CANON Canadian Anonymization Network

Case Studies and Design Patterns – The Practical Application of De-identification

OPC Contributions Program Submission

Contents

One-Page Summary 2
Project Description 3
Title 3
Overview 3
Problem Statement 3
Background 4
Project Phases 6
Project Phase 1 – Identifying General Categories of Use Cases 6
Project Phase 2 – De-Identification Case Studies and Analysis 7
Project Details 10
Target Groups, Anticipated Results and Expected Benefits 10
Key Deliverables 10
Budget 11
Timeline and Monitoring 12
Acknowledgement of OPC Contribution 12
Applicant Details 13

One-Page Summary

Data is increasingly recognized as a key driver for innovation. As stated by *Innovation, Science, and Economic Development Canada*, "Digital and data-driven technology is already empowering science, supporting innovation, and driving economic growth."[1] Of course, in order to leverage these benefits, individuals must be able to trust that their data are being used responsibly; just as the increased availability of data is critical to innovation, so too must the protection of individuals' privacy underpin a robust digital ecosystem.

Regulators (including both Innovation, Science and Economic Development Canada and the Office of the Privacy Commissioner of Canada), organizations and innovators of all kinds have expressed interest in the potential role of de-identification in enhancing privacy while encouraging innovation. However, there are certain challenges that must be overcome before we will see more widespread adoption of effective de-identification techniques. Such challenges include:

- Ambiguity in legal and policy norms and standards causing hesitancy by some organizations to de-identify and use data for innovative, economic and/or socially-beneficial purposes;
- A lack of technical guidance and/or operational resources resulting in insufficient capacity and inconsistent practices; and,
- Highly publicized re-identification incidents caused by a lack of understanding and/or implementation of de-identification best practices resulting in potential reputational risks for organizations.

De-identification as a field is similar to cybersecurity – strong, proven solutions exist and these solutions are constantly being researched and improved upon, but they are only effective when they are both known to, and properly implemented by, the organizations that rely on them.

This project will take steps towards closing this gap between theory and practice. It will proceed in two primary phases:

> **Phase One**: A level-setting exercise, which will include the development of a taxonomy for the primary use cases of de-identification along with other resources that will help to create a common understanding of the vocabulary, techniques and outcomes associated with de-identification.
>
> **Phase Two**: Working with members of the Canadian Anonymization Network, the development of a series of practical case studies which set out in-depth descriptions of how de-identification has been used in practice and for which purposes (identifying both successes and challenges encountered).

Lastly, as a key part of the analysis of Phase Two, we intend to develop a series of "design patterns" for de-identification, which set out high-level approaches that can be taken to commonly encountered data-sharing scenarios that might benefit from the application of de-identification.

[1] Innovation, Science and Economic Development Canada. (2019) *Strengthening Privacy for the Digital Age.* https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html

Project Description

Title

Case Studies and Design Patterns – The Practical Application of De-identification

Overview

Data is increasingly recognized as a key driver for innovation. As stated by *Innovation, Science, and Economic Development Canada*, "Digital and data-driven technology is already empowering science, supporting innovation, and driving economic growth."[2] The availability of data is critical to the health researcher working to prevent or cure disease, the government agency seeking to better serve its citizens, and the company creating the innovative products and services which create social value and drive the Canadian economy.

These and other benefits of data can be derived when organizations are able to make fuller use of the data in their custody and control, and even more so when data can be combined and made more widely available for use by multiple parties for economic and socially-beneficial purposes.

In order to leverage these benefits, individuals must be able to trust that their data are being used responsibly; just as the increased availability of data is critical, so too must the protection of individuals' privacy underpin a robust digital ecosystem.

As reinforced in ISED's "Canada's Digital Charter in Action: A Plan by Canadians, for Canadians": "[D]ata is a valuable resource helping to drive innovation, power machine learning, and improve services for Canadians. However, we must ensure that while we support the greater use of data we are also protecting the trust and privacy of Canadians."[3]

The Canadian Anonymization Network ("CANON") believes that de-identification provides promising opportunity to achieve both data availability and privacy protection, allowing for a privacy-respectful means of responsibly leveraging data for economic and socially beneficial purposes.

Problem Statement

There is clear interest among regulators in the exploration of the role of de-identification for enhancing privacy while encouraging innovation. This includes discussions of de-identification in Innovation, Science and Economic Development (ISED) Canada's 2019 white paper "*Strengthening Privacy for the Digital Age*". The OPC has also identified "state of the art techniques that respect privacy, such as the use of synthetic data, differential privacy and depersonalization" as an area of particular interest for its 2020-21 Contributions Program. At the same time, the Canadian Anonymization Network (CANON) has

[2] Innovation, Science and Economic Development Canada. (2019) *Strengthening Privacy for the Digital Age.* https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html

[3] Innovation, Science and Economic Development Canada. (2019) *Canada's Digital Charter in Action: A Plan by Canadians, for Canadians."* P. 9 https://www.ic.gc.ca/eic/site/062.nsf/vwapj/Digitalcharter_Report_EN.pdf/$file/Digitalcharter_Report_EN.pdf

been hearing from its members – as well as the broad array of organizations with whom CANON members interact in their day-to-day operations – a clear desire to implement strong de-identification practices to help facilitate the socially- and economically-beneficial, yet privacy-protective, use of data.

Though the technique has been widely used for years in the public and health sectors, there are certain challenges that must be overcome before we can see more widespread and consistent adoption in the private sector. Such challenges include:

- Ambiguity in legal and policy norms and standards causing hesitancy by some organizations to de-identify and use data for innovative, economic and/or socially-beneficial purposes;
- A lack of technical guidance, best practices and/or operational resources resulting in insufficient capacity and inconsistent practices; and,
- Highly publicized re-identification incidents caused by a lack of understanding and/or implementation of de-identification best practices resulting in potential reputational risks for organizations.

De-identification as a field is similar to cybersecurity – strong, proven solutions exist, and these solutions are constantly being researched and improved upon. However, de-identification is also only fully effective when the appropriate techniques are both known to, and correctly implemented by, the organizations that rely on them.

The ultimate objective of this project – and of CANON more generally – is to ensure that best practices for de-identification are both well-understood and followed within Canada.

To support this, we propose: (i) to undertake a level-setting exercise in which the vocabulary, techniques and use cases for de-identification are set out in a clear, concise manner; (ii) to develop a series of case studies which explore specific applications of de-identification techniques, examining both challenges encountered and benefits achieved; and, (iii) to develop a series of design patterns for the application of de-identification within commonly encountered scenarios.

Background

Before delving further into the details of this proposal, it is worthwhile to provide some background information on de-identification.

What is De-identification?

De-identification should be understood broadly as a process that transforms personal information into data for which there is no serious risk of re-identifying an individual, given the context in which that data will be processed.

This will generally include, but also go beyond, the specific process of removal and/or modification of “direct identifiers” (attributes that alone enable unique identification of an individual, such as name, address, or unique numeric identifiers) and “indirect identifiers” (attributes that, when combined with other data, enable unique identification of an individual).

Where identifiers are modified, rather than deleted entirely (because of the importance of the field to maintaining the informational value of the data set), various masking techniques may be used including but not limited to pseudonymization (e.g. replacing identifiers with codes), randomization (e.g. modifying attributes such that their new value differs from their true value in a random way) or aggregation (e.g. grouping values into ranges).

It can also include emerging data transformation techniques, such as the creation of synthetic data (a fictitious dataset which mimics the patterns and qualities of the original).

What is Risk-Based De-identification

An analysis of de-identification cannot, however, stop with an analysis of the transformation(s) applied to data. De-identification is, after all, a process which is intended to reduce the risk of re-identification to an acceptably low level[4] - and this risk is impacted by contextual factors such as the release environment (e.g. public release vs. release in a controlled environment to a trusted party), what external information is available to an adversary (and the potential that it can be combined with the de-identified dataset), and any incentives to re-identify data (e.g. data that is of high intrinsic value, high sensitivity, or for which an adversary would receive notoriety for re-identifying).

As put succinctly in ISED's "Canada's Digital Charter in Action: A Plan by Canadians, for Canadians": "[T]o truly be a nation of innovators, we must build a culture of innovation, one that embraces resilience and risk."[5]

This "risk-based approach" to de-identification has been recognized by regulators globally; CANON has collected these position papers at deidentify.ca/resources.

One particular framework for assessing risk-based de-identification that has been gaining acceptance globally[6] is called the "Five Safes." As summarized by Luk Arbuckle (CANON member) and Felix Ritchie[7], the Five Safes are:

> **Safe projects**: What are the data flows, is de-identification needed as a privacy protective measure?
>
> **Safe people**: Who are the anticipated data recipients, what are their motivations and capacity to re-identify, and who may they know in the data?

[4] Note that CANON takes the widely shared – including by the Supreme Court of Canada – that reducing the risk of re-identification to zero, while retaining informational value in the resulting dataset, is not practicable. Thus, information should be considered to be de-identified where there is no serious risk of re-identifying an individual.
[5] Innovation, Science and Economic Development Canada. (2019) *Canada's Digital Charter in Action: A Plan by Canadians, for Canadians."* P. 9 https://www.ic.gc.ca/eic/site/062.nsf/vwapj/Digitalcharter_Report_EN.pdf/$file/Digitalcharter_Report_EN.pdf
[6] See, for example, the Australian Computing Society (December 2019). *Privacy-Preserving Data Sharing Frameworks.* https://www.acs.org.au/insightsandpublications/reports-publications/privacy-preserving-data-sharing-frameworks.html (free with registration)
[7] Arbuckle and Ritchie (2019). *The Five Safes of Risk-Based Anonymization*. IEEE Security and Privacy, Volume 17, Issue 5. https://ieeexplore.ieee.org/document/8821469

Safe settings: What are the technical and organizational controls in place to prevent a deliberate attempt to re-identify or to prevent a data breach?

Safe data: What is the re-identification risk, considering the people and settings of the data environment, and what de-identification can be applied?

Safe outputs: Would sharing data be perceived as an invasion of privacy to data subjects, and are there any ethical concerns with how the shared data will be used?

Much like any privacy protective measure, de-identification should not be considered in the abstract; the situation in, and the data to, which it is being applied will have important impacts on its effectiveness. The Five Safes framework gives us a structured way to approach a risk-based analysis — and will form a key element of the case studies of de-identification developed through this project.

Risk-based de-identification is a broad topic, spanning a variety of techniques and applied across a myriad of situations. Unfortunately, this breadth can lead to confusion as to the intended use case(s) for de-identification and related practices, as well as to the meaning of various terms.

On the latter point, CANON has an on-going project to develop a lexicon of de-identification related terminology which draws on the experience of Canadian organizations, experts in the field, and relevant court findings – as well as CANON's library of international resources.[8] Building upon the development of this lexicon currently underway, CANON will undertake two subsequent phases of work as part of this proposed project.

Project Phases

Project Phase 1 – Identifying General Categories of Use Cases

In the first phase of the proposed Contributions Program project, we will identify, and validate, a number of concrete and practical use cases for deidentification. These categories will aim, to the extent possible, to be sector and technology-agnostic.

As a starting point, CANON has identified the following use cases:

- **Public release**: An organization de-identifies data before releasing it publicly.
- **Disclosure to a third party**: Data is put through a de-identification process before being provided to a third party (generally accompanied by contractual provisions, such as confidentiality and/or data usage agreements).
- **On-site access by a third-party**: Data is put through a de-identification process and then made available by the original data holder to a third party either on-site or through a secure, audited portal – subject to a confidentiality agreement.
- **Access via a Trusted Agent**: Data from multiple organizations are put through a de-identification process and provided to a trusted agent (such as a data trust), which either analyzes,

[8] Available at https://deidentify.ca/resources/

aggregates, or pools the data on behalf of the third party or makes it available to other parties (including the original organizations).
- **Internal R&D**: An organization puts data about its customers through a de-identification process and makes it available to internal research and development staff, separated from its business lines behind a physical, virtual and administrative firewall.

Organizations that opt to de-identify data – rather than, for instance, deleting it – have made the decision that (i) the data in question has value, and (ii) that value can be realized without the inclusion of personal information. Thus, it should be unsurprising that the primary categorizations for de-identification are likely to focus on the mechanisms by which that data is made available for further use.

The original list of use cases may be expanded – or sub-taxonomies developed, as warranted to allow for more in-depth exploration of technical criteria and constraints, based on our evolving analyses during the project.

We propose similar analyses for various types of outputs of de-identification processes (record-level; pseudonymous; non-identifiable; etc.), common applications of de-identification involving one or more use cases (open data; pooling of data from multiple sources; etc.) and emerging de-identification techniques (e.g. synthetic data).

In the first phase of the project, we will conduct a literature review and hold a series of initial interviews with CANON members to examine whether these categorizations are both meaningful and comprehensive. CANON has secured a speaking slot at the 2020 IAPP Canada Privacy Symposium, which will be used to further validate these use cases among a broader range of stakeholders.

The end deliverable of this Phase will be, in effect, a series of de-identification "flash cards" –visual and text-based descriptions of key concepts in the field. These will be a highly-accessible, yet meaningful, resource which supports the creation of a common understanding, and of an analytical framework moving forward. While a final design has not yet been developed, it is envisioned that these "flash cards" may contain a visual representation of the use case, a general understanding of the most appropriate applications of the use case, and/or a description of the additional protective measures which tend to accompany the use case.

Project Phase 2 – De-Identification Case Studies and Analysis

In Phase 2, we move from the abstract to the concrete to understand specifically how de-identification is being used by organizations. Building on the categories of use cases identified and validated in phase one, CANON will work with its membership – and other interested stakeholders[9] – to describe concrete instances in each identified category where de-identification has been used in practice.

This work is inspired by, among other things, a similar effort by the Canadian Institutes of Health Research (a project led by CANON Steering Group member Patricia Kosseim), which examined the need for using personal information for health research purposes, the rationale and intended benefits, the

[9] For example, for this project CANON has secured in-kind resources from PwC, who will describe uses of de-identification by their clients.

privacy risks and protective safeguards used, the governance structures and processes in place, as well as the various legal, policy and ethical challenges encountered, etc.[10]

The CIHR project set out 19 case studies, from which several common themes and trends were identified, forming the basis for the eventual development of national Best Practices for Protecting Privacy in Health Research. While CANON intends to adapt the specific format of its case studies to suit present purposes, we have included (for illustrative purposes only) links to the CIHR Case Studies, along with the resulting Privacy Best Practices, in order to demonstrate the intended process of development.

For our project, we intend to follow a similar structure, setting out for each case study:

- The rationale and intended benefits (including the socially- or economically-beneficial purposes being pursued);
- How the Five Safes (safe projects; safe people; safe settings; safe data; safe outputs) were addressed within the project;
- What challenges (technical, operational, legal and/or policy) were encountered (if any); and,
- What lessons were learned and can be transposed to other similar-type applications.

To allow for greater openness, organizations will be permitted to have the application(s) they describe (a) attributed to them; (b) attributed generically to their sector (i.e. "a financial services organization"); or, (c) non-attributed, with identifying characteristics removed and/or modified.

Once developed, these case studies – and an analysis thereof – will be published. The analysis will focus on:

- Why do organizations choose to de-identify personal information?
- Why is it important that de-identified data be allowed to be shared?
- What other controls are in place, above and beyond any transformation applied to the original dataset?
- What are the key areas of challenge and/or uncertainty with respect to de-identification?

Through this effort, we intend to accomplish multiple objectives. First, this comprehensive examination of multiple case studies will shed light on the process-based, and not outcome-based, nature of risk-based de-identification. While the popular narrative around de-identification will often focus on instances in which data was publicly released without any controls, these case studies should make clear that such a scenario is very much the exception, and not the rule. Rather, de-identification will generally include both the application of a transformation to the data as well as a series of controls around the transformed dataset.

Second, by identifying both the rationale for de-identification, and the challenges and uncertainties encountered by organizations seeking to apply de-identification techniques, regulators, organizations and industry groups such as CANON will be better able to both direct their resources to areas in which organizations require support, and to engage in productive dialogue.

Third, this collection of case studies will create the beginnings of a knowledge-base around the practical use of de-identification – what protections have proven most effective, how they were adapted across

[10] Canadian Institutes of Health Research. Secondary Use of Personal Information in Health Research: Case Studies, November 2002. https://cihr-irsc.gc.ca/e/1475.html

scenarios, how organizations were engaged to bring legal, operational, and technical staff in a process, and so on. As noted in the "About the Applicant" section, one of CANON's primary objectives is to create a Canadian community of practice among stakeholders that rely on effective de-identification for the success of their organizations and eventually develop a body of acceptable best practices. Open sharing of information and consistent understanding of specific contexts (for example, via case studies) is a critical first step toward this goal.

To review and validate the case studies and associated analysis, CANON proposes to host an online workshop in early-2021. It is anticipated that the audience for this workshop would include all CANON members, as well as organizations that have contributed case studies, and other interested stakeholders such as ISED and the OPC.

Design Patterns

As a key part of our analysis of the case studies, we will seek to develop a series of design patterns for de-identification, to help organizations approach de-identification going forward (until such time as a de-identification framework is developed).

A **design pattern** is a "general, reusable solution to a commonly occurring problem …. [A] description or template for how to solve a problem that can be used in many different situations."[11] Through this project, CANON proposes to develop a series of design patterns aimed at the most common scenarios encountered by Canadian private sector organizations who are looking to anonymization as a means of using or sharing data for a socially-beneficial purpose and in a privacy-protective manner.

By way of illustration, see https://privacypatterns.org/ hosted by the UC-Berkeley School of Information.

As design patterns set out an approach to a problem – rather than a specific instruction set or algorithm – they are highly flexible and can be applied by organizations of variable size and across a broad range of sectors. As well, because design patterns are focused on process rather than end-result, they are also adaptable and can be quickly modified to address any changes in a regulatory environment.

The design pattern process will also provide an opportunity to introduce new de-identification techniques and show how and where they may be integrated into (or replace) existing processes. For this project, we will focus on the use of synthetic data. Synthetic data is created by analysing a real dataset – identifying, for example, statistical distributions and relationships between variables – and creating a set of fictitious records which closely mimic many of the properties of the original. Done properly, this approach can strike the dual objective of retaining data value, permitting more open data sharing (and thus innovation), and strongly protecting individual privacy.

This phase will also allow for the possibility of identifying "anti-patterns" – commonly occurring solutions that are actually counter-productive.

As these design patterns will only be useful if they can be applied in practice, CANON proposes – as part of the aforementioned workshop – to present initial prototypes to organizations which may be best positioned to use them.

[11] https://en.m.wikipedia.org/wiki/Software_design_pattern

CANON will also commit to posting the design patterns – either as part of the broader deliverable report on de-identification case studies, or as an independent resource, as is found to be more appropriate - on its website, as well as the NIST Privacy Engineering program's collaboration space[12], to allow for continued discussion and development of these design patterns.

Project Details

Target Groups, Anticipated Results and Expected Benefits

CANON envisions this project having three primary audiences, each of which will see a different expected benefit:

- **Individual Canadians** will gain a greater understanding of why organizations may wish to share data in a privacy-protective manner, and of the extent of the protections associated with such activities.
- **Organizations** will (i) gain a better ability to communicate with both customers and regulators (via an increased in the shared understanding of terminology and context); (ii) be able to learn from the experiences of others in designing any future de-identification projects; and (iii) develop a concrete foundation from which to develop eventual de-identification best practices.
- **Governments** and **Regulators** will gain an increased understanding of the practical elements of the challenge before them in creating a policy environment that enables innovation while ensuring the protection of privacy.

Key Deliverables

This project will have the following deliverables:

- A set of "flash card"-type descriptions of common categories of de-identification use cases based on common language and understanding of context.
- A series of draft concrete case studies on deidentification, risks and benefits, challenges faced and lessons learned, etc.
- An online workshop – in which interested stakeholders, including the OPC and other relevant regulatory bodies will be invited to participate – to refine the case studies, discuss preliminary findings and test the effectiveness of prototype design patterns
- A final report, containing:
    - A minimum of eight case studies of practical applications of de-identification
    - An analysis of the overall learnings from the case studies
    - A minimum of three design patterns, which describe a high-level approach to common de-identification scenarios faced by organizations

[12] https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/browse

Budget

Salaries and Benefits

For this project, CANON proposes to hire a part-time researcher at the rate of $60/hr – 2/3 of which is requested from the OPC (and the other 1/3 which will be covered by CANON). At an average of 10hrs / week over the course of the project, the requested contribution from the OPC would be:

($40 per hour) * (10 hours per week) * (48 weeks total) = **$19,200**

This is the sole funding being requested from the OPC.

CANON's contribution to the researcher's time would cover the balance of $9,600.

***TOTAL REQUESTED CONTRIBUTION**:* $19,200

In-kind

CANON does not propose to compensate members for working with our research team to identify and document instances of the practical application of de-identification. Rather, we consider this to be part of an in-kind contribution to the project. Specifically, based on an anticipated average of 20 hours of involvement per case study, a modest industry rate of $100/hr, and a proposed 8 case studies developed, we value the in-kind contribution from CANON member organizations to be **$16,000.**

Other

As noted, to ensure the broadest possible participation – and noting that CANON members are spread throughout Canada – the proposed Case Study / Design Pattern workshop will be held online. Any incidental costs, including any costs paid to the selected web-conferencing service and the production and distribution of workshop materials, will be covered by CANON. We anticipate this to be approximately **$500**.

It is anticipated that the primary distribution of CANON's deliverables will be electronic; as such, we will produce a high-quality PDF version of the report. The anticipated cost of this production is **$2000**, which will be covered by CANON.

CANON will also have an abstract and executive summary of the final report translated into French. At an estimated $50/page, the anticipated cost for this is **$300**, which will be covered by CANON.

CANON will also cover any overhead costs.

Timeline and Monitoring

	Apr. '20	May '20	June '20	July '20	Aug. '20	Sept. '20	Oct. '20	Nov. '20	Dec. '20	Jan. '21	Feb. '21	Mar. '21
Phase One – Level Setting												
Initial Development												
Validation w/ Experts & Members												
Validation @ IAPP Canada												
DELIVERABLE												
Phase Two – Case Studies												
Identification of Case Studies												
Interviews / Research												
Preliminary Drafting and Analysis												
Preliminary Dev. Of Design Patterns												
Workshop												
Revision / Review												
Formatting / Translation												
DELIVERABLE												

In addition to the key deliverables in each phase, CANON will provide the OPC with quarterly status update reports.

Knowledge Translation Activities

As a growing industry association, CANON has a natural avenue by which to disseminate the results of this project to its membership which is composed of many of the largest Canadian data custodians.

In addition to the workshop planned above, and to ensure broadest possible access to the deliverables, we will also commit that:

- All deliverables created through this project will be made freely and publicly available to anyone via CANON's website, deidentify.ca
- Project deliverables will be linked to (and/or discussed) in an AccessPrivacy monthly update call
- CANON will seek additional opportunities to disseminate findings via webinars, presentations, etc.

Acknowledgement of OPC Contribution

An acknowledgement of the role of the OPC Contributions Program in funding this project will be included within any published material and/or presentations related to this project.

Applicant Details

Basic Information

Organization name: The Canadian Anonymization Network
Address: 100 King St. W., Suite 6200
1 First Canadian Place
Toronto, ON M5X 1B8
Billing address: <same>
Fax number: N/A
Email address: info@deidentify.ca
Principal: Dr. Khaled El-Emam
…

Legal Status

CANON is a registered not-for-profit corporation (Corporation Number 1170112-6).

Organizational Background

The Canadian Anonymization Network ("CANON") is a not-for-profit corporation whose members include large Canadian data custodians from across the public, private and health sectors. CANON's primary purpose is to promote de-identification[13] in Canada as privacy-respectful means of supporting innovation and leveraging data for socially- and economically-beneficial purposes.

The overall objectives of CANON are five-fold:

- To share and exchange information about internationally-evolving legal, policy and technical standards on de-identification.
- To develop a Canadian community of practice among stakeholders that rely on effective de-identification for the success of their organizations across the public, private and health sectors.
- To educate the community at large about the effectiveness of de-identification methods, and meaningfully contribute to discussions about risks and opportunities related to this topic.
- To identify emerging issues and challenges with de-identification, include re-identification risks and legal or policy constraints and ambiguities.
- To advocate for legislative and policy standards for de-identification that enable innovative and beneficial used of data while protecting against privacy risks.

Since its formal launch in May 2019, CANON's membership has grown to include representatives from 35 member organizations, including many of the largest data custodians in Canada across the public, private and health sectors. General membership in CANON is free for any organization with large data

[13] A note on terminology: In this proposal, CANON uses the term "de-identification" to mean to any process which transforms personal information into data for which there is a very small risk of re-identifying an individual, given the context in which that data is processed. In some jurisdictions, this may be called "anonymization"; however, we find that "de-identification" is the more used term of art in Canada.

holding; inclusion in the CANON Steering Group requires a financial or in-kind contribution to the operation of the network. As of December 2019, the following are CANON member organizations:

- AccessPrivacy
- Alberta Health Services
- Bell
- BMO
- Canada Health Infoway
- Canadian Institute of Health Information
- Cancer Care Ontario
- Cryptonumerics
- E-Health Ontario
- Employment and Social Development Canada
- Georgian Partners
- Health Canada
- Health Data Coalition of British Columbia
- HITRUST
- IBM
- IMS/IQVIA
- Integrate.ai
- Manga International
- Metrolinx
- Microsoft
- Moneris
- National Bank
- Privacy Analytics
- PwC
- RBC
- Roche
- Rogers
- Statistics Canada
- SunLife
- Symcor, Inc.
- TD Bank
- Telus
- TransUnion
- Vancouver Coastal Health
- Waterfront Toronto

While, necessarily, this Contributions Program proposal focuses on de-identification in the private sector, our ready access to expertise from the public and health sectors will provide a significantly broader perspective on the topic and allow us to draw on the experiences of organizations for which the sharing of sensitive data for the public good is a legal and/or ethical requirement.

Project Team and Resources

The lead researcher representing CANON on this project is **Dr. Khaled El Emam**, senior scientist at the Children's Hospital of Eastern Ontario (CHEO) Research Institute and Director of the multi-disciplinary Electronic Health Information Laboratory (EHIL) team. Dr. El Emam is a world-renowned expert in statistical de-identification and re-identification risk measurement. He is one of only a handful of individual experts in North America qualified to anonymize Protected Health Information under the HIPAA Privacy Rule.

Dr. El Emam will be supported by:

- CANON Steering Group members which include: AccessPrivacy by Osler; Symcor, Telus, Bell, Georgian Partners, Moneris, Rogers, SunLife, Toronto Dominion Bank and Transunion;
- CANON General members comprised of large, innovative, and responsible Canadian data custodians across multiple sectors with unparalleled wealth of practical knowledge about the benefits, and challenges, associated with de-identification; and
- CANON's part-time researcher specifically assigned to this project.

Previous Financial Support

CANON has not received any previous financial support from the Office of the Privacy Commissioner of Canada.

Declaration of Conflicts

CANON acknowledges the following real and/or perceived conflicts:

- **Dr. Khaled El-Emam** is a Director at CANON, co-founder of Replica Analytics (which provides synthetic data services), sits on the board of a number of companies including Replica Analytics and Canary Medical (a medical device company), and actively invests in digital health technology companies. He is also an advisor (data protection technologies and AI) to a number of companies and government agencies.

By way of addressing this real and/or perceived conflict, we note that CANON is a vendor-neutral network; agreements with members, Steering Group members and/or sponsors do not, and will not, require the promotion of any particular vendor's technology, nor the use of any de-identification technique that is proprietary to a vendor. For this project, CANON further commits that its analysis will only describe de-identification techniques for which detailed public information (in the form of academic research, industry white papers, etc.) is readily available.

Patricia Kosseim is a Director at CANON, **Luk Arbuckle** is among the General CANON members, and **Vance Lockton** works as a part-time consultant for CANON. All three individuals are former OPC employees. Both Patricia Kosseim and Luk Arbuckle left OPC in January and April 2018, respectively. Should this project be funded, Mr. Lockton will not be engaged by the project team until after April 1, 2020 (the formal start of the funding period) at which time he will also be outside of the one-year period to which any post-employment restrictions would apply.