International Journal of Population Data Science

Journal Website: www.ijpds.org

Essential requirements for the governance and management of data trusts, data repositories, and other data collaborations

P. Alison Paprica[1,2,3,*], Monique Crichlow[4], Donna Curtis Maillet[5], Sarah Kesselring[3], Conrad Pow[6], Thomas P. Scarnecchia[7], Michael J. Schull[2,8], Rosario G. Cartagena[2], Annabelle Cumyn[9], Salman Dostmohammad[10], Keith O. Elliston[11], Michelle Greiver[12,13], Amy Hawn Nelson[14], Sean L. Hill[15], Wanrudee Isaranuwatchai[1,16,17], Evgueni Loukipoudis[18], James Ted McDonald[5], John R. McLaughlin[19], Alan Rabinowitz[20], Fahad Razak[1,8,21], Stefaan G. Verhulst[22], Amol A. Verma[1,8,21], J. Charles Victor[1,2], Andrew Young[22], Joanna Yu[15], and Kimberlyn McGrail[3,23,24]

Submission History

Submitted:	12/03/2023
Accepted:	16/06/2023
Published:	20/09/2023

[1] Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, Ontario, M5T 3M6
[2] Institute for Clinical Evaluative Sciences (ICES), Toronto, Ontario, M4N 3M5
[3] Health Data Research Network Canada, Vancouver, British Columbia, V6T 1Z3
[4] Indoc Research, Toronto, Ontario, M5H 3W4
[5] New Brunswick Institute for Research, Data and Training (NB-IRDT), University of New Brunswick, Fredericton, New Brunswick, E3A 5A3
[6] Diabetes Action Canada, Toronto, Ontario, M5G 2C4
[7] Digital Aurora, Inc., Manchester, Vermont 05254
[8] Department of Medicine, University of Toronto, Toronto, Ontario, M5S 3H29
[9] Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1
[10] Public Health Agency of Canada, Ottawa, Ontario, K1A 0K9
[11] Seneca Creek Research, 209 Burlington Road, Suite 207, Bedford MA, 01730, USA (formerly affiliated with PHEMI)
[12] North York General Hospital, Toronto, Ontario, M2K 1E1
[13] Department of Family and Community Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, M4N 3M5
[14] University of Pennsylvania, Philadelphia, Pennsylvania, 19104
[15] Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, M5T 1R8
[16] Health Intervention and Technology Assessment Program (HITAP), Ministry of Public Health (Thailand), Daongmane, Mueang Nonthaburi District, Nonthaburi 11000, Thailand
[17] Knowledge Translation Program, St. Michael's Hospital, Toronto, Ontario, M5B 1W8
[18] Catalyx Tecnologies Inc., Vancouver, British Columbia, V6Z 3A4
[19] Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, M5T 3M7
[20] St Paul's Hospital, Vancouver, British Columbia, V6Z 1Y6
[21] Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, M5B 1T8
[22] The GovLab, New York University, Tandon School of Engineering, Brooklyn, New York, 11201
[23] Population Data BC, University of British Columbia, Vancouver, British Columbia, V6T 1Z3
[24] Centre for Health Services and Policy Research, School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, V6T 1Z3

Abstract

Introduction
Around the world, many organisations are working on ways to increase the use, sharing, and reuse of person-level data for research, evaluation, planning, and innovation while ensuring that data are secure and privacy is protected. As a contribution to broader efforts to improve data governance and management, in 2020 members of our team published 12 minimum specification essential requirements (min specs) to provide practical guidance for organisations establishing or operating data trusts and other forms of data infrastructure.

Approach and Aims
We convened an international team, consisting mostly of participants from Canada and the United States of America, to test and refine the original 12 min specs. Twenty-three (23) data-focused organisations and initiatives recorded the various ways they address the min specs. Sub-teams analysed the results, used the findings to make improvements to the min specs, and identified materials to support organisations/initiatives in addressing the min specs.

Results
Analyses and discussion led to an updated set of 15 min specs covering five categories: one min spec for Legal, five for Governance, four for Management, two for Data Users, and three for Stakeholder & Public Engagement. Multiple changes were made to make the min specs language more technically complete and precise. The updated set of 15 min specs has been integrated into a Canadian national standard that, to our knowledge, is the first to include requirements for public engagement and Indigenous Data Sovereignty.

Conclusions
The testing and refinement of the min specs led to significant additions and improvements. The min specs helped the 23 organisations/initiatives involved in this project communicate and compare how they achieve responsible and trustworthy data governance and management. By extension, the min specs, and the Canadian national standard based on them, are likely to be useful for other data-focused organisations and initiatives.

Keywords
data governance; data management; standards; public engagement; Indigenous Data Sovereignty; data trust; data repository; data collaboration; data stewardship

*Corresponding Author:
Email Address: alison.paprica@utoronto.ca (P. Alison Paprica)

https://doi.org/10.23889/ijpds.v8i4.2142

Introduction

There is growing recognition that significant public benefits could be realised through increased use, sharing, and re-use of person-level data for research and innovation [1–6]. For these benefits to be realised, data governance and management must ensure that data are stewarded properly, e.g., data are secure, privacy is protected and lawful, data are not misused, and appropriate technical controls and oversight mechanisms are in place [7–13].

Data subjects – the people whom the data are from and about – need assurance that there are system- and organisation-level controls to protect their privacy and other interests [14–22]. Organisations and initiatives that collect or disclose data (i.e., make it available or release it to another organisation or person [23]) need assurance that potential partners and data users fulfill essential requirements for responsible and trustworthy data governance and management [8, 11].

Data trusts, data repositories, and other data collaborations can take many different forms. Some are focused on particular areas of analysis or interest, such as smart cities, or birth cohorts, while others are deliberately broad. Some organisations/initiatives work with data that are obtained with consent from the data subjects and/or data that contain identifying information, others work with coded or de-identified administrative data collected and used without express consent from the data subjects. Some bring data together in a single location for analysis, others function as federated models where data remain in place and queries are sent to data. Some organisations/initiatives do not hold data themselves but play a role in helping data users understand where and how to access data held by other organisations. Despite this variety in form, our team contends that there are fundamental commonalities in functions to support responsible data use, sharing, and re-use in privacy-preserving and transparent ways that produce public value.

There are published frameworks and principles related to data governance and management including the Five Safes framework, the TRUST principles, and the FAIR principles [24–26]. Consistent with the definitions of "framework" and "principles," these resources provide fundamental structures and foundations for data governance and management systems. However, principles and frameworks do not necessarily translate into discernible requirements or standards, and it is not always possible to assess the compatibility of organisations for data sharing, data disclosure, or expanded uses of data based solely on the principles or frameworks they reference. Accordingly, our objective was to complement existing frameworks and principles by providing practical guidance that helps with the tangible aspects of responsible data governance and management by (i) identifying essential requirements that are (or should be) addressed by all data-focused organisations/initiatives, (ii) providing support to help organisations/initiatives understand how to address the essential requirements, and (iii) presenting examples of how organisations/initiatives can communicate practices related to the essential requirements to their stakeholders, including members of the public.

We are building upon a project undertaken from 2019 to 2020 by a team of people in Canada, including five of the co-authors of this paper. That team used a facilitated process to establish minimum specification requirements ("min specs") which are one of the "liberating structures" associated with complexity theory [27]. Min specs define essential requirements in the least prescriptive way possible in order to allow the maximum possible room for innovation and adaptation [28]. Combining first-hand experience stewarding data in Canada with a synthesis of concepts from the literature, the team published a peer-reviewed feature paper in 2020 which identified 12 min specs in five requirement categories: one min spec for Legal, four for Governance, three for Management, two for Data Users, and two for Stakeholder & Public Engagement [29].

There have been important developments since 2020, including increased use of the term "data trust" in academic publishing and movement toward using the term to refer to legal entities modelled on trust law, i.e., with trustees and beneficiaries [30–32].

Equally important is work focused on involving communities and community perspectives in policy and practice decisions about data. One rights-focused aspect of this is about supporting Indigenous Data Sovereignty and the right to self-determination consistent with the United Nations Declaration on the Rights of Indigenous Peoples and related legislation [33–38]. The CARE principles provide guidance for data from or about Indigenous Peoples and the First Nations principles of OCAP (ownership, control, access, and possession) have been developed specifically by and for First Nations [39–41].

There are also calls for race- and ethnicity-based data and other "disaggregated data" which have increased in the context of the disproportionate effects of COVID-19 pandemic on certain communities [42–44]. The Engagement Governance, Access and Protection (EGAP) Framework is an example of guidance for data from or about Black people [45]. The "Accessibility Ecosystem", co-developed with people with disabilities, includes a "Trusted Authority" for data governance that includes people with disabilities and a "Community Platform" to provide a simple and clear way for community members to contribute their knowledge, expertise and constructive criticism about accessibility in Ontario, Canada [46–48].

We acknowledge and support the data governance and management guidance that has been developed by communities and realised that the 2020 min specs were incomplete in that they did not direct data-focused organisations/initiatives to follow community-led guidance or involve colonised and/or historically marginalised publics in data governance and management.

Accordingly, based on these developments and understanding that the original 12 essential requirements would be improved through use, we initiated work on a second project to test and refine the min specs with a larger team, including people and organisations/initiatives from outside of Canada and people and organisations/initiatives focused on data from outside of the health sector.

Approach and aims

To increase the diversity of the types of organisations involved, all participants from the 2019–2020 project were invited to

participate and encouraged to invite additional organisations and individuals to join the team for this second project.

A kick-off meeting was held in April 2021 with 79 participants. The objectives of the meeting were to review the 12 min specs published in 2020 and discuss and agree on a plan which included:

1. Data Collection. Establish and use a template to collect information about how various data-focused organisations/initiatives address the 12 min specs.

2. Analysis. Establish five analysis sub-teams, one for each of the five min specs categories, to identify commonalities and differences in terms of how min specs were fulfilled and to identify revisions and additions to strengthen the min specs.

3. Synthesis of findings. Bring together and refine preliminary findings from the analysis sub-teams through meetings and collaboration on online documents.

4. Final outputs and knowledge translation. Preparation of this paper and supporting the integration of the min specs into a Canadian national standard – CAN/DGSI 100-7:2023, Data governance – Part 7: Operating model for responsible data stewardship – developed and published by the the Digital Governance Standards Institute (formerly the CIO Strategy Council) [49].

Work was led by a core team consisting of three Co-Principal Investigators [PAP, MJS, KM], plus the analysis sub-team leads [MC, DCM, KM, CP, PAP, TS] and a coordinator [SK]. In addition to core team meetings, there were 14 meetings between April 2021 and December 2021; two meetings of the whole project team, two meetings of the entire analysis team (48 people involved in completing and analysing templates), and nine analysis sub-team meetings. Recordings of meetings were made available to support the participation of people across multiple time zones.

Completed templates were received from the 23 participating organisations/initiatives identified in Box 1 and Appendix A.

Results

Analyses and discussion led to the articulation of 15 min specs (see Box 2 for min specs in short form, and the subheadings below, Table 1, and Appendix B for the full text of each of the min specs).

The total count of min specs increased from 12 to 15 because of the addition of one new Governance min spec focused on Indigenous Data Sovereignty, one new Management min spec focused on metadata and data documentation, and the division of a single Stakeholder & Public Engagement min spec into two separate min specs for stakeholder engagement and public engagement, respectively.

Among other changes, we stopped using the term "data trust" as an umbrella label for all forms of data infrastructure for two reasons. Foremost, our objective was to create guidance that could be applied widely, and there is a growing literature that uses the term "data trust" to refer specifically to legal entities based on trust law that have defined beneficiaries and trustees with fiduciary duties [30–32]. Secondly, members of the Health Data Research Network Canada Public Advisory Council had negative reactions to the term "data trust", noting the potential for it to be misinterpreted as being associated with financial services.

Our team was not able to identify any single term or label that would encompass the many different approaches and operational models for data use, data sharing, and data re-use that are in intended to be in-scope for the min specs. For example, the term "data repository" could be misleading for distributed analytics approaches that share code and queries instead of bringing data together. Therefore, the revised set of 15 min specs uses the language "data trusts, data repositories, and data collaborations" instead of relying on an uncommon use of the term "data trust." In addition to this change in language, we made multiple edits to make the language of the min specs more technically complete and precise. Appendix B summarises the modifications that were made to the original 12 min specs and the main reason(s) for each change.

The discussion of results below presents the 15 min specs alongside information about how they are being addressed by the 23 participating organisations/initiatives of this project. We also include examples of materials that can support data trusts, data repositories, and other data collaborations in addressing and communicating the min specs (see Table 1), a discussion of the overlap between the 15 min specs, and recommendations for implementation and future work.

Legal

1) The data trust, data repository, or data collaboration must fulfill all legal requirements including, as required, authority(ies) to collect, retain, use, disclose, and/or destroy data.

The foundational min spec is that a data trust, data repository, or data collaboration needs to comply with relevant legal requirements. The 23 participating organisations'/initiatives' testing of the min specs revealed that, in practice, legal authorities and constraints can take many forms including laws and regulations, governance documents (e.g., corporate objects, research ethics board approvals), management documents (e.g., binding terms and conditions in data sharing agreements, policies, processes, and procedures that address legislation compliance with respect to data privacy) and project-specific documents (e.g., the details of informed consent obtained from individuals participating in a research study).

We found that fulfillment of the legal min spec can be described as a 'tiered approach' to the law regime. Canadian organisations/initiatives that completed templates described a hierarchy, with applicable laws at the top, followed by the use of data sharing agreements, which are, in turn, reinforced through research ethics approval and upheld through local policy and procedures. For participating organisations/initiatives in the United States of America, the fulfillment of legal requirements was also addressed through a tiered approach that involved layers of agreements from a broad Letter of Intent (LOI) between partners, to Enterprise Memorandum of Understanding (EMOU) speaking to the nature of the data sharing, to Data Sharing Agreements with

Box 1: Names and locations of participating organisations/initiatives which tested the min specs

1. Actionable Intelligence for Social Policy (AISP) – USA
2. Centre for Addiction and Mental Health (CAMH) BrainHealth Databank – Canada
3. Canadian Partnership for Tomorrow's Health (CanPath) – Canada
4. Canadian Institute for Health Information (CIHI)– Canada
5. Canadian Research Data Centre Network (CRDCN)– Canada
6. Centre hospitalier universitaire de Sherbrooke (CHUS) Biobank – Canada
7. Digital Cardiac Health Platform – Canada
8. GEMINI – Canada
9. Hartford Data Collaborative (HDC) – USA
10. Health Data Research Network Canada (HDRN Canada) – Canada
11. Health Intervention and Technology Assessment Program (HITAP) – Thailand
12. ICES – Canada
13. Manitoba Centre for Health Policy (MCHP)– Canada
14. National Diabetes Repository of Diabetes Action Canada – Canada
15. New Brunswick Institute for Research Data and Training (NB-IRDT) – Canada
16. Newfoundland and Labrador Centre for Health Information (NLCHI) – Canada
17. Ontario Brain Institute – Canada
18. Ontario Health Data Platform, Queen's University site – Canada
19. PHEMI – BC Health Innovation Hub – Canada
20. Population Data BC (PopData)– Canada
21. Primary Care Ontario Practice-based Learning and Research Network (POPLAR) – Canada
22. Rhode Island Ecosystem – USA
23. ThinkData Works - External Data Catalog – Canada and the UK

Box 2: Fifteen min specs in short-form

1. Legal:
   Fulfills all legal requirements.
2. Governance:
   a) Includes a publicly stated purpose, and
   b) accountable governance body(ies), and
   c) is transparent, and
   d) acknowledges and respects Indigenous Data Sovereignty, and
   e) is adaptive and responsive.
3. Management:
   a) Policies, processes, and procedures cover the entire data lifecycle including
   b) cybersecurity and data protection, and
   c) risk management, and
   d) metadata and data documentation.
4. Data Users:
   a) Must complete privacy and security training, and
   b) acknowledge consequences for non-compliance.
5. Stakeholder & Public Engagement:
   a) There is ongoing engagement with stakeholders,
   b) including ongoing engagement with members of the public, and
   c) tailored engagement with subpopulations or groups that have a particular interest in, and/or that would be affected by, decisions or activities.

individual partners and finally Data Use Licences (DULs) addressing disclosure of de-identified data [50]. This tiered 'agreements approach' would be responsive to the sectoral nature of privacy law in the USA.

We also found that a single organisation or initiative might draw upon multiple authorities. For example, a single project or data repository could include (i) data collected and used with informed opt-in consent from the data subjects, (ii) data collected and used based on opt-out consent from the data subjects, and (iii) population-wide de-identified administrative data collected and used without express consent from the data subjects. It was also clear that the required legal authorities varied depending on the purpose and/or objectives of the data-focused organisation/initiative. For example, some

organisations did not have a role in data collection, and therefore did not have or need the authority to collect data, but did have the authority to use data for certain purposes and the/or authority to share data or provide access to it.

Given this complexity, we recommend that data trusts, data repositories, and other data collaborations involve legal experts in setting their policies, processes, and procedures when there is any question regarding whether a planned use or user of data would be lawful. We also recommend that organisations/initiatives demonstrate their commitment to min spec 1 by citing the specific statutory authority(ies) and other documents that serve as the legal foundation for data collection, retention, use, disclosure, and/or destruction (see Table 1 for examples and additional guidance).

Governance

2a) The data trust, data repository, or data collaboration must have a stated purpose that specifically addresses why its activities are necessary and/or beneficial.

2b) The data trust, data repository, or data collaboration must have an accountable governance body that is answerable for its decisions.

2c) The data trust, data repository, or data collaboration must be transparent about its purpose, governance body membership, data holdings, policies regarding who has access to what data for what purposes, and other information that is requested.

2d) The data trust, data repository, or data collaboration must acknowledge and respect Indigenous Data Sovereignty.

2e) Governance must be adaptive and responsive to risks, opportunities, and the concerns of stakeholders.

International studies have found that many members of the public see data as an asset that should be used for public benefit provided that risks are addressed, and specific conditions are met [14]. It is our view that governance is the best way to ensure that data trusts, data repositories, and other data collaborations meet all legal requirements and align with social licence.

The term governance can be used as an umbrella term that could cover every min spec category, e.g., as in the ISO 24143 definition of Information Governance as a "strategic framework for governing information assets across an entire organisation in order to enhance coordinated support for the achievement of business outcomes and obtain assurance that the risks to its information, and thereby the operation capabilities and integrity of the organisation, are effectively identified and managed" [51]. However, the Governance min specs we have articulated focus more narrowly on requirements that relate to the locus of accountability for decision making [10]. Accordingly, the main objectives of the Governance min specs are to ensure that there is at least one identified governance body that is answerable for its decisions and that the responsibilities of the governance body(ies) are clear, including to non-experts and the people whose data are being collected, used, shared and/or re-used.

Our testing of the min specs revealed that the original 2020 set of Governance min specs did not fulfill these objectives because the min specs were too vague or high-level. For example, to demonstrate how they were addressing the original Governance min spec focused on transparency, most of the 23 participating organisations/initiatives provided links to their websites without specifying where they had proactively published important information such as what kinds of data are held, who has access to which data for what purposes, or how people who have concerns or questions can have them addressed. On a related point, some organisations and initiatives identified a high-level accountable governance body, such as a university or hospital Board of Directors, that did not have an obvious connection to data-related policies and decisions.

For that reason, the refined Governance min specs provide less leeway and are more directional than the 2020 set of Governance min specs, e.g., in requiring that governance bodies are answerable for their decisions and clarifying the information that should be publicly available. This added specificity in the requirements is expected to make it easier for organisations/initiatives involved in data use, sharing, and re-use to know what to communicate proactively about their responsible and trustworthy governance practices (see Table 1 for additional guidance).

We also added a fifth Governance min spec focused on Indigenous Data Sovereignty [34, 35]. This new min spec, 2d, complements and goes beyond the Stakeholder & Public min spec 5c) which requires tailored and direct engagement with subpopulations and groups, such as American Indians and Alaska Natives in the United States of America and First Nations, Inuit, and Métis Indigenous Peoples in Canada. Consistent with the right to self-determination under the United Nations Declaration on the Rights of Indigenous Peoples, our team saw a need for a distinct Indigenous Governance min spec focused on Indigenous Data Sovereignty [33]. In practice, we expect this new requirement will require data trusts, data repositories, and other data collaborations to determine whether they hold data from or about Indigenous Peoples or communities, and if yes, to draw upon existing guidance as they work with, and take direction from, distinct Indigenous Nations and communities to establish data governance that supports Indigenous Data Sovereignty [38–41, 52, 53].

Management

3a) There must be well-defined policies, processes, and procedures covering the entire data lifecycle.

3b) There must be policies, processes, and/or procedures for cybersecurity and data protection safeguards which are reviewed and updated regularly.

3c) There must be policies, processes and/or procedures to identify, assess, and manage risks on an ongoing basis.

3d) There must be policies, processes and/or procedures to create and maintain metadata and data documentation which provides sufficient information for potential users to find, understand, use, and reuse data holdings.

In contrast with the Governance min specs which present requirements for strategic oversight and accountability, Management min specs focus on requirements for the day-to-day operations and operational decisions of data trusts, data repositories, and other data collaborations. For responsible data stewardship to be achieved, policies, processes, and procedures must be in place, e.g., to ensure that data custodians have appropriate controls and oversight mechanisms are in place to guarantee data security, privacy and intended use. The main addition to the Management min specs is an explicit requirement to have policies and procedures related to metadata and data documentation.

In practice, fulfilling the first management min spec, 3a, encompasses a lot of work, with min specs 3b, 3c and 3d calling out a subset of the many policies, processes, and procedures that are required for responsible and trustworthy data management. Our testing of the min specs revealed that, behind these management min specs, there are often numerous documents and multiple full-time staff responsible for setting up and overseeing policies, processes, and procedures related to the Five Safes, FAIR, CARE and other frameworks and principles [24, 26, 39]. For example, min spec 3d) will often be fulfilled by establishing a living online machine-readable data dictionary to make data "findable" as an important step toward fulfilling the "F" in the FAIR principles [26].

Given the breadth of activities that the Management min specs cover, communications about them must balance the need for transparency with not providing an overwhelming amount of detail. Several of the 23 participating/initiatives did this by stating a commitment to each of the Management min specs alongside a few examples of their relevant policies, processes, and procedures (see Table 1). We suggest that organisations/initiatives involved in data use, sharing, and reuse go one step further and, providing that it does not compromise data protection safeguards, make digital repositories of policies, processes, and procedures publicly accessible as several of the organisations/initiatives in Table 1 have already done. This practice could help members of the public understand how organisations/initiatives address risks, such as those related to privacy. Published repositories of policies, processes, and procedures would also help data trusts, data repositories and other data collaborations learn from each other so that they can identify and spread best practices.

Most of the 23 participating organisations/initiatives have policies, processes and procedures beyond those required by min spec 3b), 3c) and 3d). This is to be expected given the diversity of organisations/initiatives involved in this project and may reflect proportionate management controls based on the sensitivity of the data.

Data users

4a) Data users must complete privacy and security training before they access data.

4b) Data users must acknowledge that there may be consequences for non-compliance.

Data user min specs have been established because there will always be vulnerability to privacy and security at the point where data users interact with the data. Privacy and data security can be compromised because of insufficient data protection (e.g., users being able to re-identify individuals in de-identified datasets or project outputs), unintentional mistakes (e.g., using data for a purpose that is outside what was approved by a research ethics board), and/or deliberate malicious activities (e.g., users taking screen shots and using screen scraping techniques to transfer data that cannot be downloaded out of secure environments).

Most of the 23 participating organisations/initiatives directly or indirectly required privacy and security training for data users, and some also offered training on other topics (e.g., how to work with data). The majority of organisations/initiatives required proof that data users had completed privacy and security training provided by another organisation such as TCPS2 [54] or CITI [55, 56], while some had developed their own in-house privacy and security training with quizzes. Some organisations/initiatives required refresher training at prescribed intervals or when the content of the training changed.

Despite the fact privacy and security training for data users appears to be widespread among the 23 participating organisations/initiatives, we identified it as an area where there is room for significant improvements for two reasons. The first is that several team members reported that the current approach to training can become a "check-box" exercise through which the same data user repeats multiple, similar privacy and security training programs because that is easier than trying to make the case that previously completed training is sufficient. The second issue is that some existing privacy and security training focuses on making people aware of the details of laws and policies as opposed to testing how data users would act in situations that are likely to arise which pose threats to data privacy or security.

Our team discussed how scenario-based training could complement existing training materials and decrease the likelihood of (non-malicious) privacy and security breaches in practice. For example, in addition to requiring that health data users complete privacy and security training from TCPS2 or another authorised source, the Vector Institute presents potential health data users with realistic scenarios to clarify grey-zones and interpretation of data user requirements. For instance, the Vector training includes the question "Is it OK to provide a new team member with your login credentials while they wait for their paperwork to be processed?", with the answer being "No." Similarly, Sage Bionetworks' certification process requires data users to pass a 15-question quiz with scenarios to ensure that they understand their responsibilities and the rules and policies that govern data sharing on the Synapse platform [57, 58]. Scenarios in the Synapse quiz prompt users to demonstrate that they understand they are prohibited from attempting to re-identify individuals in datasets and that they must report suspected data breaches or data misuses promptly.

All but one of the 23 participating organisations/initiatives required (or has plans to require) that data users sign an agreement or, equivalently, acknowledge binding terms when they access data. However, only a small number of organisations/initiatives emphasise the consequences if data users do not comply with terms and policies. Our team identified this as a gap that needs to be filled by data trusts, data repositories, and other data collaborations because, in

Table 1: Examples of publicly available materials to support organisations/initiatives in addressing the min specs

Min specs category	Min spec	Organisations and initiatives with publicly available information to support fulfillment of the min spec
All Min Specs	The set of 15 min specs	Hartford Data Collaborative [61], ICES [62], NB-IRDT [63] Also, from organisations/initiatives external to project team: Canadian CIO Strategy Council [64]
Legal	1 The data trust, data repository, or data collaboration must fulfill all legal requirements including, as required, authority(ies) to collect, retain, use, disclose, and/or destroy data	AISP [50, 65], ICES [66], The GovLab [C4DC searchable library of data sharing agreements [67]]
Governance	2a) The data trust, data repository, or data collaboration must have a stated purpose that specifically addresses why its activities are necessary or beneficial	AISP [65], CIHI [68], ICES [69], GEMINI [70], MCHP [71], POPLAR [72]
Governance	2b) The data trust, data repository, or data collaboration must have an accountable governance body that is answerable for its decisions	AISP [65], CAMH BrainHealth Databank [73], CIHI [74], CRCDN [75], Diabetes Action Canada [76], ICES [77, 78], Ontario Health Data Platform [79]
Governance	2c) The data trust, data repository, or data collaboration must be transparent about its purpose, governance body membership, data holdings, policies regarding who has access to what data for what purposes, and other information that is requested	CIHI [80], MCHP [81, 82], Diabetes Action Canada [83, 84]
Governance	2d) The data trust, data repository, or data collaboration must acknowledge and respect Indigenous Data Sovereignty	HDRN Canada [85], ICES [86] Also, from organisations/initiatives external to project team: CARE Principles [39], First Nations Information Governance Centre OCAP®Guidance [34, 40, 41], Georgetown Center on Policy & Inequality [53], National Congress of American Indians [38, 52]
Governance	2e) Governance must be responsive and adaptive	HDRN Canada [87, 88], ICES [89] Also, from organisations/initiatives external to the project team: Ontario Government "AODA Accessibility Ecosystem" [46, 48]
Management	3a) There must be well-defined policies, processes, and procedures covering the entire data lifecycle	AISP [90], CIHI [91–93], ICES [62] Also, from organisations/initiatives external to the project team: Health Data Research UK Trusted Research Environments [94], Ritchie Five Safes Framework [24]
Management	3b) There must be policies, processes, and/or procedures for cybersecurity and data protection safeguards which are reviewed and updated regularly	AISP [65], CIHI [95], ICES [96]
Management	3c) There must be policies, processes and/or procedures to identify, assess, and manage risks on an ongoing basis	AISP [65], CIHI [97], ICES [96]
Management	3d) There must be policies, processes and/or procedures to create and maintain metadata and data documentation which provides sufficient information for potential users to find, understand, use, and reuse data holdings	HDRN Canada [98], Diabetes Action Canada [99, 100], ICES [101] Also, from organisations/initiatives external to project team: HDR UK [102], Maelstrom Research [103]

Continued

Table 1: Continued

Min specs category	Min spec	Organisations and initiatives with publicly available information to support fulfillment of the min spec
Data Users	4a) Data users must complete privacy and security training before they access data	AISP [56], CIHI [104], GEMINI [105], ICES [62] Also, from organisations/initiatives external to the project team: CITI [56], FAIR [106], TCPS2 [54]
Data Users	4b) Data users must acknowledge that there may be consequences for non-compliance	GEMINI [107], ICES [108] Also, from organisations/initiatives external to project team: Sage Bionetworks/Synapse [109] UK Biobank [59]
Stakeholder & Public Engagement	5a) There must be ongoing engagement with stakeholders.	AISP [65], ICES [110, 111], POPLAR [112, 113]
Stakeholder & Public Engagement	5b) Stakeholder engagement must include ongoing engagement with members of the public	CAMH BrainHealth Databank [73], Diabetes Action Canada [114], HDRN Canada [115–117], ICES [118, 119], The GovLab [120] Also, from organisations/initiatives external to the project team: ENGAGE(Québec) [121] NICE [122], UK Government [123]
Stakeholder & Public Engagement	5c) Where there is a reasonable expectation that specific subpopulations or groups would have a particular interest in, and/or be affected by, an activity or decision, there must be direct engagement tailored for that subpopulation/group	AISP [44], HDRN Canada [124], ICES [86], The GovLab [125] Also, from organisations/initiatives external to project team: EGAP Framework [45], First Nations Information Governance Centre [40, 41], National Congress of American Indians [38] Ontario Government "AODA Accessibility Ecosystem" [46, 48]

the absence of consequences, data user agreements/terms do not "have teeth" and may not be perceived as meaningful.

The most common consequences for data user non-compliance identified by participants in this project were (i) withdrawing privileges to access data and (ii) notifying other parties of the non-compliance (e.g., funders, data providers). This is consistent with the consequences in the UK Biobank material transfer agreement which notes that the data-holding organisation 'may prohibit the Applicant Principal Investigator and other researchers from the Applicant's Institution from accessing any further data; and/or, it may inform relevant personnel within the Applicant PI's Institution, funders of the Applicant and/or governing or other relevant regulatory bodies.' [59] Our team agrees that describing a range of consequences is appropriate because consequences should be proportionate to the sensitivity of the data and the nature of the non-compliance, i.e., different consequences for unintentional mistakes vs. malicious unauthorised use of data.

Generally, the 23 participating organisations/initiatives found it easy to communicate how they address Data User min specs in one or two sentences (see Table 1). As was the case for the Management min specs, there would be advantages if data trusts, data repositories, and other data collaborations went one step further and made training materials and data user agreements public in downloadable formats. Doing so would support learning across the data ecosystem and help members of the public understand how accountability with data users is achieved.

Stakeholder & Public engagement

5a) There must be ongoing engagement with stakeholders.

5b) Stakeholder engagement must include ongoing engagement with members of the public.

5c) Where there is a reasonable expectation that specific subpopulations or groups would have a particular interest in, and/or would be affected by, an activity or decision, there must be direct engagement tailored for that subpopulation/group.

In order to achieve their purposes, data trusts, data repositories, and other data collaborations rely on support, not to mention data, from their stakeholders. Trust and ongoing support from members of different publics is critical because most of the data that are held, used, shared, and re-used are data generated by the activities of people. In addition, data trusts, data repositories, and data collaborations have stakeholders with different needs and interests, including the individuals and organisations that use data, knowledge users, such as policymakers who act on evidence generated from data, and organisations that set laws or standards related to data.

Accordingly, we have articulated three distinct min specs focused on the ongoing engagement of different stakeholder groups. All of the 23 participating organisations/initiatives involved in this project reported that they fulfilled min spec 5a by engaging with stakeholders that collect and share data and/or stakeholders who are in a position to use the knowledge

generated based on the data holdings. The most common approach was to include people who had experience working in governments, government agencies, and universities in a Board of Directors, Steering Committee, Advisory Committee, and/or another governance body.

Similarly, and consistent with international best practices, many of the 23 participating organisations/initiatives reported they addressed min spec 5b) by including members of the public alongside other stakeholders in governance bodies, and/or by creating Advisory Committees consisting solely of members of the public. In addition, some team members shared examples of supplementary public consultations on specific topics.

A minority of the 23 participating organiations/initiatives have already taken steps to address min spec 5c) which requires targeted public engagement and involvement with groups and subpopulations that would have a particular interest in, and/or be affected by, certain activities and decisions. Where this practice was reported, it often referred to models for engagement and involvement of Indigenous Peoples, new immigrants, youth, or groups of patients (e.g., with a particular condition or with rare diseases) or historically marginalised populations. Several organisations/initiatives noted that they are in the process of developing or implementing equity frameworks to guide their data governance, management, and engagement activities.

As shown in the examples in Table 1, organisations/ initiatives can demonstrate that they address the Stakeholder & Public Engagement min specs by publishing information about their activities in plain language and creating easy ways for members of the public and stakeholders to get involved through open and transparent processes that encourage the involvement of people with different backgrounds, abilities, experiences, and perspectives. We also recommend that organisations make use of guidance, such as the International Association for Public Participation (IAP2) spectrum, which prompts thinking about the level of that engagement, how engagement will be supported, and how public input will be used to support the mission and vision of the data trust, data institution, or data repository [60]. Part of that is being transparent and specific about what they mean by "engagement" e.g., by providing examples of how they act on recommendations and advice, including from members of the public.

Discussion

Connections across and between min specs

Our analysis sub-teams identified several ways that min specs categories are complementary or connected. For example, the Legal, Governance, and Management min specs are connected in that data trusts, data repositories, and other data collaborations must acknowledge that the approval from their governance bodies is not, in itself, sufficient; management's policies, processes, and procedures must also fulfill legal requirements that are established externally.

There is also a connection between Governance and Stakeholder & Public Engagement min specs in that involvement of stakeholders and members of the public in governance bodies was a common way for organisations to achieve trustworthy governance. Similarly, there is a connection between Data Users and Stakeholder & Public Engagement min specs in that Data Users should be engaged in an ongoing manner as an important stakeholder group who can be drivers of innovation and continuous quality improvement.

Other connections included the fact that Data User training should include training on Indigenous Data Sovereignty, and that Management policies, processes, and procedures will often include requirements related to Data Users and/or Stakeholder & Public Engagement.

We do not view the fact that there is some overlap and complementarity between min specs as a weakness. Rather, these connections are an indication that the min specs reinforce each other to cover the complex reality of trustworthy and responsible data governance and management.

Implementation

We have a sense of the work effort required to apply the min specs from the 23 organisations/initiatives that tested them as part of this project. Generally, most of these organisations/initiatives found that they had existing materials related to most of the min specs, and that summarising how they address the min specs required a few days of work. Table 1 presents examples of publicly available materials to support organisations/initiatives in addressing the min specs and examples of how organisations/initiatives can communicate about their practices related to the min specs. While most of the materials are from the organisations/initiatives directly involved in this project, we have also included some materials from external organisations and teams. This is an indication of the generalisability of the min specs and the fact that the min specs complement vs. compete with existing frameworks.

Table 1 is not meant to imply that all organisations/ initiatives will immediately fulfill or address each of the 15 min specs. Notably, many organisations/initiatives are just beginning to address the new min spec 2d focused on Indigenous Data Sovereignty. Additionally, work to increase the equity of data use, sharing and re-use, as per min spec 5c, will require years of effort, and resources and support for colonised and/or historically marginalised communities that are establishing their own data governance and management requirements. Consistent with the guidance published in the 2020 feature paper, we continue to recommend that data-focused organisations/initiatives start by documenting, in simple plain language, how they address the min specs that they already fulfill, then begin work to address min specs that are outstanding.

We have already taken steps to socialise the min specs and support their use, including through the preparation of this paper and presentations at conferences. The min specs have also been included in formal, public, feedback provided to Canadian guidance that is being developed for data repositories [126] and have been incorporated into the curriculum of the Data Stewardship executive program offered to international participants at no cost by The GovLab [127].

Our team has also gone beyond traditional research knowledge mobilisation efforts by supporting work to integrate

the min specs into a new Canadian national standard, CAN/DGSI 100-7:2023, Data governance – Part 7: Operating model for responsible data stewardship [49]. Canada's Digital Governance Standard Institute (formerly the CIO Strategy Council) became aware of the min spec work from the 2020 publication, and subsequently joined the team for the second project and invited researcher members of this team to participate in the development of a standard based on the min specs. As is the case for all standards developed by the Digital Governance Standards Institute, the process for CAN/DGSI 100-7 was open (any interested party could join the technical committee to help develop the standard, and all members of the committee have a vote), transparent (drafts of all standards were posted online for feedback) and free (there were no fees to participate in the development of the standard, and published standards are free for non-commercial use). CAN/DGSI 100-7 was published in July 2022 and amended to improve its clarity in July 2023. To our knowledge, CAN/DGSI 100-7 is the first national standard to require respect for and acknowledgement of Indigenous Data Sovereignty and public engagement related to data.

This new Canadian standard may serve as a model for other national standards and/or as high-level guidance that complements existing ISO standards and other frameworks. Based on the 23 organisations/initiatives experience with the min specs, we identified some potential immediate uses and benefits including:

- Information about how an organisation or initiative addresses the 15 min specs could be converted into information on public websites, e.g., "Frequently Asked Questions" (FAQ).
- The min specs could help organisations understand, at a high-level, the practices and authority(ies) of potential data partners.
- The min specs could serve as the principles for a memorandum of understanding for data sharing between organisations.
- Canadian organisations/initiatives that record how they address the min specs may simultaneously demonstrate conformity with the Canadian national standard CAN/DGSI 100-7, which may increase their perceived trustworthiness by stakeholders, including members of the public.
- Conformity with the min specs, and CAN/DGSI 100-7, could be a first step towards conformity with ISO standards.

Limitations and future work

Though we consider the updated set of 15 min specs to be improved relative to those published in 2020, we expect the min specs to continue to evolve and be improved as data practices and technologies advance.

By design, there is a high threshold for a requirement to become a min spec. It is not enough to say that a new requirement might improve data-related practices, it would only become a min spec if it is not possible to meet the working acceptable standard for trustworthy and responsible data governance in its absence. Our team identified practices and requirements outside of the min specs which we believe should be encouraged, e.g., requirements for data traceability, the ability to execute withdrawal of data upon request of the data subject, and standards for reporting data quality. We will continue to track these suggestions and be open to expanding the min specs to include them, and additional essential requirements identified by other groups.

While we tasked ourselves with developing min specs that would have utility beyond the 23 participating organisations/initiatives, we also acknowledge that there are many perspectives that have not been incorporated in the set of 15 min specs published here. Most significantly, though the min specs include requirements related to Indigenous People and historically marginalised populations, our team does not have the expertise or knowledge to recommend detailed operational guidance for those Nations or communities. Our hope is that min specs 2d and 5c, which direct organisations/initiatives to engage, involve, and take direction from colonised and/or marginalised populations, will help create space, time, and resources for future community-led and co-led work on data governance and management for those populations.

Additionally, though we had more diverse, and more international, perspectives integrated into this updated set of min specs, the project team still consisted mostly of people in Canada and the USA that work in the public or not-for-profit sectors. As such, the current manuscript does not reflect the perspectives and approaches of people involved in data-focused organisations/initiatives in other countries. We anticipate that there would be benefit from future work to bring in the perspectives of more project team members from outside of Canada and the USA.

It is also important to learn if and how the min specs add value to commercial organisations that establish data trusts, data repositories and other data collaborations, including public-private partnerships. A subset of the members of our team has begun planning for a separate project that will bring together private sector organisations and people involved in data-focused public-private partnerships to test the 15 min specs and obtain advice on how they might be applied and improved to support use by companies.

Conclusions

Our review and updating of the work published in 2020 suggests there is a commonly agreed core set of 15 min specs for responsible and trustworthy data governance and management. Collectively, the publicly available materials identified or generated by this project demonstrate that there are many different ways the min specs can be fulfilled, provide a mechanism for data-focused organisations/initiatives to learn from each other, and provide transparency through multiple examples of how data trusts, data repositories and other data collaborations address the min specs.

Transparent communications about how the requirements are addressed is important because it can help enhance public trust and stakeholder support for uses of data. We see the potential for additional benefits to be realised if organisations/initiatives involved in data use, sharing and

reuse go one step further by providing downloadable policies, processes, and procedures with plain language content that helps stakeholders, data users and members of the public understand how organisations/initiatives support high-quality data use, sharing, and re-use in privacy-preserving and transparent ways that produce public value.

Our hope is that the min specs can be a useful checklist both for new data trusts, repositories, and data collaborations and as an ongoing touchstone for existing organisations/initiatives to communicate and address essential requirements as the min specs evolve.

Acknowledgements

Acknowledged individuals and organisations: Megan Ahuja, Kimberly Begley, Charles Burchill, Lisa A. Dietrich, Frank Gavin, Mary Horodyski, Della Jenkins, Theodore Konya, Denise Mak, Kirk Nylen, Matthew MacNeil, Kwame McKenzie, Parisa Osivand, Sujitha Ratnasingham, Donna Roche, Joseph Scheuhammer, Andrea Smith, Eric Sutherland, Jutta Treviranus, Jennifer D. Walker, Nicole Yada, the Black Health Equity Working Group, the Health Data Research Network Canada Public Advisory Council, and Sage Bionetworks.

This work was funded by the Strategy for Patient-Oriented Research (SPOR) National Data Platform grant from the Canadian Institute of Health Research (CIHR NDP-160882) and in-kind contributions from team members' employer organisations.

Statement of conlficts of interest

No conflicts to declare.

Ethics statement

This study did not require ethical approval as it did not involve human participants or the use of personal data.

References

1. The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, Council of Canadian Academies. Accessing Health and Health-Related Data in Canada. [Internet]. Ottawa ON: Council of Canadian Academies; 2015 [cited 2023 Jan 30]. Available from: https://cca-reports.ca/wp-content/uploads/2018/10/healthdatafullreporten.pdf.

2. Health Canada. Unleashing Innovation: Excellent Healthcare for Canada – Executive Summary. [Internet]. Ottawa ON: Government of Canada; 2015 [cited 2023 Jan 30]. Available from: https://healthycanadians.gc.ca/publications/health-system-systeme-sante/report-healthcare-innovation-rapport-soins/alt/report-healthcare-innovation-rapport-soins-eng.pdf.

3. Verhulst, Stefaan G., Andrew Young, Michelle Winowatan, and Andrew J. Zahuranec. Leveraging private data for public good [Internet]. New York NY: The GovLab;2019 [cited 2023 Feb 15]. Available from: https://thegovlab.org/static/files/publications/data-collab-report_Oct2019.pdf.

4. Task Force on Artificial Intelligence for Health (AI4Health). Building a Learning Health System for Canadians [Internet]. Toronto ON: Canadian Institute for Advanced Research; 2020 [cited 2023 30]. Available from: https://cifar.ca/wp-content/uploads/2020/11/AI4Health-report-ENG-10-F.pdf.

5. Hall, Wendy; Pesenti, Jérôme. Growing the artificial intelligence industry in the UK. [Internet]. London Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy; 2017 [cited 2023 30]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.

6. Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N. The other side of the coin: Harm due to the non-use of health-related data. International journal of medical informatics 2017 Jan;97:43–51. https://doi.org/10.1016/j.ijmedinf.2016.09.010

7. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. Journal of Medical Ethics 2015 May;41:404–9 . https://doi.org/10.1136/medethics-2014-102374

8. Mangravite LM, Sen A, Wilbanks JT, Team SB. Mechanisms to Govern Responsible Sharing of Open Data: A Progress Report. [Internet]. Manubot; 2020 [cited 2023 Feb 10]. Available from: https://sage-bionetworks.github.io/governanceGreenPaper/.

9. Cumyn A, Dault R, Barton A, Cloutier AM, Ethier JF. Citizens, research ethics committee members and researchers' attitude toward information and consent for the secondary use of health data: Implications for research within learning health systems. J. Empir. Res. Hum. Res. Ethics 2021 Jul.;16:165-78. Available from: https://doi.org/10.1177/1556264621992214

10. Khatri V, Brown CV. Designing data governance. Communications of the ACM [Internet] 2010 Jan;53:148–52. Available from: https://doi.org/10.1145/1629175.1629210

11. Verhulst SG, Zahuranec AJ, Young A, Winowatan M. Wanted: data stewards. (Re-) defining the roles and responsibilities of data stewards for an age of data collaboration [Internet]. New York: The GovLab; 2020 [cited 2023 Feb 10]. Available from: https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf

12. Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B. Bringing code to data: Do not forget governance. J. Med. Internet Res. 2020 Jul;22:e18087. https://doi.org/10.2196/18087

13. Knoppers, B.M. Framework for Responsible Sharing of Genomic and Health-Related Data. HUGO J. 2014 Dec;8:3. https://doi.org/10.1186/s11568-014-0003-1

14. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med. Ethics 2016 Dec;17:1–24. https://doi.org/10.1186/s12910-016-0153-x

15. Cumyn A, Ménard J-F, Barton A, Dault R, Lévesque F, Ethier J-F. Transparency and the Secondary Use of Health Data: A Scoping Review of What Should Be Communicated to the members of the public. How and at What Conditions. (forthcoming/in press) https://doi.org/10.2196/45002

16. Aitken M, Tully MP, Porteous C, Denegri S, Cunningham-Burley S, Banner N, Black C, Burgess M, Cross L, van Delden JJ, Ford E. Consensus statement on public involvement and engagement with data intensive health research. Int. J. Popul. Data Sci. 2019;4. https://doi.org/10.23889/ijpds.v4i1.586

17. Paprica PA, de Melo MN, Schull MJ. Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study. Can. Med. Assoc. Open Access J. 2019 Jan;7:E40-6. https://doi.org/10.9778/cmajo.20180099

18. Teng J, Bentley C, Burgess MM, O'Doherty KC, McGrail KM. Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada. Int. J. Popul. Data Sci. 2019 May;4:1103. https://doi.org/10.23889/ijpds.v4i1.1103

19. McCradden MD, Sarker T, Paprica PA. Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research. BMJ Open 2020 Oct;10:e039798. https://doi.org/10.1136/bmjopen-2020-039798

20. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. Sci. Public Policy 2016 Oct;43:713–23. https://doi.org/10.1093/scipol/scv075

21. Cumyn A, Barton A, Dault R, Safa N, Cloutier AM, Ethier JF. Meta-Consent for the Secondary Use of Health Data Within A Learning Health System: A Qualitative Study of the Public's Perspective. BMC Med. Ethics 2021 Jun;22:81. https://doi.org/10.1186/s12910-021-00647-x

22. Hill EM, Turner EL, Martin RM, Donovan JL. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. BMC Med. Res. Methodol. 2013 Dec;13:1–10. https://doi.org/10.1186/1471-2288-13-72

23. Information and Privacy Commissioner of Ontario. Disclosure [Internet]. Disclosure [cited 2023 Feb 10]; Available from: https://www.ipc.on.ca/health-organizations/collection-use-and-disclosure-of-personal-health-information/disclosure/#:...:text=The%20term%20"disclose"%20means%20to,health%20information%20custodian%20or%20person.

24. Ritchie F. The 'Five Safes': a framework for planning, designing and evaluating data access solutions. Zenodo [Internet] 2017; Available from: https://doi.org/10.5281/zenodo.897821.

25. Lin D, Crabtree J, Dillo I, Downs RR, Edmunds R, Giaretta D, De Giusti M, L'Hours H, Hugo W, Jenkyns R, Khodiyar V. The TRUST Principles for digital repositories. Sci. Data 2020 May;7:144. https://doi.org/10.1038/s41597-020-0486-7

26. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 2016 Mar;3:1–9. https://doi.org/10.1038/s41597-020-0486-7

27. Liberating Structures. Min Specs [Internet]. [cited 2023 Feb 10]; Available from: https://www.liberatingstructures.com/14-min-specs/.

28. Zimmerman B, Lindberg C, Plsek PE. Edgeware: Insights from complexity science for health care leaders. Irving, TX: VHA Publishing; 1998.

29. Paprica PA, Sutherland E, Smith A, Brudno M, Cartagena RG, Crichlow M, Courtney BK, Loken C, McGrail KM, Ryan A, Schull MJ. Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen Canadian organizations and initiatives. Int. J. Popul. Data Sci. 2020;5:1353. https://doi.org/10.23889/ijpds.v5i1.1353

30. Zhang X. A commentary of Data trusts in MIT Technology Review 2021. Fundam. Res. 2021 Nov;1:834–5. https://doi.org/10.1016/j.fmre.2021.11.016

31. Milne R, Sorbie A, Dixon-Woods M. What can data trusts for health research learn from participatory governance in biobanks? J. Med. Ethics 48:323–8. https://doi.org/10.1136/medethics-2020-107020

32. Rinik C. Data trusts: more data than trust? The perspective of the data subject in the face of a growing problem. Int. Rev. Law Comput. Technol. 2020 Sep;34:342–63. https://doi.org/10.1080/13600869.2019.1594621

33. United Nations. United Nations Declaration on the Rights of Indigenous Peoples [Internet]. 2007

Sep. [cited 2023 Feb 10]; Available from: https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html.

34. First Nations Information Governance Centre. First Nations data sovereignty in Canada. Stat. J. IAOS 2019 Mar;35:47–69. https://doi.org/10.3233/SJI-180478

35. Rainie SC, Kukutai T, Walter M, Figueroa-Rodríguez OL, Walker J, Axelsson P. Indigenous data sovereignty [Internet]. In: Davies T, Walker SB, Rubinstein M, Perini F., editor. The state of open data: Histories and horizons. Cape Town and Ottawa: African Minds and International Development Research Centre; 2019 [cited 2023 Feb 10]. Available from: https://library.oapen.org/bitstream/handle/20.500.12657/24884/The_State_of_Open_Data_9781928331957_web.pdf?sequence=1#page=315.

36. Disaggregated demographic data collection in British Columbia: The grandmother perspective [Internet]. British Columbia's Office of the Human Rights Commissioner; 2020 [cited 2023 Feb 15]. Available from: https://bchumanrights.ca/publications/datacollection/#:...:text=in%20British%20Columbia%3A-,The%20grandmother%20perspective,that%20effectively%20addresses%20systemic%20inequalities.

37. Rowe RK, Carroll SR, Healy C, Rodriguez-Lonebear D, Walker JD. The SEEDS of Indigenous population health data linkage. Int. J. Popul. Data Sci. 2021 Jun;6:1417. https://doi.org/10.23889/ijpds.v6i1.1417

38. Support of US Indigenous Data Sovereignty and Inclusion of Tribes in the Development of Tribal Data Governance Principles [Internet]. The National Congress of American Indians Resolution #KAN-18-011; 2018 [cited 2023 Feb 15]. Available from: https://www.ncai.org/attachments/Resolution_gbuJbEHWpkOgcwCICRtgMJHMsUNofqYvuMSnzLFzOdxBIMlRjij_KAN-18-011%20Final.pdf.

39. Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodriguez-Lonebear D, Rowe R, Sara R. The CARE principles for indigenous data governance. Data Sci. J. 2020;19:1–12. https://doi.org/10.5334/dsj-2020-043

40. Ownership, Control, Access and Possession (OCAP^{TM}): The Path to First Nations Information Governance. [Internet]. Ottawa: The First Nations Information Governance Centre.; 2014 [cited 2023 Feb 15]. Available from: https://fnigc.ca/wp-content/uploads/2020/09/5776c4ee9387f966e6771aa93a04f389_ocap_path_to_fn_information_governance_en_final.pdf.

41. First Nations Information Governance Centre. Barriers and Levers for the Implementation of OCAP^{TM}. Int. Indig. Policy J. [Internet] 2014;5:1–11. Available from: https://ir.lib.uwo.ca/iipj/vol5/iss2/3/

42. Statistics Canada. Disaggregated Data Action Plan: Why it matters to you [Internet]. 2021 Dec. [cited 2023 Feb 15]; Available from: https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2021092-eng.htm

43. Health Data Research Network Canada. Data for Equity? [Internet]. [cited 2023 Feb 15]; Available from: https://www.hdrn.ca/en/data-equity.

44. Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Berkowitz, E., et al. A Toolkit for Centering Racial Equity Throughout Data Integration [Internet]. Actionable Intelligence for Social Policy, University of Pennsylvania; 2020 [cited 2023 Feb 15]. Available from: https://aisp.upenn.edu/centering-equity/.

45. Black Health Equity Working Group. Engagement, Governance, Access, and Protection (EGAP): A Data Governance Framework for Health Data Collected from Black Communities in Ontario [Internet]. Alliance for Healthier Communities; 2021 [cited 2023 Feb 15]. Available from: https://blackhealthequity.ca/wp-content/uploads/2021/03/Report_EGAP_framework.pdf.

46. Information and Communications Standards Development Committee. Review of the Information and Communications Standards: 2020 final recommendations report [Internet]. Toronto: Government of Ontario; 2020 [cited 2023 Feb 15]. Available from: https://www.ontario.ca/page/review-information-and-communications-standards-2020-final-recommendations-report.

47. Treviranus J. The Accessibility Ecosystem Proposal [Internet]. 2022 Oct. [cited 2023 Feb 16]; Available from: https://wecount.inclusivedesign.ca/views/the-accessibility-ecosystem-proposal/.

48. Treviranus J. Accessibility Ecosystem, in The Accessibility Ecosystem Proposal [Internet]. 2002 [cited 2023 Feb 16]; Available from: https://files.ontario.ca/msaa_1/msaa-icsdc-infographic-1-en-2020-11-03.pdf.

49. CAN/DGSI 100-7:2023, Data Governance – Part 7: Operating model for responsible data stewardship [Internet]. Digital Governance Council; 2023 [cited 2023 Aug 3]. Available from: https://dgc-cgn.org/standards/find-a-standard/standards-in-data-governance/responsible-data-stewardship/.

50. Hawn Nelson A, Kemp D, Jenkins D, Benitez JR, Berkowitz E, Burnett TC, Smith K, Zanti S, Culhane D. Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration. [Internet]. University of Pennsylvania: Actionable Intelligence for Social Policy; 2022 [cited 2023 Feb 16]. Available from: https://aisp.upenn.edu/resource-article/finding-a-way-forward-how-to-create-a-strong-legal-framework-for-data-integration/.

51. ISO 24143:2022(en) Information and documentation — Information Governance — Concept and principles

[Internet]. The International Organization for Standardization; [cited 2023 Feb 16]. Available from: https://www.iso.org/obp/ui/#iso:std:iso:24143:ed-1:v1:en.

52. Lucero JE, Emerson AD, Beurle D, Roubideaux Y. The holding space: a guide for partners in tribal research. Prog. Community Health Partnersh. Res. Educ. Action 2020;14:101–7. https://doi.org/10.1353/cpr.2020.0012

53. Evans-Lomayesva G, Lee J. Advancing American Indian & Alaska Native Data Equity. [Internet]. Georgetown Center on Poverty and Inequality; 2022 [cited 2023 Feb 16]. Available from: https://www.georgetownpoverty.org/issues/advancing-american-indian-alaska-native-data-equity/.

54. Panel on Research Ethics. TCPS 2: CORE-2022 (Course on Research Ethics) [Internet]. [cited 2023 Feb 16]; Available from: https://tcps2core.ca/welcome

55. Data Management and Security for Student Researchers: An Overview [Internet]. 2021 Nov. [cited 2023 Feb 15]; Available from: https://about.citiprogram.org/course/data-management-and-security-for-student-researchers-an-overview/?h=data.

56. Hawn Nelson A, Culhane D. Social-Behavioral-Educational (SBE) Comprehensive [Internet]. 2020 Feb. [cited 2023 Feb 15]; Available from: https://about.citiprogram.org/course/human-subjects-research-2/.

57. Wilbanks J, Friend SH. First, design for data sharing. Nat. Biotechnol. 2016 Apr;34:377–9. https://doi.org/10.1038/nbt.3516

58. Synapse. Certified User Quiz [Internet]. [cited 2023 Feb 15]; Available from: https://www.synapse.org/#!Quiz:Certification (registration required to view content)

59. Material Transfer Agreement [Internet]. 2021 Jun. [cited 2023 Feb 15]; Available from: https://www.ukbiobank.ac.uk/media/p3zffurf/biobank-mta.pdf.

60. IAP2 Spectrum of Public Participation. [Internet]. 2018 12 [cited 2023 Feb 15]; Available from: https://cdn.ymaws.com/www.iap2.org/resource/resmgr/pillars/Spectrum_8.5x11_Print.pdf.

61. How the Hartford Data Collaborative addresses fifteen essential requirements for responsible data stewardship [Internet]. 2023 18 [cited 2023 Feb 15]; Available from: https://static1.squarespace.com/static/5d8b7b3eabff3c4f1954d802/t/63c9b2638614cc5609a3a0d3/1674163135114/hdc-minspecs.

62. ICES. Data Repository Requirements [Internet]. https://www.ices.on.ca/accountability-and-reporting/data-repository-requirements.

63. NB-IRDT. Data privacy and security [Internet]. [cited 2023 Mar 6]; Available from: https://www.unb.ca/nbirdt/data/privacy/index.html.

64. CIO Strategy Council. CIO Strategy Council Publishes National Standard For Responsible Data Stewardship [Internet]. 2022 Jul. [cited 2023 15]; Available from: https://dgc-cgn.org/cio-strategy-council-publishes-national-standard-for-responsible-data-stewardship/.

65. Gibbs L, Hawn Nelsom A, Dalton E, Cantor J, Shipp S, Jenkins D. IDS Governance: Setting Up for Ethical and Effective Use. [Internet]. Philadelphia: University of Pennsylvania: Actionable Intelligence for Social Policy; 2017 [cited 2023 Feb 15]. Available from: https://aisp.upenn.edu/wpcontent/uploads/2016/07/Governance.pdf.

66. ICES. Privacy at ICES [Internet]. [cited 2023 Feb 15]; Available from: https://www.ices.on.ca/Data-and-Privacy/Privacy-at-ICES.

67. Contracts For Data Collaboration (C4DC). Library: List of All Example Agreements [Internet]. [cited 2023 Feb 15]; Available from: https://contractsfordatacollaboration.org/library/.

68. CIHI's Strategic Plan: 2022 to 2027 [Internet]. Canadian Institute for Health Information (CIHI); [cited 2023 Feb 15]. Available from: https://www.cihi.ca/sites/default/files/document/cihi-strategic-plan-2022-2027-en.pdf.

69. ICES. Mission, Vision & Values [Internet]. [cited 2023 Feb 15]; Available from: https://www.ices.on.ca/About-ICES/Mission-vision-and-values.

70. GEMINI. About Us: Helping physicians, health care teams and hospitals improve patient care [Internet]. [cited 2023 Feb 15]; Available from: https://www.geminimedicine.ca/about.

71. MCHP. About Manitoba Centre for Health Policy (MCHP) [Internet]. [cited 2023 Feb 15]; Available from: https://umanitoba.ca/manitoba-centre-for-health-policy/.

72. Primary Care Ontario Practice-Based Learning and Research Network (POPLAR),. Mission & Vision [Internet]. [cited 2023 Feb 15]; Available from: https://www.poplarnetwork.ca/mission-vision.

73. Centre for Addiction and Mental Health (CAMH) BrainHealth Databank. Our Team [Internet]. [cited 2023 Feb 15]; Available from: https://www.camh.ca/en/science-and-research/discovery-fund/brainhealth-databank/our-team.

74. Canadian Institute for Health Information (CIHI). Accountability [Internet]. [cited 2023 Feb 15]; Available from: https://www.cihi.ca/en/about-cihi/governance-and-accountability/accountability.

75. CRDCN. Organisational chart of the CRDCN [Internet]. 2009 Oct. [cited 2023 Feb 15]; Available from: https://crdcn.ca/app/uploads/2021/10/1-4-CRDCN-Organisation-chart-October-2009.pdf.

76. Diabetes Action Canada. Governance Portal [Internet]. [cited 2023 Feb 22]; Available from: https://diabetesaction.ca/governance-portal/.

77. ICES. Board of Directors [Internet]. [cited 2023 Feb 15]; Available from: https://diabetesaction.ca/governance-portal/.

78. ICES. Key Contacts at ICES [Internet]. [cited 2023 Feb 15]; Available from: https://www.ices.on.ca/About-ICES/Key-Contacts.

79. Ontario Health Data Platform (OHDP). Governance Structure [Internet]. [cited 2023 Feb 15]; Available from: https://ohdp.ca/governance-structure/.

80. Board of Directors Governance Handbook [Internet]. Ottawa, ON: Canadian Institute for Health Information (CIHI); 2021 [cited 2023 Feb 15]. Available from: https://www.cihi.ca/sites/default/files/document/governance_handbook_en.pdf.

81. MCHP. The Manitoba Population Research Data Repository [Internet]. [cited 2023 Feb 15]; Available from: https://umanitoba.ca/manitoba-centre-for-health-policy/data-repository.

82. MCHP. Department of Community Health Sciences [Internet]. [cited 2023 Feb 15]; Available from: https://umanitoba.ca/manitoba-centre-for-health-policy/data-repository.

83. Diabetes Action Canada. National Diabetes Repository Research Governance Framework and Guidelines [Internet]. 2019 Oct. [cited 2023 Feb 16]; Available from: https://repository.diabetesaction.ca/wp-content/uploads/2019/10/Research-Governance-Operational-Framework-101119.pdf.

84. Diabetes Action Canada. Documents [Internet]. [cited 2023 Feb 16]; Available from: https://repository.diabetesaction.ca/documents/.

85. Health Data Research Network (HDRN) Canada. Working With Indigenous Data [Internet]. [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/en/dash/working-with-indigenous-data.

86. ICES. Indigenous Portfolio [Internet]. [cited 2023 Feb 16]; Available from: https://www.ices.on.ca/About-ICES/Collaborations-and-Partnerships/Indigenous-Portfolio.

87. Health Data Research Network (HDRN) Canada. Health Data Research Network Canada Strategic Plan 2021-2026 [Internet]. 2021 [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/en/about/strategic-plan.

88. Health Data Research Network (HDRN) Canada. HDRN Canada Board of Directors Terms of Reference [Internet]. 2020 May [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/sites/default/files/2022-02/HDRN%20Canada%20Board%20of%20Directors%20Terms%20of%20Reference.pdf.

89. ICES. ICES Strategic Plan 2020/21-2022/23 [Internet]. 2020 [cited 2023 Feb 16]; Available from: https://www.ices.on.ca/.../media/Files/Corporate-Reports/2020/Strategic-Plan-2020-2023.ashx?la=en-CA.

90. Actionable Intelligence for Social Policy (AISP). Quality Framework for Integrated Data Systems [Internet]. 2021 [cited 2023 Feb 16]; Available from: https://aisp.upenn.edu/quality-framework-for-integrated-data-systems/#elementor-action%3Aaction%3Dpopup%3Aopen%26settings%3DeyJpZCI6IjQ3MjciLCJ0b2dnbGUiOmZhbHNlfQ%3D%3D.

91. Canadian institute for Health Information (CIHI). Privacy Policy on the Collection, Use, Disclosure and Retention of Health Workforce Personal Information and De-identified Data, 2011 [Internet]. 2011 Jun. [cited 2023 Feb 16]; Available from: https://www.cihi.ca/sites/default/files/document/hw_privacy_policy_2011_en.pdf.

92. Canadian Institute for Health Information (CIHI). Privacy Policy on the Collection, Use, Disclosure and Retention of Personal Health Information and De-Identified Data, 2010 - Updated November 2022 [Internet]. Ottawa, ON: CIHI; 2022 [cited 2023 Feb 16]. Available from: https://www.cihi.ca/sites/default/files/document/privacy-policy-en.pdf.

93. Canadian Institute for Health Information (CIHI). Privacy Impact Assessment Policy [Internet]. Ottawa, ON: CIHI; 2020 [cited 2023 Feb 16]. Available from: https://www.cihi.ca/sites/default/files/rot/2020-privacy-impact-assessment-policy-en-web.pdf.

94. Health Data Research (HDR) UK. New principles published to improve public confidence in access and use of data for health research through Trusted Research Environments [Internet]. 2021 Dec. [cited 2023 Feb 16]; Available from: https://www.hdruk.ac.uk/news/new-principles-published-to-improve-public-confidence-in-access-and-use-of-data-for-health-research-through-trusted-research-environments/.

95. Canadian Institute for Health Information (CIHI). Privacy and security [Internet]. [cited 2023 Feb 16]; Available from: https://www.cihi.ca/en/about-cihi/privacy-and-security.

96. Smith M, Ross KA, ICES Privacy & Legal Office. Report to the Information and Privacy Commissioner of Ontario: Three-Year Review as a Prescribed Entity under PHIPA [Internet]. Toronto: ICES; 2020 [cited 2023 Feb 16]. Available from: file:///C:/Users/skesselring/Downloads/ICES-Privacy-Report.pdf

97. Canadian Institute for Health Information (CIHI). Privacy and Security Risk Management Framework [Internet]. Ottawa, ON: CIHI; 2020 [cited 2023 Feb 16]. Available from: https://www.cihi.ca/sites/default/files/document/privacy-security-risk-management-framework-en.pdf.

98. Health Data Research Network (HDRN) Canada. Data Access Support Hub (DASH) [Internet]. [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/en/dash.

99. Diabetes Action Canada. Data Dictionary beginning alphabetically with "Allergy Intolerance" [Internet]. [cited 2023 Feb 16]; Available from: https://repository.diabetesaction.ca/wp-content/uploads/2019/11/vimo-v2.htm#.

100. Diabetes Action Canada. Table 1: Rural-urban distribution [Internet]. [cited 2023 Feb 16]; Available from: https://repository.diabetesaction.ca/wp-content/uploads/2020/11/ndr_ses.html.

101. ICES. ICES Data Dictionary [Internet]. [cited 2023 Feb 16]; Available from: https://www.ices.on.ca/Data-and-Privacy/ICES-data/Data-dictionary.

102. Health Data Research Innovation Gateway. Datasets [Internet]. [cited 2023 Feb 16]; Available from: https://web.www.healthdatagateway.org/search?search=&datasetSort=latest&tab=Datasets.

103. Maelstrom. Individual Studies [Internet]. [cited 2023 Feb 16]; Available from: https://www.maelstrom-research.org/individual-studies?query=study(limit(0,50),sort(name),in(Mica_study.className,Study))

104. Canadian Institute for Health Information (CIHI). Privacy and Security Training Policy [Internet]. Ottawa, ON: CIHI; 2022 [cited 2023 Feb 16]. Available from: https://www.cihi.ca/sites/default/files/document/privacy-security-training-policy-en.pdf.

105. GEMINI. Access to GEMINI Data [Internet]. [cited 2023 Feb 16]; Available from: https://www.geminimedicine.ca/access-data.

106. GO FAIR. GO FAIR Workshop Series [Internet]. [cited 2023 Feb 16]; Available from: https://www.go-fair.org/resources/go-fair-workshop-series/.

107. GEMINI. GEMINI End User Agreement [Internet]. 2021 Feb. [cited 2023 Feb 16]; Available from: https://1e448661-e373-4273-8b91-ef3034f3eea8.filesusr.com/ugd/8a116f_f1da2254cea24b5199ef9e3e085ff2f4.pdf.

108. ICES. Conflict of Interest Policy [Internet]. Toronto, ON: ICES; 1994 [cited 2023 Feb 16]. Available from: file:///C:/Users/skesselring/Downloads/Conflict-of-Interest-Policy.pdf.

109. Synapse. Synapse Commons Data Use Procedure [Internet]. Seattle: SageBionetworks; 2022 [cited 2023 Feb 16]. Available from: https://s3.amazonaws.com/static.synapse.org/governance/SynapseCommonsDataUseProcedure.pdf (registration required to view content)

110. ICES. Research Impact Stories [Internet]. [cited 2023 Feb 16]; Available from: https://www.ices.on.ca/Newsroom/Impact-Stories.

111. ICES. Applied Health Research Questions (AHRQ) [Internet]. [cited 2023 Fed 16]; Available from: https://www.ices.on.ca/DAS/AHRQ.

112. Primary Care Ontario Practice-Based Learning and Research Network (POPLAR). Data Security & Privacy [Internet]. [cited 2023 Feb 16]; Available from: https://www.poplarnetwork.ca/data-security.

113. Primary Care Ontario Practice-Based Learning and Research Network. Contribute EMR Data [Internet]. [cited 2023 Feb 16]; Available from: https://www.poplarnetwork.ca/contributeemrdata.

114. Willison DJ, Trowbridge J, Greiver M, Keshavjee K, Mumford D, Sullivan F. Participatory governance over research in an academic research network: the case of Diabetes Action Canada. BMJ Open 2019 Apr;9:e026828. https://doi.org/10.1136/bmjopen-2018-026828

115. Health Data Research Network (HDRN) Canada. Public Advisory Council [Internet]. [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/index.php/en/public/pac.

116. Health Data Research Network (HDRN) Canada. For the Public [Internet]. [cited 2023 Feb 16]; Available from: https://www.hdrn.ca/en/public.

117. Burt J, Cumyn A, Dault R, Paprica PA, Blouin C, Carter P, et al. Social licence for uses of health data: A report on public perspectives [Internet]. Vancouver, BC: Health Data Research Network (HDRN) Canada; 2022 [cited 2023 Feb 16]. Available from: https://www.hdrn.ca/en/public/reports.

118. ICES. Public Advisory Council [Internet]. [cited 2023 Feb 22]; Available from: https://www.ices.on.ca/About-ICES/Public-Engagement/Public-Advisory-Council.

119. Paul J, Davidson R, Johnstone C, Loong M, Matecsa J, Guttmann A, Schull MJ. Public engagement can change your research, but how can it change your research institution? ICES Case Study. Int. J. Popul. Data Sci. 2020 Sep;5:1364. https://doi.org/10.23889/ijpds.v5i3.1364

120. The Data Assembly. Public Deliberation On The Re-Use Of Data [Internet]. [cited 2023 Feb 22]; Available from: https://thedataassembly.org/.

121. ENGAGE. About [Internet]. [cited 2023 Feb 22]; Available from: https://engageplus.org/en.

122. National Institute for Health and Care Excellence (NICE). Get Involved [Internet]. [cited 2023 Feb 22]; Available from: https://www.nice.org.uk/get-involved.

123. Hopkins H, Kinsella S, Evans G, Reid S. Putting Good into Practice: A public dialogue on making public benefit assessments when using health and care data [Internet]. London: Hopkins Van Mil; 2021 [cited 2023 Feb 22]. Available from: https://www.gov.uk/government/publications/putting-good-into-practice-a-public-dialogue-on-making-public-benefit-assessments-when-using-health-and-care-data.

124. Health Data Research Network (HDRN) Canada. Virtual Library [Internet]. [cited 2023 Feb 22]; Available from: https://www.hdrn.ca/index.php/en/data-equity/virtual-library.

125. The GovLab. The 100 Questions Initiative [Internet]. [cited 2023 Feb 22]; Available from: https://thegovlab.org/project/project-the-100-questions-initiative.

126. Paprica A. Comments – Alison Paprica: Comments in Association with TCSP 2 Consultation [Internet]. [cited 2023 Feb 22]; Available from: https://ethics.gc.ca/eng/consultation-alison_paprica.html.

127. The GovLab. Data Stewards Academy [Internet]. [cited 2023 Feb 22]; Available from: https://course.opendatapolicylab.org/.

Abbreviations

AISP:	Actionable Intelligence for Social Policy
CAMH:	Centre for Addiction and Mental Health
CanPath:	Canadian Partnership for Tomorrow's Health
CARE Principles:	Collective Benefit, Authority to Control, Responsibility, and Ethics
CHUS:	Centre hospitalier universitaire de Sherbrooke
CIHI:	Canadian Institute for Health Information
CITI:	Collaborative Institutional Training Initiative
CRDCN:	Canadian Research Data Centre Network
DGSI:	Digital Governance Standard Institute (of Canada)
DULs:	Data Use Licences
EGAP Framework:	Engagement Governance, Access and Protection Framework
EMOU:	Enterprise Memorandum of Understanding
FAIR Principles:	Findability, Accessibility, Interoperability, and Reusability
HDC:	Hartford Data Collaborative
HDRN Canada:	Health Data Research Network Canada
HITAP-NHSO:	Health Intervention and Technology Assessment Program National Health Security Office, Thailand
ISO:	International Organization for Standards
IAP2:	International Association for Public Participation
ICES:	Institute for Clinical Evaluative Sciences
LOI:	Letter of Intent
MCHP:	Manitoba Centre for Health Policy
Min specs:	minimum specification requirements
NB-IRDT:	New Brunswick Institute for Research Data and Training
NLCHI:	Newfoundland and Labrador Centre for Health Information
OCAP:	Ownership, Control, Access, and Possession
PI:	Principal Investigator
PopData:	Population Data British Columbia (BC)
POPLAR:	Primary Care Ontario Practice-based Learning and Research Network
TCPS2:	The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans
TRUST principles:	Transparency, Responsibility, User focus, Sustainability and Technology
USA:	United States of America

Appendix A: Organisations and initiatives that tested the min specs

Data trust, Data repository, or Data collaboration	Description
Actionable Intelligence for Social Policy (AISP)	Actionable Intelligence for Social Policy (AISP) helps state and local governments collaborate and responsibly use data to improve lives. AISP does not hold data. AISP is a network and peer learning community of 36 distinct state and local data collaborations, each of which holds cross-sector data on health, education, and social service provision. Collectively, the AISP network covers over 50% of the US population. Website: https://www.aisp.upenn.edu/
The Centre for Addiction and Mental Health (CAMH)-wide BrainHealth Databank	The Centre for Addiction and Mental Health (CAMH)-wide BrainHealth Databank initiative accelerates research and improves care by collecting and studying the full spectrum of data that individuals choose to share to advance mental health. The BrainHealth Databank holds data from diverse multi-modal research projects and integrated clinical programs in Ontario, Canada, and includes records for over 25,000 participants. Website: https://www.camh.ca/en/science-and-research/discovery-fund/brainhealth-databank
Canadian Partnership for Tomorrow's Health (CanPath)	The Canadian Partnership for Tomorrow's Health (CanPath) is a large-scale, prospective cohort platform used by the research community to study the biology, behaviours, and environments of Canadians to advance knowledge of the causes and control of chronic diseases and cancer. CanPath holds and regularly updates data provided with consent by more than 300,000 Canadians who are actively involved and followed in every province. CanPath data include surveys, health measures, environmental exposures, and linkages to health system records. Many participants also provided physical measures and biological samples from which genomic data can be derived. Website: https://canpath.ca/
Canadian Institute for Health Information (CIHI)	CIHI is an independent, not-for-profit organisation funded by the Canadian federal government and Canadian provinces and territories to deliver comparable and actionable information to accelerate improvements in health care, health system performance and population health across the continuum of care. CIHI is a secondary data collector of health information and holder of health data. CIHI data holdings include population-level hospital data for almost the entire population of Canada as well as data from other health care facilities, long-term care homes, regional health authorities, medical practitioners and governments. Website: https://www.cihi.ca/en
Canadian Research Data Centre Network (CRDCN)	The CRDCN is a partnership between Statistics Canada and a consortium of more than 30 Canadian universities. Through the CRDCN, approved researchers in university, government, and other sectors are able to access linked but anonymised social, economic and health datasets prepared by Statistics Canada that, in some cases, include entire populations of provinces or territories. Website: https://crdcn.ca/
Sherbrooke University Hospital Center (CHUS) Biobank	The CHUS Biobank is located in Quebec, Canada, and provides data for approved research projects from researchers. Specimens include blood, tissue, fluid, stool, microbiome and living derivatives. Data include demographic data, health status (weight, blood test, illnesses, drugs, medical images, etc.), lifestyle habits (diet, alcohol consumption, physical activity, etc.), medical and family history, clinical data as well as data from genome-wide sequencing. Website: https://biobanking.org/biobanks/view/123

Continued

Appendix A: Continued

Data trust, Data repository, or Data collaboration	Description
Digital Cardiac Health Platform (DCHP)	The Digital Cardiac Health Platform (DCHP) is a secure and high-performance data integration platform for secondary use of clinical data located in Ontario, Canada. Website: https://www.uhn.ca/PMCC/Documents/Strategic-Expansion-of-Digital-Cardiac-Health-Platform-into-UHN-Digital-Health-Platform.pdf
GEMINI	The GEMINI Study collects routinely-generated administrative and clinical data for research and quality improvement. GEMINI holds data from over 30 participating hospitals in Ontario, Canada, and includes records for over 2 million hospitalisations. Website: https://www.geminimedicine.ca/
Hartford Data Collaborative (HDC)	The HDC is part of the Connecticut Data Collaborative. It includes a network of non-profit organisations, government agencies, and philanthropic partners that facilitates data sharing and data integration among its partners in the Hartford, Connecticut area. HDC does not hold data, it links data for others. HDC works with data on demographics, health, education and social service provision. Website: https://www.ctdata.org/about-hdc
Health Data Research Network (HDRN) Canada	HDRN Canada brings together provincial, territorial and federal organisations which hold and manage data. HDRN Canada does not hold data. HDRN Canada is a network of provincial, territorial and federal data centres which hold data covering the entire Canadian population. Data covers topics including health, education, demographics, and immigration. Website: https://www.hdrn.ca/
Health Intervention and Technology Assessment Program - National Health Security Office (HITAP-NHSO), Thailand Ministry of Public Health	The Health Intervention and Technology Assessment Program (HITAP) and National Health Security Office (NHSO) worked together on a research data initiative in Thailand. This initiative was created to explore the impact of COVID-19 on the healthcare system such as health service utilisation. Data includes demographics, inpatient and outpatient information. Over 1 billion records are held in the NHSO database which represent health service utilisation record from a universal coverage scheme which provides health care to approximately 80% of Thai citizens. Website: https://www.hitap.net/research/179967
ICES (formerly the Institute for Clinical Evaluative Sciences, now using the name ICES in public communications)	The ICES Data Repository includes over 100 health and population-related data assets in which direct personal identifiers (such as name and health card number, if applicable) have been removed and replaced with a confidential code. Most holdings are administrative data at population-level, including decades of longitudinal health records for 21 million people who are, or have previously been, eligible for publicly funded health care services in Ontario, Canada. The ICES Data Repository also includes some data for publicly funded services outside of health, population survey, investigator-led research project data sets (e.g., clinical trial data sets), and registries that have been linked to ICES' core administrative data holdings. Website: https://www.ices.on.ca/

Continued

Appendix A: Continued

Data trust, Data repository, or Data collaboration	Description
Manitoba Centre for Health Policy (MCHP)	The Manitoba Centre for Health Policy (MCHP) is a population health data centre located at the University of Manitoba in Winnipeg, Canada. MCHP conducts research on health and the social determinants of health. MCHP holds population-level data for approximately 1.37 million people, as of 2022, derived from Manitoba Government departments, provincial laboratories, clinical programs, community and social outreach organisations, and Indigenous governance bodies. Website: https://umanitoba.ca/manitoba-centre-for-health-policy/
National Diabetes Repository	The National Diabetes Repository is a scalable secure analytics platform and multi-jurisdictional data repository. Through enhanced informatics, the repository applies privacy-aware methods to safely contain data and make it available to approved investigators for analytics. The data are de-identified and include primary care electronic medical records (e.g., Encounters, Prescriptions, Vital Stats) from the Canadian provinces of Ontario, Alberta, Manitoba, Newfoundland and Labrador, and Quebec. The data represent over 150,000 individuals from over 1,500 physicians. Website: https://repository.diabetesaction.ca/
New Brunswick Institute for Research Data and Training (NB-IRDT)	The NB-IRDT at the University of New Brunswick is a provincial data custodian that hosts and provides researcher access to linkable and prepared person-level data on health, education, labour market training, social assistance, immigration, and other related topics. Most of the datasets at NB-IRDT are population-level, providing information on almost all of New Brunswick's approximately 800,000 residents. Website: https://www.unb.ca/nbirdt/
Newfoundland and Labrador Centre for Health Information (NLCHI)	NLCHI is a crown corporation responsible for managing provincial health data and information assets and providing data access and health analytics services, whose staff also have expertise in data security and privacy. NLCHI holds population-level data for the province's roughly 520,000 residents. Website: https://www.nlchi.nl.ca/
Ontario Brain Institute – Brain CODE	The Ontario Brain Institute's Brain-CODE is a large-scale informatics platform that manages the acquisition and storage of multidimensional data collected from participants with a variety of brain disorders. Brain-CODE currently hosts data from over 20,000 participants with a wide array of data types, including clinical, neuroimaging and molecular. Website: https://www.braincode.ca/content/about-brain-code
Ontario Health Data Platform (OHDP)	The Ontario Health Data Platform (OHDP) is a highly secure high performance computing environment that was established by the government of Ontario support scientific investigation of COVID-19 by enabling private and secure record level linkage to personal health information. The OHDP is comprised of datasets containing personal health information collected from ICES and/or Ontario Health. The OHDP holds population-level data for the province of Ontario. Website: https://ohdp.ca
PHEMI Health DataLab	The PHEMI Health DataLab is a cloud-based system for privacy, security & governance. PHEMI has one central system to manage data stored in many locations (cloud or hybrid cloud), process data in many locations (cloud or hybrid cloud) while controlling access and governing data all in one place. Website: https://www.phemi.com/data-privacy-management/

Continued

Appendix A: Continued

Data trust, Data repository, or Data collaboration	Description
Population Data BC (PopData)	PopData is a data and education resource facilitating interdisciplinary research on the determinants of human health, well-being and development in British Columbia (BC), Canada. PopData supports access to population-level data for BC's approximately 5 million residents and from over 30 data sets from both federal and provincial sources. Website: https://www.popdata.bc.ca/
Primary care Ontario Practice-based Learning and Research Network (POPLAR)	POPLAR is an initiative of Ontario's six University Departments/Sections of Family Medicine and the Alliance for Healthier Communities. POPLAR securely collects and de-identifies electronic medical record (EMR) data to support practices in delivering optimal care across Ontario, and strengthen practice-based clinical research and quality improvement processes. Currently, over 1,000 family physicians are contributing EMR data for over 1.5 million patients to the POPLAR database. Website: https://www.poplarnetwork.ca/?
Rhode Island Ecosystem	The Rhode Island Ecosystem is an integrated data system that links data, at the person and family level, across various state agency and non-profit datasets to drive holistic improvements in the wellbeing of Rhode Island residents. The Rhode Island Ecosystem holds population-level data for the state's roughly 1 million residents. Website: https://eohhs.ri.gov/initiatives/data-ecosystem
ThinkData Works - External Data Catalog	ThinkData Works' Data Catalog is a scalable, cloud-agnostic end-to-end data management platform that helps organisations securely find, govern, and deliver data. With configurable data ingestion, active metadata management, role-based distribution and streaming integration to data science toolkits, the ThinkData Catalog ensures data preparedness for governance reporting and advanced analytics. With multi-cloud warehouse support and data virtualisation capabilities, the ThinkData Catalog lets clients access and collaborate on shared data while adhering to data residency requirements. The Catalog supports regional deployment on isolated infrastructure, letting users define where and what data is hosted on the platform. Website: https://www.thinkdataworks.com/platform/data-catalog

Appendix B: Main changes to min specs and reasons for those changes

Refined 2023 min spec (changes relative to the 2020 min spec are identified with italic font)	**Min spec published in 2020**	**Main reason(s) for change(s)**
Legal 1) The data trust, *data repository, or data collaboration* must fulfill all legal requirements including, *as required,* authority*(ies)* to collect, *retain, use, disclose, and/or destroy* data	The data trust must fulfill all legal requirements, including the authority to collect, share and hold data	"Authority" changed to "authorities" because multiple authorities may apply. Expanded the list of activities that there must be authority(ies) for. Replaced the term "share" with "disclose" to be inclusive of activities to make data available/accessible and/or release it to another organisation or person.
Governance 2a) The data trust, *data repository, or data collaboration* must have a stated purpose *that specifically addresses why its activities are necessary or beneficial*	The data trust must have a stated purpose	Original min spec was vague and could be fulfilled by any organisation/initiative regardless of its purpose. Revised min spec requires that a data-related purpose is either necessary (e.g., required by law) or producing some public benefit.
Governance 2b) The data trust, *data repository, or data collaboration* must have an accountable governance body *that is answerable for its decisions*	The data trust must have an accountable governing body	Added "answerable" to ensure that the governance body(ies) have mechanisms in place to respond to questions and concerns.
Governance 2c) The data trust, *data repository, or data collaboration* must be transparent *about its purpose, governance body membership, data holdings, policies regarding who has access to what data for what purposes, and other information that is requested*	The data trust must be transparent in its activities	Original min spec was too vague; responses often did not provide the information that data partners, stakeholders and members of the public would typically seek.
Governance 2d) The data trust, data repository, or data collaboration must acknowledge and respect Indigenous Data Sovereignty	Not applicable	New requirement added consistent with the right to self-determination under the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP).
Governance 2e) Governance must be adaptive *and responsive to risks, opportunities, and the concerns of stakeholders*	Governance must be adaptive	Added "adaptive" to ensure that, in addition to establishing proactive measures, the governance body(ies) has/have mechanisms in place to change. Added detail on what organisations/initiatives might need to adapt and respond to, i.e., response to risks, opportunities, questions, and concerns of stakeholders.
Management 3a) There must be well-defined policies, processes, and *procedures covering the entire data_lifecycle*	There must be well-defined policies and processes for the collection, storage, use and disclosure of data	Added "procedures" and "covering the data lifecycle" to make the min spec more technically complete.
Management 3b) There must be policies, processes, and/or *procedures* for *cybersecurity and* data protection safeguards which are reviewed and updated regularly	Policies and processes must include data protection safeguards which are reviewed and updated regularly	Added "procedures" to make the min spec more technically complete. Given public concerns about the topic, specifically included the term "cybersecurity" (though it could be considered part of data protection broadly).
Management 3c) There must be *policies*, processes *and*/or *procedures* to identify, assess, and manage risks *on an ongoing basis*	There must be an ongoing process to identify, assess and manage risks	Added policies and procedures to make the min spec more technically complete and to make it clear that risk management is an ongoing (vs. one-time) requirement.

Continued

Appendix C: Continued

Refined 2023 min spec (changes relative to the 2020 min spec are identified with italic font)	**Min spec published in 2020**	**Main reason(s) for change(s)**
Management 3d) There must be policies, processes and/or procedures to create and maintain metadata and data documentation which provides sufficient information for potential users to find, understand, use, and reuse data holdings	Not applicable	New requirement based on the practical observation that you cannot have a functional data trust, data repository, or data collaboration without metadata and other documentation about the data that it collects, retains, uses, discloses, and/or destroys.
Data Users 4a) Data users must complete *privacy and security* training before they access data	All data users must complete training before they access data	Specified that privacy and security training is mandatory. For consistency with other min specs language, removed the word "all".
Data Users 4b) Data users must *acknowledge* that there *may be* consequences for non-compliance	All data users must agree to a data user agreement that acknowledges that data use will be monitored and includes consequences for non-compliance	Changed the min spec to reflect the practice that many organisations have data users acknowledge terms vs. sign agreements. For consistency with other min specs language, removed the word "all". Changed text to read "may be" consequences vs. implying that there will always be consequences because depending on the severity and intentionality of the non-compliance, there may not be consequences. Removed text about the data user agreeing to monitoring/auditing policies/practices because it is understood that non-compliance would be identified by monitoring/ auditing and, in practice, monitoring/auditing policies and procedures are often described in other documents vs. data user agreements/terms.
Stakeholder & Public Engagement *5a) There must be ongoing engagement with stakeholders.*	There must be early and ongoing engagement with stakeholders including members of the public	Separated stakeholder and public engagement requirements to clarify that both are essential and because they are, by their nature, different.
Stakeholder & Public Engagement *5b) Stakeholder engagement must include ongoing engagement with members of the public*	There must be early and ongoing engagement with stakeholders including members of the public	Separated stakeholder and public engagement requirements to clarify that both are essential and because they are, by their nature, different.
Stakeholder & Public Engagement 5c) Where there is a reasonable expectation that specific subpopulations or groups would have a particular interest in, *and*/or would be affected by, an activity *or decision*, there must be direct engagement tailored for that subpopulation/group	Where there is a reasonable expectation that specific subpopulations or groups would have a particular interest in, or would be affected by, an activity of the data trust, there must be direct engagement tailored for that subpopulation/group	Added the requirement to engage and involve affected groups at key decision points.

Note: the revised min specs use the language "data trust, data repository, or data collaboration" instead of relying on an uncommon interpretation of the term "data trust."