# 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 in order to facilitate access while protecting privacy.

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 (defined below), the intended Recipients (defined below), the data sets that will be matched[1], 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] While it is recognized that matching MDM/R data with other data sets at the source level is an important component of enhancing the usability of the data, who and how any matching services would be provided would be determined as part of the implementation of Foundation, should it go forward.

## 1. Definition: De-identified Information[2]

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.

> "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 related MDM/R Data Mart, and matching this data 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. For the Foundation Project, the IESO, as the Smart Metering Entity, is the Data Custodian of the MDM/R, and related MDM/R Data Mart, data. As referenced in this framework:
    - "Requestors" include any organization or individual submitting a request for data.
    - "Recipients" refer to any organization or individual who will gain access as a result of a request being fulfilled.
    - "Individual Users" refer to authorized individuals who will gain access to the data within the Requestor, Recipient, Data Custodian or their authorized agents.
- It is assumed that the Data Custodian will have the authority to collect, use and disclose the inputs to the de-identified data (the underlying personal information). The Data Custodian will comply with applicable privacy legislation in collecting, using and disclosing personal information. As such, customer consent will not be required for fulfilling any of the requests for de-identified data.

[2] 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

- 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 will comply with the appropriate terms and conditions, including protection of privacy, and in the event the data Requestor is permitted to disclose the provided de-identified information to other Recipients, it will include appropriate terms and conditions of service in its agreements with such Recipients.
- For the purposes of these recommendations around risk assessment, an example of a typical vulnerability, threat and adversary would be:
    - Vulnerability – e.g. inadequate controls for authorized Individual Users to log on to a system containing the de-identified and/or personal information.
    - Threat – e.g. authorized Individual User accessing de-identified and/or personal information in a manner that may compromise controls.
    - Adversary – e.g. unauthorized organizations or individuals seeking to compromise the integrity of the information.

## 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 de-identified data while maintaining public trust, privacy, integrity, security, availability, and confidentiality.
- Restrict access to personal information and to de-identified data to authorized Individual Users within the Data Custodian, Requestor, Recipients and/or their authorized agent, 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, and, if applicable, their authorized Individual Users.
- Define the rules and procedures for prioritizing, managing and fulfilling/declining and logging all data access requests.
- 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 Requestor-Data Custodian agreements to obligate data Requestors to maintain the confidentiality of the information that they receive. The agreements should specify at a minimum that the Requestor receiving the de-identified data:
    - Must not re-identify, or attempt to re-identify, or allow to be re-identified, any individuals
    - Must not link any other data elements to the data without verifying that the data will remain de-identified
    - Must have controls in place that provide reasonable assurance that only the Requestor's authorized Individual Users will have access to the de-identified data and that it will remain de-identified
    - Must agree that all of the data Requestor's authorized agents and its authorized Individual Users with access to the de-identified data shall abide by the confidentiality and privacy terms and conditions of the data Requestor-Data Custodian agreement to safeguard the de-identified information

- 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.

## 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.
- Identify and classify data fields that are direct identifiers and quasi-identifiers.
- Perform an assessment of both internal (such as accidental release of data, inadvertent access by employees) and external (such as the risk of successful external hacking of systems and databases, and risk of hackers, marketers, journalists to re-identify de-identified data) risks.
    - 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
    - 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
    - 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?
        - Document risk decision
    - 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.