Reference	Comment	Proposed Change
About CANON	CANON is a not-for-profit corporation that supports a network of some of Canada's largest data custodians across the public, private and health sectors. Its mission is to promote and enhance de-identification as an effective privacy-enhancing technology. CANON believes that de-identification is an area which would benefit significantly from a code of practice or framework that: • Increases the public's confidence and trust in de-identification as a means of allowing responsible data use and sharing for innovative economic and socially beneficial purpose • Enables organizations to effectively minimize re-identification risks while still preserving the utility of data; and, • Allows for clarity and common understanding when: • Organizations describe their personal information management practices to individuals; • Organizations enter into contractual arrangements with processors based on an undertaking of de-identification; and, • Organizations seek to demonstrate to regulators that they are acting in an accountable manner and cite de-identification as a mitigating factor. We believe that there is an opportunity to work collectively with the CIO Strategy Council towards the creation of an accessible, operational resource that supports and enhances Canada' position as a leader in both privacy protection and in innovation through the use of de-identification. We are, therefore, pleased to provide comments with respect to the CIO Strategy Council's proposed part 3 to its Data Governance Standards: "Privacy enhancing data-identification framework" (the "Proposed Standard"). In addition to providing comments and suggested changes to specific sections of the Proposed Standard, we would like to make the following general comments:	
General Comments	The stated scope of the Proposed Standard is to "provide a framework for identifying and mitigating the likelihood that an individual is directly identified from data". We suggest that this explanation of "scope" be expanded to include further detail as to the intended users of this standard, as well as the types of data sharing situations that it is designed to cover. In addition, it would be helpful to add an explanation as to how this Proposed Standard fits within the broader group of data governance standards under development by CIOSC.	
General Comments	Although there is a bibliography attached to the Proposed Standard and notes throughout the document referencing other standards, no normative references are provided in Section 2. For example, the introduction section of the Proposed Standard says that "The standard provides organizations with an implementation framework to govern the appropriate use of data de-identification techniques"; however, standards like ISO/IEC 20889:2018 that specifically address de-identification techniques are not included in the "normative reference" section. We suggest that the committee consider adding the appropriate normative references.	
General Comments	The Draft Standard says that it is "intended to be used for conformity assessment". Key to conformity assessment are requirements that are measurable. It is unclear how a user would demonstrate in any objectively measurable way their compliance with a substantial number of the stated requirements. More detail is provided in our comments with respect to specific sections, but for example, section 4.1.1. sets out a number of factors that "shall be considered" as part of a contextual analysis. While good guidance might suggest that these are useful factors to consider when evaluating potential risk of de-identification, there is no way to objectively measure this "consideration".	
General Comments	Consideration should be given to whether this Proposed Standard can be applied to the most current practices that organizations use to generate, use and disclose non-identifiable data. CANON recently published a report for the Office of the Privacy Commissioner that describes a number of the most common practices for generating non-identifiable data.	

Overall	It is important that this document does not conflict with other standards and frameworks, and that it be interoperable with them. Alignment of language/definitions would be helpful.	~~Suggest conforming and coross referencing the terms and definitions section with the lexicon produced by CANON~~. Include some context or backgrouond as to how this standard interoperates with the existing ISO standard for de-identification. (20889) **Consider also lexicon created by CANON for potential inclusion in appendix.**
Overall	Generally, this document does not seem to refer to use by an organization that does not involve sharing or release. Is that the intention i.e. to not have this apply to the internal generation and use of non-identified data or is the intention to capture based on if it was disclosed/breached? The document is not express on this.	Consider scope of standard.
Overall	Align language with PIPEDA and other longstanding privacy guidance/standards.	Align requirements with Consent, Choice, Limit Collection, Limit Use and onwards transfer, accountability, redress, Limit retention
Overall	Missing discussion on temporal aspects of data and data sharing. For example in 4.1.1 - the enumerated list should include considerations such as whetheher the data is from an existing dataset, a historic dataset, going forward as a streming dataset, any timeframes associated with data sharing.	Include as a consideration how often assessments are performed based on changes in circumstances.
Overall	Datasheets for datasets provide a meaningful approach to characterizing datasets and should be considered for inclusion to help better characterize the attributes that should be consdered.	Include commentary on Datasheets for datasets
Overall	With an increasing use of data to fuel AI / ML / DL solutions it is essential that the document include discussion on the considerations associated with these technologies.	Document should include considerations for AI, especially AI exploits - Membership Inference, Model Inversion, etc.
Overall	Document does not consider business models or monetization techniques for shared data.	Document seems to pre-suppose that all data should be de-identified. There are legitimate cases where de-identification is not necessary. Suggest that this document requires a broader explanation as to how de-identification fits within the broader data governance practice.
Overall	Threat Modelling	Threat modelling should be further defined as should the IT Security Controls.
Overall	Threat Modelling	Threat modelling should also consider security, privacy and AI releated threats, which may also be threats to re-identification because these will help the average practicioner to properly manage risk in their overall threat modelling activities.
Overall	Ephemeral Consideration	The requirements shoud explicitly include ephemeral considerations - should include necessisty to consider agreement longevity, periodic review, expiry of agreements, etc.
Overall	The document contains many qualitative assertions for mandatory items which cannot be used for conformance.	Remove subjective mandatories e.g. Measure the trustworthiness of organizations by measuring disclosures - how does this impact hospitals impacted by Ransomware?
Overall	Privacy regulations permit consistent use of data. This standard does not include this concept.	Include accomodation / decision making for consistent use.
Overall	Adversarial testing presupposes limited access to large PII datasets. Recent data breaches have exposed billions of records which are available to advesaries in addition to other readily available resources.	Reconsider the guidance that assessor should attempt to consider what repositories are available to the data recipient. For public release of data, the assessor should consider that broad identifying information is available.
Overall	Document largely avoids legal authority to collect or share data.	Include legal authorities.
Overall	Notes do not neccssarily complement the text.	Edit the notes and consider if the note should form part of the direction in the standard.

Overall	Readability in document sections should be improved. It is difficult to interpret how the component subitems fit together to build a comprehensive view of how the guidance fits together.	Add additional narrative to the "General" sections (4.1, 5.1, 6.1) to describe the upcoming sections so that implementers can appreciate how the guidance fits together.
Scope and Definition		
Page i	The word "Framework" in the title is ambiguous.	Consider alternative title. Perhaps Risk Manegement Process is a better alternative.
page ix	Scope of the standard is unclear	The scope of the standard should be further defined to define the intended: 1. Audience - Is this for implementors, regulators, auditors, etc? 2. Types of data handling - While the text suggests sharing, should this direction be used for broader data handling? Does it cover internal sharing, external sharing, sharing between businesses, sharing between government and business, and release to open data communities? 3. What kinds of data is being addressed? Is it standard tabular data/statistics? image data? Video Data? Streaming data? 4. Does the guidance cover public data that may have been retrieved and used by the organization?
page ix	Please see this section of the introduction: "In almost all cases de-identification requires, at the very least, an evaluation of the external information available to an individual or group that may inappropriately reveal or uncover personal information (which we will refer to as an adversary, whether an individual is identified intentionally or not), and how they may combine it to reveal or uncover personal information. In short, removing personal information from data requires: . . ."	Need to include/build in the context/risk-based approach i.e. information for which there is no serious possibility in the circumstances that it could be associated with an identifiable individual. This seems to be focused on what information externally could possibly exist and be combined without an express reference to the reasonable likelihood.
Normative References	Extracts from De-identification Guidelines for Structured Data, ICO of Ontario, 2016 appear to have been used for mandatory requirements.	Include the guidelines as a normative reference.
Definitions	Many definitions are drawn from external references. These references should be included in the bibliography.	Include full refernces in the bibliography.
Definitions	Definition of "custodian": This definition includes a reference to Electronically Stored Information which is a defined term in an ISO standard - getting to this definition is difficult since the relevant ISO standard is not easily accessed and in a lot of cases has to be purchased. We recommend having a definition for electronically stored information within the document.	Craft a definition that is independent of ISO standards, but that aligns with , or is interoperable with, the ISO standard.
Definitions	Definition of personal information	"person" is a defined term, so no need to say "natural person"
Definitions	Definition of privacy impact assessment	Suggest the following: overall process of identifying, analyzing, evaluating, consulting, communicating and planning the treatment of potential privacy impacts ~~with regard to the processing of personal information,~~ which may be framed within an organization's broader risk management framework
Part 4, Context Assessment	De-identification is one mechanism in support of an organization's overall data governance structures and processes. Reference should be made to an organization's broader data governance program. This will not only remove duplication across the data governance series of documents but also avoid the potetnial for contradictory guidance between documents.	Suggest drafting a definition that is independent of ISO standards but that aligns/is interoperable.
Part 4, Context Assessment	Document suggests that sharing party has knowledge of all the capabilities of the "recipient". This is not possible.	Reconsider the expectation on the sharing party to understand motives,capabilities, intent of the recipient.

Part 4, Context Assessment	There is significant overlap between Section 4.1 and 4.2.1, making section 4.2.1 redundant.	Remove redundancies.
4.1.1	Include the broader context of these activities in the context of data section.	Replace first sentence with the following: "The organization shall evaluate the context of data sharing or release to help properly scope the application of safeguards including de-identification."
4.1.1	Should another consideration be the incentive to gain access to the data. Can this data be sold for attractive amounts? Can the data provide sigificant leverage or benefit to somenoe?	Consider adding additional consideration.
4.1.1.a.	Environment is not defined	define "environment" to scope what organizations must consider. Is this a compute environment, a business context, geographical view, etc?
4.1.1.a.	should this also say: "or could be released" or "will be used"	
4.1.1.b.	note typo: "come"	reword - The characteristics of the data including context of collection, legal authorities, accepteable use, etc. Suggest leveraging Datasheets for Datasets here. https://arxiv.org/pdf/1803.09010.pdf. Replace "come from" with "orginated".
4.1.1.d.	"nature of data" seems too broad	change to "sensitivity of data"
4.1.1.d.	Ethical considerations are not defined.	Add definition for what ethical considerations shall be in scope. The drafters may consider fairness, security/privacy, safety, inclusiveness, accountability and transparancy.
4.2.1	Include the broader context for threat modelling.	Suggest the following replacement for 4.2.1: "As part of an organization risk management program, organizations shall evaluate security and privacy risks for data. When de-identification is considered as a safeguard, organizations shall iclude the following in their assessment:"
4.2.1.a	Determining an organization's or individual's "motive" is unmanageable.	remove "motive" and replace with "intended use"
4.2.1.b	Determining with any certainty the other data sets that organizations have access to is not realistic in most instances. This is increasingly difficult for the public space (e.g. OpenData) due the to many large data sets that have been stolen and shared.	Reword to "Consider any unique cabilities the organization may hold".
4.2.1.c	The term "infrastructure" suggests the technical environment for the recipient. This seems out of place with the governance structure.	Remove "governance structures" from subsection 4.2.1.c and create new subsection 4.2.1.e that addresses "governance" separately.
4.2.1.c	Infrastucture Controls/ IT Security Controls	Suggest that in place of evaluating the controls in the destination domain, the sharing organization shall define the IT Security controls required in the destination domain.
4.2.1.d	alignment with privacy language	Reword to "the nature of the data and the ~~potential~~ impact ~~if shared or relased~~ as a result of onwards transfer or release."
4.2.2	Include the broader context of threat modelling. Organizations already use multiple threat models. What is specific to de-identification?	Replace 4.2.2 with: "As part of their risk management processes, organizations shall use a structured threat modelling apporach such as STRIDE, PASTA, Trike or MAL. Threat modelling for deintification shall consider the following unique threats:"

4.2.2.a	Internal? "Environmental" is also deliberate in many cases.	Should we add a combination of both deliberate and accidental threats as a consideration" For example, consider the situation where a former employee who was an insider, or who still has access to insiders, can work to create an environmental threat.
4.2.2	Given that "threat modelling" is already defined, this is confusing.	track the definition here.
4.2.3	See comment re 4.2.1.a - Determining an organization's or individual's "motive" is unmanageable.	reword to "a recipient's ~~motives~~ ability to …."
4.2.3	Scope of use is perhaps the overriding obligation	reorder to move line e to line a. Reword item e to "Definining scope of the agreement and acceptable use of the data."
4.2.3		Replace "controlled" with "mitigated"
4.2.3	This section assumes that disclosure is to a third party organization. It does not contmplate the use of information within an organization.	Should the standard consider "recipients" within an organizaiton? Note that "data recipient" is a defined term. Change "recipient" to "data recipient"
4.3.1	Include reference to the broader data governance structure . Also consider and align with current obligations under law. This section seems much broader than current requirements.	Reword to "As part of an organizaton's risk management process, organizations shall assess the impact of disclosure of the data. Specific considerations added by the use of de-identification techniques include:…"
4.3.1	Consider the overlap with 4.1.1. and 4.2.1	Remove overlapping provisions
4.3.1.a	Untestable requirements - The is no measurement for "as possible" or "where practical".	Reword this requirement to be testable.
4.3.1.b		define the terms "highly sensitive", "intimate" and "vulnerable populations"
4.3.1.c.	Trustworthiness and motivations are not testable.	Reword to include measureable characteristics - this may include certifications, legal authorities, violations.
4.3.1.c.	Previous privacy breaches are not indicative of an organization's current capacity to safeguard data.	Remove "and previous privcy breaches."
4.3.2	This section is impractical, especially from a business perspective.	Reword first sentence to - "The organziation should conduct an environmental scan of current information sharing practices, including legal and/or data protection authority decisions."
4.4.1	CIOSC 101 is not in the references	Include CIOSC 101 in the references
4.4.1	This statement does not call out how de-identification would fit in to a ethics board or framework.	Add wording to enumerate how deidentification should be incorporated into an organization's ethics program.
4.4.1	What is the intent / obligation? It is not practical to approach an ethics board for every decision realting to assessing data recipients.	Need more information on expectations.
5.1.1	Another goal of de-identification is to reduce privacy realted reisks and reduce the scale of mitigation strategies that are necessary.	Add these goals to the draft.
5.1.1	Nonstandard language. Missing the terms that are common among privacy practicioners. For this to be useful for them, it should use terminology that is familiar to them.	Suggest using terms for generally accepted privacy principles that have made their way into PIPEDA, GDPR and Privacy Act (e.g. Limit disclosure)

5.1.1		reword - "Organizations shall conduct a technical analysis of the data to be shared. Since complexity has a direct impact on the analysis, organizations should reduce the complexity of datasets using techniques such as severing."
5.1.1	Reword to call out the mandatory requirement in the last sentence of the second para	reword to "The organization shall evaluate features of the data that are central to identifiability **as described in this section**"
5.2.1	Awkward wording - "Organizations should have a sense of…"	reword to " Organzations should assess the external background information available to adversaries, including but not limited to…"
5.2.1	Public and private data sets - Assume access to all	This section is unrealistic given the legitmate ease of accessibility to PII data banks (clearances, D/L, etc) as well as the existance of dictionaries from data breaches. Reconsider the philosophy behind considering repository availiability.
5.2.1.d	Seems to duplicate above requirements like consideration of "privacy related harms"	remove
5.2.2.	The treatment of data / notes considering "vulnerable" populations requires reconsideraton	This is covered in section 4.3. Remove
5.3	Data type	The data types (and the entire document) suggest static data (and tabular and does not consider streaming data, video, photography, DNA, HCI info, etc
5.4.1.c	Should include "identifying" - this information is both identifying and sensitive	add "identifying"
5.5.1	Careful is not testable.	remove "careful"
5.5.1.a	Data Quality is not defined.	Define "Data Quality"
5.5.1	Data characteristics	Consider adding broader types of data - for example, any gaps in the data, pseudonymized elements, pre-processing, meta data aspects, etc.
5.6.1	This mandatory requirement is vague and requires additional qualification. What characteristics of the types of attack are evaluated? What context is required for the rationale? What is the form of the results?	Add further detail to this madatory requirement.
5.6	data availability	The process should assume that outside adversaries have full access to broad dictionaries.
5.6	Threat actors	In Data sharing, more emphasis should be placed on the shared "orgnization" malicious insider, malicious organization and (finally) the standard of care, before publishing to open data/research.
5.6.1	Attack modelling is very narrow and should also consider other adversaries.	Provide further explanation as to types of adversaries listed here.
5.6.1	The table requires additional context to be understandable and to be relevant for this de-identification document. As it stands, it simply reinforces the definitions of prosecutor and journalist attack. It is a truism that uncontrolled data sharing is higher risk than controlled data sharing.	Provide further explanation as to how this table assists in de-identification.
5.7	Organizations do not traditionally enumerate risks as maximums or averages.	Reframe risk levels to traditional "high", "medium" and "low".
6.1.1	This requirement is not measurable.	Reword to allow for repeatable testing.
6.1.1	What are the "methods for assessing identifiability" referenced here?	Provide standards-based references for methods. Potentially add 6.1.2 - Organizations should use identifiability assessments methods as described in reference X, Y, Z.

6.2.1	Determining "all the sources of the data that the would-be adversary might have access to" is not realistic. This is increasingly difficult for the public space (e.g. OpenData) due to the many large data sets tha have been stolen and shared.	Redraft to make requirement achievable.
6.2.1	Requirement to "document in a framework of plausible events." What is a "plausible event"? Are these events that exceed a percentage liklihood or do outlier events need consideration as well?	Define "Plausible Event"
6.2.1	When exploring the potential datasets available to a recipient organziation, are sharing organizations required to look at individual business units or the entire organization? For example, when sharing with government, do organizations consider all of government or departments or branches? Do they consider legal firewalls of business lines? (e.g. tax act, stats act, etc) If sharing between businesses, does the sharing organization consider one part of the recipient business or the broader business (e.g Sidewalk, google, GCP). The same could be said for banking - consumer banking, consumer lending, business banking, investments, etc.	Provide clarifying language about how assessments are to be conducted in addition to the scoping of the work to be performed.
6.2.1	The guidance in the note is dangerous. Just because "other organizations have been releasing similar data for a while without any apparent problems" does not necessarily mean that there is a reduction in risk. Consider Clearview case in 2019.	Remove this note.
6.2.2	Data traditionally doesn't have vulnerabilities - Data has characteristics that can lead to uniqueness and identifiability.	Reword to "…avoid focusing too closely on the characteristics of the data."
6.3	"well-established benchmarks involving strong precedents for selecting an indentifiabililty threshold" are not measurable.	Redraft to include testable measures.
6.3.2 a, b, c, d	How do these questions relate to the lead para (6.3.2)? Do these form the complete set of considerations for which subjective ratings will be provided?	Link the considerations more clearly to the mandatory requirement if necessary.
6.3.2	Which individuals are considered here? Are these data subjects? Are they organizational persons? The recipient org?	Clarify the subject to which the benefits should be considered
6.3.2	Benfits to society is a broad consideration which can be open for debate	Remove benefit to society question or remove "shall".
6.4.1	Adversary testing, while well advised, needs additional scoping and completion. Section does not address the different scenarios for data sharing, internal to business, B2B, B2G, Release. Also requires additional characterization of the expected available datasets	This section should be reworked to consider the different sharing models (internal, external, to gov, release) as well as the evolving information environment of available datasets. Approach will be different depending on sharing scenario.
6.5.1	Mitigation language suggests an iterative approach for remediation. This is consistent with traditional security and privacy risk assessments. This lifecyle approach should be reflected throughout the document.	Add in the lifecycle appraoch throughout the document - requirements->design->deployment-> maintenance-> refinement->retirement
6.6.1	Security / Privacy model is dated	Update security / privacy model to modern tools and environments (cloud vs on prem)
6.6.1.c	No definition for "published"	add definition for "published"
6.6.1.d	Information will be shared between orgnizations (and people therein)	Info is shared between legal entities wirh people bound by organizational agreements. Reword and add in legal arrangements.
6.7	"changing the data"	This section should be reworded. Also, the document must reconsider the temporal aspects of any information sharing process - initiation, negotiation, agreement, sharing, review, discontinuance.

6.7	See comment on the lifecycle approach for section 6.5.1	Add a lifecycle / iterative appraoch to the data safeguards.
6.7.1	Too much focus on the data	Should focus on the solution
6.7.2	Misses safeguards / assists	Should include links/pointers to mechanisms such as those included in ISO 20889:2018
Part 7	governance	This section should form part of an idependent standard and not be hidden in a de-id standard. Data governance includes privacy, security, responsible AI, ownership, business value etc.
7.5.1	This should synch with legislative requirements.	Review with legislation
7.6.1	Requirement is broad and impractical. Also may conflict with legal requirements.	Remove