Foundation Working Group

Proposed Recommendations on De-identifying Information for Disclosure to Third Parties

The Foundation Working Group (FWG) engaged in discussions around protecting privacy while enabling sharing of the MDMR data. With the recommendations to add geo-location information to the MDMR, the resultant data in the MDM/R would be considered personal information. Therefore a framework for de-identifying personal information was determined to be required.

An expert panel presented and provided resources to the FWG in respect of the de-identification of personal information for disclosure to third parties to inform these discussions.

Data de-identification is not a static process, where a single set of procedures, techniques and tools is established and exactly repeated for every de-identification operation. Therefore, the following is a proposed recommended framework, based on accepted principles and practices, for de-identifying information. Each de-identification operation will be performed on a case-by-case basis, taking into account the purpose of the disclosure, the requestor, the intended users, the data sets that will be matched, and the risks of re-identification, among other considerations.

The proposed recommended data de-identification framework will be formulated in the following sections. This framework would need to be adhered to as part the implementation of a data de-identification service.

1. Definition: De-identified Information
2. De-identification Context and Assumptions
3. Principles for De-identification
4. Requirements for a De-Identification Process
5. Steps for the Data De-Identification Process

1. Definition: De-identified Information[1]

The definition of 'de-identified information' the FWG has adopted is based on the definition in *De-identified Protocols: Essential for Protecting Privacy* from the Privacy by Design Framework.

[1] Source: Cavoukian, Ann, Ph.D., and Khaled El Emam, Ph.D. "De-identification Protocols: Essential for Protecting Privacy." 25 June 2014. Web: https://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identifcation_essential.pdf

"De-identified information is information that cannot be used to identify an individual, either directly or indirectly. Information is de-identified if it does not identify an individual, and it is not reasonably foreseeable in the circumstances that the information could be used, either alone or with other information, to identify an individual."

2. De-Identification Context and Assumptions

The following context and assumptions are made for the de-identification framework:

- This framework only applies to de-identifying data contained in the MDM/R and matching it with data sets from other sources, that would be provided by the requesting party.
- All de-identification of the data will be done by the Data Custodian or its authorized agent. The IESO, as the Smart Metering Entity, is the Data Custodian of the MDM/R data.
- It is assumed that the Data Custodian will have the authority to collect, use and disclose the de-identified data. As such, customer consent will not be required for fulfilling any of the requests for de-identified data.
- The Data Custodian will define the services and terms and conditions of service, including the consequences for non-compliance, for fulfilling requests to match data sets and de-identify personal information.
- The data requestor and any other users will comply with the appropriate conditions of service.

3. Principles for De-Identification

The following principles are recommended for the de-identification of personal information:

- Enable appropriate collection, use, storage, and disclosure of identified and de-identified data while maintaining public trust, privacy, integrity, availability, and confidentiality.
- Restrict access to identifiable and de-identified data to authorized organizations and users for authorized purposes.
- Establish a documented, clear, transparent process for de-identification with objective measures of risk.

- Establish a documented governance structure with clear accountabilities of all parties involved, including consideration of requirements of those parties to fulfill those accountabilities.
- Manage risk, in accordance with an authorized framework and risk thresholds of providing data access. Minimize re-identification risks while meeting the legitimate purpose of disclosure, while acknowledging that risks cannot be entirely eliminated.
- Leverage available standards and best practices for the data de-identification process and techniques, as they evolve. Leverage appropriate systems and competent personnel for the design, implementation and operation of the de-identification process.

4. Requirements for a De-Identification Process

The following defines some of the requirements that must be included as part of the data de-identification process for every request that is processed:

- Define the roles, responsibilities, and scope of the data custodian, data requestors and recipients. [Canadian Institute for Health Information (CIHI) 5.1, UK Information Commissioner's Office (ICO) 8][2]
- Define the rules and procedures for prioritizing, managing and fulfilling/declining and logging all data access requests. [CIHI 5.1, ICO 8]
- Define criteria for determining what is an appropriate use of the de-identified information.
- Define when a Privacy Impact Assessment ("PIA") and/or a Threat/Risk Assessment ("TRA") are/is required and associated disclosure requirements.
- Define a process for evaluating data re-identification risks.
- Define comprehensive and enforceable data user-custodian agreements to obligate users to maintain the confidentiality of the information that they receive. The agreements should specify at minimum that the recipient of de-identified data be required to:
    - Not re-identify, or attempt to re-identify, or allow to be re-identified, any individuals
    - Not link any other data elements to the data without verifying that the data will remain de-identified

[2] Refer to the Appendix for the identification of these references.

- Have in place controls that provide reasonable assurance that only authorized organizations and personnel have access to systems and data and that de-identified data remains de-identified
- Agree that all personnel or parties with access to the information shall abide by all of the conditions of the agreement

- Periodically conduct a review of the process and methodology against current and foreseeable threats and known vulnerabilities. This review may include reviews of the design and operating effectiveness of the process, as well as embedding monitoring controls in the process. [ICO 3, WP22]

5. Steps for the Data De-identification Process

The de-identification process must include, but is not limited to, the following steps for each data request or request category:

- Determine whether the proposed use is an acceptable use of the data for each data access request. [CIHI 7.1, US Department of Health and Human Services (HHS) 2.8, ICO 2]
- Identify and classify data fields that are direct identifiers and quasi-identifiers. [CIHI 6.2]
- Perform an assessment of both internal and external risks [CIHI 3, HHS 2.6, CIHI 13, HHS Table 1, HHS 2.7]
    - Identify plausible adversaries and plausible threats on the data in question (i.e. what would be the consequences if the data were disclosed without mitigations?)
    - Determine re-identification risk threshold or tolerance
    - Evaluate re-identification risk including, but not limited to, the considerations below:
        - Consider what reasonable resources with reasonable ability could accomplish in attempting to re-identify the information
        - Consider not only what data is being provided, but also other data available or that may become available in the foreseeable future that could be used to re-identify the information
    - Establish acceptable risk threshold based on the specific requestor and recipients of the data

- Perform Data De-identification to Mitigate Risk [CIHI 6.1, ICO3, CIHI7.1, HHS 2.9, ICO, WP22]
    - Define a set of data transformations (e.g. generalization, perturbation, data masking) for de-identification to reduce the risk to an acceptable level
    - De-identify data by applying appropriate data transformations, considering current standards and best practices for de-identification techniques and potential threats.
    - Verify that the original data has been properly de-identified
- Perform Residual Risk Assessment of the De-identified Data [CIHI 8.1]
    - Assess residual risk and data usability after application of risk mitigations, considering plausible adversaries and threats
        - Has the acceptable risk threshold established for the specific requestor and recipients of the data been met?
    - Determine whether to fulfill or decline the request
        - Compare evaluated risk with the acceptable risk threshold
        - If decision is made to decline the request, provide an explanation of the reasons for declining to the requestor
        - If decision is made to fulfill the request, provide a description to the requestor of the de-identification method used to enable the recipients to make accurate use of the de-identified data, without exposing specifics that could lead to re-identification. [ICO 3, Working Paper 22 (WP22)]

APPENDIX

References

- Canadian Institute for Health Information (CIHI)
- Privacy by Design (Ontario's Information & Privacy Commissioner of Ontario) (www.privacybydesign.ca)
  - "De-Identification Developments", 2013.
  - "De-Identification Protocols: Essential for Protecting Privacy", 2014
  - "Manual For The Review And Approval Of Prescribed Persons And Prescribed Entities", 2010
- UK Information Commissioner's Office (ICO) Code of Practice
  - Principles and methods for de-identification
- US Department of Health and Human Services (HHS)
- Working Paper 22 (WP-22) from Federal Committee on Statistical Methodology
  - Specific techniques for de-identification, Cited in HIPPA