strict separation of roles between Team A and Team B, and Team B will not have access to the common salt that generated the hashed identifiers.

[1] Pseudonymisation refers to the replacement of identifying data with made up values. More details on pseudonymisation can be found in PDPC's Guide to Basic Data Anonymisation Techniques.

(vi) **Fusing and anonymisation of datasets**: Team B will then fuse the pseudonymised datasets provided by all data collaboration partners. If the fused dataset is not assessed to be anonymised, Team B may further process the fused dataset to ensure it is anonymised. If the fused dataset cannot be anonymised, Team B will generate a synthetic dataset from the fused dataset.

(vii) **Disclosing anonymised and/or synthetic datasets**: The fused dataset and/or synthetic dataset will be disclosed to another organisation for purposes of hosting the datasets for others to access and analyse for a set of pre-defined purposes. The datasets will be hosted in an environment that disallows downloads.

3. The Commission understands that there may not be consent obtained for the collection, use and disclosure of individuals' personal data in a pseudonymised format for the purpose of this data collaboration arrangement.

**PDPC's Guidance**

4. Specifically, the Commission provided guidance on the following:

(i) Whether consent is required for the data collaboration partners **to disclose** pseudonymised data to Team B to fuse the data and to anonymise the fused datasets (Data Extraction and Pseudonymisation stage);

(ii) Whether consent is required **for processing** of pseudonymised data to fuse and create anonymised datasets (Data Fusion and Data Anonymisation stage); and

(iii) Purposes for which the fused dataset and/or synthetic dataset **may be used and disclosed** (Data Distribution stage).

5. For the avoidance of doubt, the guidance set out in this document has been scoped to address the situation as described in paragraphs 1 to 3 above, based on the information provided.

6. The Commission's guidance relates to the application of the Personal Data Protection Act 2012 ("PDPA"), and is not applicable to the public agency[2] involved in the data collaboration arrangement.

[2] Refer to Section 4(1)(c) of the PDPA; further, data-related activities of public agencies are governed under the PSGA.

**(a) Whether consent is required for data collaboration partners to disclose pseudonymised data to Team B to fuse the data and to anonymise the fused datasets (Data Extraction and Pseudonymisation stage)**;

7. The Commission notes that under the proposed data collaboration arrangement, Team B is fusing and anonymising the pseudonymised data on behalf and for the purposes of the data collaboration partners. Given so, data collaboration partners are responsible for ensuring that there are appropriate safeguards (e.g. legal and procedural controls) in place to ensure the fusing and anonymisation of data done by Team B are scoped to their purposes and individuals cannot be re-identified from the anonymised and/or synthetic datasets, in accordance with section (c) below.

8. Where Team B is from a public agency, it is excluded from the application of the Data Protection Provisions of the PDPA. Nonetheless, other data collaboration partners' disclosure of personal data to Team B is subject to the PDPA. Under the PDPA, consent is required for the organisations to collect, use and disclose personal data for a purpose (unless an exception applies). However, **consent is not required to process and convert personal data into anonymised data**. The Commission is of the view that **the data collaboration partners' disclosure of pseudonymised data to Team B is for the purposes of fusing the data and anonymising the fused dataset, for which separate consent is not required.** As consent is not required for the fusing and anonymisation of data, consent is consequently not required for the data collaboration partners' disclosure of pseudonymised data to Team B to perform the fusing and anonymisation.

9. **The analysis in the preceding paragraph applies in the situation where Team B is not a public agency and is a data intermediary[3] under the PDPA.** In this case, it will be subject to the Protection[4] and Retention Limitation[5] Obligations under the PDPA. In the event the data intermediary uses or discloses personal data in a manner that goes beyond the processing required by the data collaboration partners (e.g., using or disclosing data obtained from the data collaboration partners for other purposes), it will be required to comply with all Data Protection Provisions under the PDPA. Among other requirements, it must ensure that consent has been obtained to use or disclose the personal data for these other purposes.

[3] A "data intermediary" is defined in the PDPA as an organisation that processes personal data on behalf of and for the purposes of another organisation. The arrangement where an organisation is to act as a data intermediary of another organisation should be set out clearly in a contract that is evidenced or made in writing, including the responsibilities and liabilities of each organisation in relation to the processing of the personal data.

[4] Section 24 of the PDPA.
[5] Section 25 of the PDPA.

**(b) Whether consent is required for processing of pseudonymised data to fuse and create anonymised datasets (Data Fusion and Anonymisation Stages)**

10. The Commission notes that the proposed data collaboration arrangement will incorporate the following measures as safeguards during the course of the collaboration:

    (i) Use of hashed identifiers based on a common salt;

    (ii) Separation of Team A (which defines the common salt) and Team B (which fuses the data and anonymises the fused dataset). This ensures that Team A will not gain access to pseudonymised datasets shared by the other data collaboration partners, and Team B will not have access to the common salt used to generate pseudonymised datasets; and

    (iii) Contractual agreement between data collaboration partners clearly specifying Team B's obligations and responsibilities in respect of the fusing and anonymisation of datasets, on behalf of and for the purposes of the data collaboration partners.

11. The Data Protection Provisions under the PDPA do not apply to anonymised data. "Anonymisation" refers to the process of converting personal data into data that cannot be used to identify any particular individual, and can be reversible or irreversible[6].

12. The Commission notes that at the Data Fusion and Data Anonymisation stages, the outcome of Team B's processing is to create fused and/or synthetic datasets that are anonymised. Based on the position stated in paragraph 8, **data collaboration partners do not need to obtain consent to fuse the data and to anonymise or create synthetic dataset from the fused dataset.**

**(c) Purposes for which the fused dataset and/or synthetic dataset may be used and disclosed (Data Distribution stage)**

13. The Data Protection Provisions under the PDPA do not apply to the collection, use or disclosure of fused and/or synthetic datasets that are anonymised, and such datasets can be used and disclosed for any purposes. This would include **the use and disclosure of fused and/or synthetic datasets that are anonymised before the Data Distribution Stage, as well as for research, data mining, data analytics, development of commercial products and services, developing Artificial Intelligence machine learning models, etc.**

[6] Refer to PDPC's Advisory Guidelines on the PDPA for Selected Topics on Anonymisation.

14. In general, data collaboration partners should assess the risks of re-identification of individuals from the resultant datasets. They should put in place effective measures to ensure there is no serious possibility of re-identifying individuals from the datasets and any other information that is (or is likely to be) accessible. If there is a possibility that an individual can be re-identified from the fused and/or synthetic dataset, the Commission will not consider that dataset to be anonymised, and consent would be required for the collection, use or disclosure of such datasets. Please refer to PDPC's Advisory Guidelines on the PDPA for Selected Topics (Chapter 3 on Anonymisation) for further information on assessing and managing the risks of re-identification of anonymised data.

END OF DOCUMENT