

Rare disease data stewardship in Canada

Alexander Bernier^{a*}

^aCentre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, QC H3A 0G1, Canada

*alexander.bernier@mail.mcgill.ca

Abstract

The Canadian Genomics Partnership for Rare Diseases, spearheaded by Genome Canada, will integrate genome-wide sequencing to rare disease clinical care in Canada. Centralized and tiered models of data stewardship are proposed to ensure that the data generated can be shared for secondary clinical, research, and quality assurance purposes in compliance with ethics and law. The principal ethico-legal obligations of clinicians, researchers, and institutions are synthesized. Governance infrastructures such as registered access platforms, data access compliance offices, and Beacon systems are proposed as potential organizational and technical foundations of responsible rare disease data sharing. The appropriate delegation of responsibilities, the transparent communication of rights and duties, and the integration of data privacy safeguards into infrastructure design are proposed as the cornerstones of rare disease data stewardship.



ETHICO-LEGAL APPROACHES TO PAN-CANADIAN RARE DISEASE DATA STEWARDSHIP

Key words: data sharing, health ethics, research ethics, genetic privacy, government regulation, information dissemination, multi-institutional systems, organizational policy, rare diseases

Introduction

The Canadian Genomics Partnership for Rare Diseases (CGP-4RD), also known as “All for One,” is a novel Genome Canada initiative that will foster data sharing in Canada across the clinical, research, and health administration branches of Canada’s health care system for improved rare and idiopathic

OPEN ACCESS

Citation: Bernier A. 2020. Rare disease data stewardship in Canada. FACETS 5: 836–863. doi:[10.1139/facets-2020-0050](https://doi.org/10.1139/facets-2020-0050)

Handling Editor: Iain E.P. Taylor

Received: June 20, 2020

Accepted: August 25, 2020

Published: November 5, 2020

Copyright: © 2020 Bernier. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Published by: Canadian Science Publishing

disease care. At the outset, the partnership intends to provide Canadians with rare or idiopathic diseases access to genome-wide sequencing as part of routine clinical care. Enabling the reuse of the clinical data generated for research and health care system improvement purposes, both within Canada and internationally, is a central objective of the One for All initiative ([Genome Canada 2019](#)). This policy analysis seeks to elucidate the principal ethical and legal obligations of health care institutions in sharing health data and genomic data, arising from Canadian law and research ethics guidance. The analysis proposes best practices for compliance with these ethical and legal obligations. Ensuring compliance can initially appear challenging as these obligations are not harmonized across Canada's federal, provincial, and sectoral boundaries ([Bernier and Knoppers 2020](#)).

Genome Canada further intends the creation of a “national rare disease cohort” composed of 30 000 sequenced samples provided by rare disease patients and their biological relatives, a “national platform” to share genomic rare disease data directly and guide independent actors in sharing rare disease data, and a “clinical implementation” branch to guide participating sites in obtaining necessary ethical, legal, and regulatory permissions to participate ([Genome Canada 2018](#)). To support this initiative, data stewardship approaches are proposed herein that incorporate legal, organisational, and technological aspects to ensure that rare disease data reuse and sharing complies with Canadian ethics and law. The proposals hope to ensure that Canadian rare disease data can be used interoperably across countries, provinces, and sectors regardless of its initial provenance.

Though this analysis is specific to Canada, its conclusions can be generalized to other jurisdictions that are subject to multifarious and potentially conflicting data privacy laws, such as the Member States of the European Union—and increasingly, the United States and the nations of the Pacific Rim ([Daley et al. 2015](#); [Kim et al. 2018](#); [Hintze 2019](#); [Peloquin et al. 2020](#)). Considerable technical barriers to rare-disease data interoperability have been highlighted in scientific literature, including difficulties in ensuring compatible ontologies are used to describe phenotypic information and disease-specific information ([Thompson et al. 2014](#); [Blais 2016](#); [Boycott et al. 2017, 2019](#); [Gainotti et al. 2018](#); [Bubela et al. 2019](#); [Grebe et al. 2020](#)). This policy analysis limits itself to the discussion of ethical and legal requirements applicable to rare disease data sharing, rather than discussing the technical considerations relevant to rare disease data interoperability.

In Part I, potential medical and scientific benefits of secondary use of rare disease data are described. In Part II, the different sources of ethico-legal obligations imposed on clinicians and researchers in using health data are outlined. Part III elucidates the common ethico-legal requirements that govern health data use in Canada and assesses their implications for the secondary use of Canadian-source rare disease data. In Part IV, organisational data stewardship models are proposed to promote responsible and equitable access to rare disease data by health care actors and to ensure ethico-legal compliance despite ethico-legal variation and regulatory uncertainty. Part V describes the potential use of tiered access controls and technological safeguards to permit broad data reuse while preserving the privacy of concerned individuals. The analysis is intended to guide clinicians, researchers, health care institutions, and policymakers in the creation of rare disease networks and consortia, and in the integration of Canadian research outputs thereto.

Part I: Potential benefits of secondary rare disease data use

The rare disease context

It is necessary to qualify what is meant in discussing “rare diseases” in this policy analysis. The definition and inclusion threshold for rare disease varies internationally and across different scientific disciplines and regulatory contexts—that is, the definition of a rare disease in health care legislation can differ from its medical or scientific meaning within a particular context ([Clarke et al. 2014](#);

[Richter et al. 2015](#)). Generally, rare diseases individually affect only a small number of patients, but collectively affect a significant proportion of the population. Often, national-level regulatory frameworks set a maximum threshold of affected persons for rare disease status; so long as the number of affected persons does not exceed the threshold, the disease is considered rare. Such thresholds often range from 50 000 to 200 000 affected persons in a national population ([Clarke et al. 2014](#); [Richter et al. 2015](#); [Mukherjee 2019](#); [Haendel et al. 2020](#)) or fall within the range of 1 out of 1500 to 2500 persons in the national population ([Valdez et al. 2016](#); [Wakap et al. 2020](#)). A small number of definitions also incorporate qualitative considerations such as a disease's severity and a disease's hereditary or genetic cause ([Richter et al. 2015](#)).

For the purposes of this policy analysis, rare disease is intended to refer generally to the diseases falling into the above definitions and is not restricted to one regulatory or scientific definition. Especially relevant are those diseases that are sufficiently rare and complex that the use of locally sourced data will often be insufficient to perform research, and uncommon enough that cooperation between primary care providers and external specialists is required to offer successful treatment. The chosen definition does not flow from a scientific discipline but was selected as challenges in ensuring holistic ethical and legal compliance in sharing data are most acute where such factors are present. Moreover, the One for All initiative to which this policy analysis pertains intends the sharing of “omics” data across different health care activities and further serves to justify the chosen scoping criteria for rare disease.

Canadian clinicians, researchers, laboratories, pharmaceutical companies, and health care institutions all generate data relating to rare diseases. The data are created heterogeneously as part of individual treatment, research, and health care planning activities. Increased access to data in health care is generally beneficial. In the rare disease context, eliminating data siloes is not only beneficial, but also necessary. Collectively, data-driven approaches to rare disease care address the paucity of locally available data about rare diseases by repurposing relevant data sets or findings from other sources ([Hilgers et al. 2016](#); [Walkley et al. 2016](#); [Boycott et al. 2020](#)). While not directly addressed in this article, it should be noted that the increasing use of whole genome and exome sequencing in the clinical rare disease context will provide vast amounts of data that could be helpful to current and future patients and their families suffering from rare diseases.

Prospective benefits of rare disease data reuse

Evidence-based medicine often relies on the use of randomized, controlled trials that leverage research cohorts of sufficient size to generate research findings that are statistically significant. Assembling statistically significant cohorts of rare disease patients at one trial site can be burdensome due to the rarity of such diseases and the geographic dispersion of affected individuals and because of ambiguities in rare disease diagnosis that create challenges in assembling stable trial cohorts, as rare diseases are often rediagnosed, which can lead to exclusion from an ongoing study if individual participation therein was predicated on the assumption that the original misdiagnosis was correct ([Hee et al. 2017](#); [Logviss et al. 2018](#); [Zakrezewska et al. 2018](#)). Reusing existing data to perform rare disease research or to better identify prospective trial participants can mitigate difficulties in assembling the large clinical cohorts necessary to perform statistically significant clinical research among local populations for a particular rare disease ([Blais 2016](#)).

The secondary reuse of existing health data or rare disease data has been proposed to facilitate diagnosis, drug discovery, and clinical trial validation, and to confirm drug effectiveness in the context of rare diseases ([Boycott and Ardigo 2018](#); [Das et al. 2018](#); [Thorogood 2020](#)). Scientists have combined rare disease data with data repurposed from other sources to reduce the cost of generating primary data in rare disease research, decreasing the time and expense required to develop novel

drugs and to receive regulatory approvals (Frieden 2017; Das et al. 2018; Li and Song 2020). Such secondary data use has been performed to ensure the critical mass of rare disease data necessary to perform statistically significant quantitative research is available by replacing control groups with historical data or using pre-existing health data to determine whether it is appropriate to rely on a population sample smaller than the usual minimum sample size required for the results of a clinical trial to be considered statistically significant (Bolognani and Pisano 2016; Jansen-van der Weide et al. 2018; Wright et al. 2018; Li et al. 2019; Mulberg et al. 2019; Taroni et al. 2019; Chow and Huang 2020).

Data-intensive methodologies such as data mining and machine learning are increasingly being advocated and utilized to perform rare disease diagnosis, drug discovery, drug validation, and basic science research, among other clinical and research activities (Choudhury et al. 2017; Mears et al. 2017; Dawkins et al. 2018; Jia et al. 2018; Shen et al. 2018, 2019). The use of data mining techniques and big data analysis has been described as especially promising for rare disease research, as the complex causal drivers of rare disease might be best discovered using big data methodologies that can generate potentially meaningful insights from large quantities of genomic, multi-omic, and other health data that would be impossible for human researchers to parse.

Incorporating data from registries and biobanks in rare disease research has been suggested to be more effective than sole reliance on traditional clinical trials (Kinkorová 2016; Garcia et al. 2018). Moreover, the joint use of data from registries and biobanks, or the creation of combine rare disease registry-biobank hybrids in rare disease research has been suggested to be more effective still, as the correlative data regarding disease origin gleaned from longitudinal health registries can be paired with fundamental genomic and basic science insights flowing from research using biobank samples and the data derived from samples (Kinkorová 2016; Luo et al. 2016; Garcia et al. 2018).

Even absent the use of novel, data-driven approaches to rare disease care, efficient pan-Canadian and international data sharing in the rare disease context can be required to enable collaboration among primary care physicians, specialist physicians, and other institutions, such as laboratories, that participate in rare disease care (Blay et al. 2016; Bowdin et al. 2016; Groft and Posada de la Paz 2017; Ormondroyd et al. 2017; Valdez et al. 2017; Gainotti et al. 2018; Byrne et al. 2020). Failure to perform efficient collaboration and knowledge sharing early in the rare disease care process can result in a worse prognosis and decreased patient satisfaction (Hannemann-Weber 2011; Hannemann-Weber et al. 2011). Genomic, multi-omic, and personalized medicine approaches to rare disease care increasingly require the integration and linkage of individual health records to the same individual's health records from other sources or to relevant health records of other individuals or larger sets of research or administrative health care data (Liu et al. 2016, 2019; Burgun et al. 2017; Halfmann et al. 2017; Pogue et al. 2017; Volmar et al. 2017).

For these reasons, enabling health data sharing in Canada and across international boundaries is critical for rare disease patients to benefit from better health care and to profit from new research outputs. Many potential benefits for rare disease patients to be obtained in using Canadian health data in combination with other data sets from other health sector uses have been described nonexhaustively. The ethical and legal principles that guide health data use in the Canadian research and health care contexts will now be considered.

Part II: Competing sources of ethico-legal obligations in rare disease data stewardship

Part I addressed the potential benefits of integrating rare disease data from varied clinical, research, and health care administration activities and of integrating Canadian-source rare disease data to

international data sharing networks. The different sources of ethical and legal obligations applicable to health data sharing in Canada are now considered. Divergence in the content thereof across jurisdictional and sectoral boundaries can cause ethico-legal compliance to appear daunting for health data custodians responsible for pan-Canadian data stewardship. For rare disease patients, the content of data privacy laws and research ethics guidance can sometimes appear to unduly restrict their participation in research. The use of personal data for research, clinical, or quality assurance activities in health care is usually contingent on the informed consent of participants to the use of their data for specified purposes or on the anonymization of the data.

It must, however, be noted that the rare disease context challenges certain fundamental assumptions of data privacy law and research ethics. First, rare disease patients are often minors or are affected by their respective disease in a manner that may impact their capacity to consent or create additional complexities in gathering informed consent. Second, there is significant overlap across clinical care, research, and quality assurance activities in the rare disease context, which can create difficulties in delineating a singular purpose for secondary data use. Last, because of the relative uniqueness of each rare disease, the data of rare disease patients can be impossible to successfully anonymize, defeating anonymization-driven approaches to ethico-legal compliance. Further, rare disease patients often want to remain identifiable (even sharing photographic images for diagnostic purposes), so that novel insights about their disease can easily be relayed to them or to their families. For all of these reasons, certain paradigms of ethical and lawful health data stewardship can be difficult to reconcile with the realities of rare disease health care (Nguyen et al. 2019). Below, I consider how variation in the ethical and legal obligations of different Canadian health care institutions can affect the formulation of a holistic data stewardship strategy for rare disease consortia active across Canada.

Canada's singular federal and multiple provincial governments have implemented laws that govern the use of data within their respective spheres of responsibility (Power 2017; Inions et al. 2018). The federal government has implemented one data privacy law for federal public-sector entities, the *Privacy Act*, and another for Canadian private sector entities engaged in commercial activities, the *Personal Information and Protection of Electronic Documents Act* (PIPEDA) (*Privacy Act* 1985; *Personal Information Protection and Electronic Documents Act* 2000; Power 2017; Inions et al. 2018). Provinces have all implemented data privacy laws governing their public-sector activities, and most have also implemented private sector privacy laws. Some have also implemented health sector privacy laws (Dove and Phillips 2015; Power 2017; Inions et al. 2018; Thorogood 2018; Office of the Privacy Commissioner of Canada 2020a). Researchers are usually required to comply with the Tri-Council Policy Statement, an ethics guidance document that governs all research funded by Canada's federal research funding agencies that involves human participants or identifiable human data or biological materials (Canadian Tri-Council 2018).

This proliferation of laws creates a general difficulty in establishing the combination of laws applicable to specific data use in Canada. Further, creating large clinical care or research collaborations across multiple sites is challenging because data use obligations can differ for each participating institution. Sharing data between institutions operating in different sectors and located in different provinces can create challenges in ensuring ethical and legal compliance, as the number of ethical and legal obligations that must potentially be complied with increases as does the number of implicated institutions (Council of Canadian Academies 2015; Thorogood et al. 2016; Office of the Privacy Commissioner of Canada 2018a).

Health-sector privacy laws will often supersede the application of more general privacy laws (*Personal Health Information Protection Act*, Ontario 2004a; *Personal Health Information Act*, Newfoundland 2008a; *Personal Health Information Privacy and Access Act*, New Brunswick 2009). Provincial private sector privacy laws will sometimes supersede the application of the general federal private sector

PIPEDA for data uses internal to the concerned province, if the federal government has decreed that the concerned provincial law is substantially similar in its content to the concerned federal law ([Office of the Privacy Commissioner of Canada 2004, 2020a; Bastarache 2012; Gratton and Hoffman 2014; Saulnier and Joly 2016](#)). For uses of health data that cross provincial boundaries, the federal PIPEDA generally applies. There is, however, disagreement among federal privacy regulators and Quebec's provincial privacy regulators as to whether provincial privacy laws, the federal PIPEDA, or both apply to data uses that cross provincial or national boundaries ([Décret 1368-2003 2003; Nikser 2006; Article 29 Data Protection Working Party 2014; Power 2017](#)). Further, the determination of which privacy laws apply to certain entities can depend on their engagement in commercial activities or their private or public character. Health-sector entities that operate at the intersections of commercial, governmental, and health care activities can find themselves regulated by disparate and overlapping laws or can face ambiguities regarding the laws applicable to them ([Rodgers v Calvert 2004; Wyndowe v Rousseau 2008; Willison 2009; Gratton and Hoffman 2014; Office of the Privacy Commissioner of Canada 2015a, 2017, 2019a; British Columbia Medical Association 2017](#)).

For these reasons, health data sharing initiatives in Canada do not often result in the creation of holistic infrastructures that gather and centralize large data sets from multiple sources for broad reuse. Instead, participating institutions often negotiate the sharing of limited data sets among themselves. Such releases are sometimes performed in the prospective collection of designated data from consenting participants for broad reuse among collaborators ([Rothstein et al. 2016; Villanueva et al. 2018](#)) in the selective release of data by a health institution to a designated recipient for one specified purpose ([Council of Canadian Academies 2015](#)) or in the establishment of networks between participating nodes that subject the release of specified data sets to a central management process by a designated body rather than requiring the prolonged negotiation of data release conditions anew for each prospective release ([Shabani et al. 2015; The ICGC Data Access Compliance Office and the ICGC International Data Access Committee 2016; Dyke 2020](#)). Thus far, potential benefits for rare disease patients arising from increased sharing of data from Canadian health institutions to other health-sector entities in Canada and abroad have been described. I have considered the potential for the multiplicity of laws and ethical instruments regulating health data use in Canada to prevent the free movement of rare disease data and to challenge the adoption of a singular data stewardship strategy for all institutions participating in a data exchange or a data sharing network. These difficulties having been acknowledged, the common obligations reflected in Canada's research ethics regime, and most data privacy laws in Canada will be delineated.

Part III: Ethico-legal requirements applicable to health data use and health data sharing in Canada

Consent to collect, use, and disclose personal health information

To use health data or share it with another person or institution, the explicit consent of the individual concerned by the data is often required, as health data are considered to be particularly sensitive ([Townsend v Sun Life Financial 2012](#)). For this reason, relying on a person's implied consent to use their health data is generally not appropriate. The consent must in certain instances be recorded in writing ([Canadian Tri-Council 2018](#)) or be drafted according to a legislatively required template, as is the case in the province of Alberta ([Health Information Act, Alberta 2000a; Inions et al. 2018](#)).

For research, broad consent to unspecified future research use in combination with ongoing data governance is generally regarded as an acceptable best practice in Canada ([Canadian Tri-Council 2018](#)). The majority of international genomic data sharing consortia, as well as many biobanks, have relied on a model of broad consent supported by ongoing governance ([Villanueva et al. 2018](#)). Explicit consent is generally not required to transfer health data to a contracted third-party service provider.

It is, however, required to transfer health data to a third party intending to use the data for their own purposes.

Canadian research ethics requires that consent to clinical care and consent to research be clearly distinguished from one another ([Canadian Tri-Council 2018](#)). The informed consent requirements for research activities are generally more onerous than for clinical activities, as participation in research usually does not lead to direct health benefits for the research participants. Therefore, all plausible risks should be disclosed to an individual considering participation in a research consortium. Participants must also be informed of intended data linkages and intended data reuse, as well as any anticipated data sharing practices ([Canadian Institutes of Health Research 2005](#); [Canadian Tri-Council 2018](#)). In the rare disease context, it is especially important to follow these requirements in collecting data for later research reuse during clinical care or in performing clinical trials with the intent of using the collected data for secondary research purposes. Overall, it is recommended that rare disease research networks harmonize their consent-gathering practices across the clinical and research settings in participating nodes and obtain affirmative consent to future data use for broad secondary purposes. Local consent requirements must be respected at each participating site.

Other justifications in ethics and law to collect, use, and disclose personal health information

Canadian ethics guidance and law allows the use and sharing of health data without individual consent for a limited number of purposes. The permissible uses vary across provinces and across categories of entities (e.g., clinics, hospitals, universities). Further, many of these require that data releases remain internal to the province in which the data was originally collected. Permissible purposes generally include quality assurance; secondary research reuse subject to certain conditions, to provide care to a person in immediate danger; and to provide health care to the individual concerned by the data (*Act Respecting Health and Social Services*, Quebec 1991a; *Freedom of Information and Protection of Privacy Act*, British Columbia 1996a; *Personal Health Information Act*, Manitoba 1997; *Health Information Protection Act*, Saskatchewan 1999; *Health Information Act*, Alberta 2000b; *Personal Information Protection Act*, British Columbia 2003; *Personal Health Information Protection Act*, Ontario 2004b; *Personal Health Information Act*, Newfoundland 2008b; Power 2017; Inions et al. 2018). Allowing such secondary reuse is left to the discretion of the health data custodian—generally, the hospital administrator or other official responsible for the safekeeping of the data upon primary collection. Data collections, uses, and disclosures absent consent must be individually authorized by the concerned health data custodian on a case-by-case basis. In practice, permissible data disclosures are often denied even in circumstances that could benefit the public good, for fear of sanctions being imposed if the data release were to engender harmful outcomes ([Council of Canadian Academies 2015](#)).

Because the reasons for which such disclosures can be made vary considerably across different laws and because health data custodians can discretionarily choose to refuse data releases, it is difficult to imagine exceptions to consent serving as a common basis for the broad secondary reuse of clinical and research data across Canada for general purposes. Such mechanisms are most appropriate to the release of data for specific and delimited purposes, rather than to justify the wholesale transfer of health records to a centralized rare disease data bank or justify broad reuse by third parties. Ministers, public health authorities, and a small number of other actors have wide powers to requisition and centralize health data for broad reuse and to designate health registries and prescribed entities responsible for such activities. Naturally, these powers are not readily at the disposal of health sector entities independently desiring to share or centralize health data ([von Tigerstrom et al. 2000](#);

von Tigerstrom and Ries 2009; Power 2017; Inions et al. 2018; Institute for Clinical Evaluative Sciences 2019; Bernier and Knoppers 2020).

In performing health data sharing, and in receiving and reusing health data, entities must ensure that a clear justification for the proposed use exists in law—and, if performing research, in research ethics guidance (Canadian Tri-Council 2018). To prospectively create a general-purpose central repository of data, obtaining explicit consent is a recommended best practice. To repurpose existing data to create a general-purpose central repository of data, the involvement of authorities may be required. For health networks that intend to perform health data exchanges on a restrictive basis, having the involved institutions ensure the ethical and legal compliance of the exchange absent the explicit consent of the concerned individual will generally be required. Many other formalities govern the use of health data in Canada. The more common ones are briefly discussed below.

Returning research results to individuals

Canadian research ethics guidance can require the creation of an incidental findings plan for research, which delineates whether research participants should anticipate the return of individual research results or incidental findings. Generally, the return of findings that could have actionable medical implications for participants must be performed if participants request such return. Informing research participants of life-threatening primary or secondary research findings can be required even if the participants have refused the return of research results or of incidental findings (Canadian Tri-Council 2018; Thorogood et al. 2019). Clinicians, researchers, and consortia often develop proportionate mechanisms for keeping research participants engaged and informed even if direct participant recontact or return of results are prohibitively resource-intensive or expressly precluded in participant consents. Such mechanisms can include the use of online publications or letters to provide general project-related information to interested parties (Canadian Institutes of Health Research 2005; Gainotti et al. 2016). Whatever approach is favoured, information should be provided to research participants in a concise and comprehensible form suitable to a lay audience. If possible, methods of providing direct notice to interested parties should reflect their expressed preferences as to the frequency and specificity of communications (Christofides et al. 2019).

Maintaining records for governments and performing privacy impact assessments

In addition to the conditional requirement to return medically relevant or actionable information to participants, governments often impose requirements to create reports and records prior to or after collecting, using, and disclosing health records. For instance, federal public sector entities and provincial public sector entities in British Columbia and Ontario, as well as health sector entities in Alberta, can be required by law to complete a privacy impact assessment (PIA) prior to using personal or health data. Such an assessment is a holistic analysis of intended data use and data governance practices, the potential risks thereof, and anticipated mitigation measures to safeguard against the risks identified (Bayley and Bennett 2012; Office of the Privacy Commissioner of Canada 2020b). Ontario is in the process of amending its health information laws to require the maintenance of records of data use that can be subject to government audit (Legislative Assembly of Ontario 2020). It is consequently recommended that health institutions perform a PIA before engaging in large-scale health data exchange projects and maintain records of their data use activities throughout the entire life of the project.

Requirements to use contracts, research management plans, and funding agreements

Some laws require that data releases to third parties providing data services (storage, analysis, etc.) or using the data for their own purposes, to be governed by data sharing agreements, the minimum required content of which is sometimes enshrined in law (Dove and Phillips 2015). For example, such agreements are required by laws in Alberta, British Columbia, Ontario, and Quebec, (*Act Respecting Health Services and Social Services*, Quebec 1991b; *Freedom of Information and Protection of Privacy Act*, British Columbia 1996b; *Personal Health Information Protection Act*, Ontario 2004c; Williams and Weber-Jahnke 2010; Wakulowsky 2011; Inions et al. 2018). Research plans compliant with statutory requirements must also be submitted to the institutions authorizing data releases for research in some provinces, such as Ontario (Wakulowsky 2011; Dove and Phillips 2015). Federal funding agencies are in the process of implementing requirements to submit data management plans at the grant application stage as a prerequisite to conducting research involving human participants, and they generally require data management plans to be maintained by researchers (Government of Canada 2016, 2018a, 2018b). It is a recommended best practice to use formal contracts that can be audited and enforced to transfer data to third parties for storage or service-related purposes and to draft and adhere to a data governance and data management plan in performing a novel, large-scale use of health data.

Minimizing data collection and limiting the purpose of use

Whether data use is justified by individual consent or another basis in law, overarching principles of data minimization and purpose limitation delimit the allowable uses of health data. Uses of data are generally restricted to the purposes originally consented to or otherwise provided for in law. Only the minimum data required to fulfill such purposes can be utilized or shared (Power 2017; Canadian Tri-Council 2018; Inions et al. 2018; Office of the Privacy Commissioner of Canada 2019b). Certain laws also permit the reuse of data for purposes compatible with the original purpose of collection (Canada Health Infoway 2017; Office of the Privacy Commissioner of Ontario 2017). Entities using health data should clearly establish their intended purposes in collecting data and refrain from collecting or storing any data that is wholly unrelated to their stated purposes.

Keeping data secure

Canadian ethics and privacy law impose duties to safeguard data using physical protection measures, technological protection measures, staff training, organisational measures, and administrative measures (Canadian Institutes of Health Research 2005; Office of the Privacy Commissioner of Canada 2015b; Flaumenhaft and Ben-Assuli 2018; Inions et al. 2018). Securitization is not a process of ensuring formal compliance with standard-form requirements, but a holistic process that must be accounted for in all aspects of health infrastructure design (Cavoukian 2005; Information and Privacy Commissioner of Ontario 2010; Bender et al. 2017; Inions et al. 2018). Security measures adopted should be proportionate to the data's sensitivity and the circumstances of its use (Office of the Privacy Commissioner of Canada 2003).

Entities using data are generally required to destroy or return all primary copies of the health data used after their research or caregiving activities have been completed. Prior to the requisite destruction of information, data should be preserved to ensure that it cannot be lost or destroyed inadvertently, which requires that secure backups be maintained. Once it is warranted, destruction of information should be performed permanently and irreversibly. It is a general good practice to maintain electronic audit logs of data access, modification, and destruction activities within an informatics system (Dove et al. 2015; Power 2017; Inions et al. 2018). The province of Ontario is in the process of

amending its health information laws to require that such data audit trails be maintained ([Legislative Assembly of Ontario 2020](#)).

Remaining accountable for the use of data

In using data, disclosing data to third parties, or transferring data to third-party service providers without relinquishing control thereof, entities using health data are required to remain accountable for the use thereof ([Office of the Privacy Commissioner of Canada et al. 2012](#); [Kosseim et al. 2014](#); [Power 2017](#); [Inions et al. 2018](#); [Office of the Privacy Commissioner of Canada 2019c](#)). Some Canadian privacy laws, including Quebec's private sector privacy law, require that the data recipient be held accountable to the same standard of privacy obligations as is incumbent on the releasing party (*Act Respecting the Protection of Personal Information in the Private Sector*, Quebec 1993; [Inions et al. 2018](#)). Others require that a holistic principle of accountability be respected, without ascribing specific formal content to the principle ([Office of the Privacy Commissioner of Canada et al. 2012](#)). Accountability requires that organisations take responsibility for their management of personal data throughout the entire organisation and beyond its bounds and inform the public of their data use practices.

Measures recommended by Canadian regulators include hiring staff such as Privacy Officers and—if required by the size of the organization—creating a Privacy Office ([Office of the Privacy Commissioner of Canada et al. 2012](#)). Engaging in intra-institutional receipt and management of privacy complaints and the performance of internal and external organisational and technological audits is also a recommended best practice for remaining accountable in using data. Creating inventories of all personal data held by an organisation has also been recommended by authorities. Implementing policies for the use of data, the management of data, the destruction of data, and ensuring that staff are cognizant of their contents can also be required ([Office of the Privacy Commissioner of Canada et al. 2012](#); [Office of the Privacy Commissioner of Canada 2018b, 2019d](#)). Last, accountability can require the proactive implementation of breach management measures, transparent communication with the public, and recurrent risk-assessment and risk-management ([Office of the Privacy Commissioner of Canada et al. 2012](#); [Inions et al. 2018](#)).

Respecting individual and public rights in data

The obligations of health care entities to safeguard privacy are counterbalanced against certain individual and public rights in data. Canadian laws enshrine the following requirements. First, individuals sometimes have a right to request access to their own data and to require its correction, destruction, or amendment under specified circumstances (*The Access to Information and Privacy Unit 2008*; [Power 2017](#); [Inions et al. 2018](#)). Individuals have certain other rights in their data, such as the right to be notified of its use and to have inaccuracies in their data rectified. Second, public institutions can be required to release data to members of the broad public that request such data. Public access requests must generally be honored by the institution unless the release would infringe individual privacy, certain protected State interests or commercial interests, or prove unduly onerous to comply with ([Power 2017](#); [Inions et al. 2018](#)). Third, parties conducting clinical trials can be required to release anonymised clinical trial data to the public once de-identification processes have been applied thereto ([Health Canada 2019](#)). If exceptions to the obligation to release data are invoked, health care entities and public institutions can often be required to de-identify or sever the personal data from the concerned record and proceed to release the resulting anonymised record (*The Access to Information and Privacy Unit 2008*; [Scassa and Conroy 2016](#)). Institutional policies should be developed to consistently address requests to access, amend, or destroy data. These policies should ensure that individuals can easily submit requests to the organization, and the requests will be conveyed to the concerned actors and addressed within the time limits prescribed by law. Most laws

require such requests to be addressed within one month of their being made, subject to limited exceptions (Power 2017; Inions et al. 2018).

Performing breach notification

In some Canadian provinces, individuals must be notified if their data are subject to a security breach. Other provinces instead require that notice be made out to the concerned enforcement body, usually the Privacy Commissioner, who will decide whether and how individuals should be notified (Cavoukian 2005; Wilkinson 2010; Power 2017; Inions et al. 2018). The federal PIPEDA requires records of breaches to be kept for two or more years after the breach occurs (Breach of Security Safeguards Regulations 2018). Breach notification can also be performed voluntarily. Privacy Commissioners have historically collaborated with voluntarily disclosing institutions to mitigate the harms of data breaches (Cavoukian 2009).

Creating a robust internal protocol for consistent breach follow-up is a recommended practice for health institutions (Inions et al. 2018). Sound data governance, and the aforementioned principle of accountability, require that internal procedures be developed to report potential breaches to health institution leadership and that staff be adequately trained to identify and address prospective data breach risks (Office of the Privacy Commissioner of Canada 2018b, 2019d). The potential benefits of rare disease data reuse, and the general ethical and legal obligations governing health data use in Canada having been addressed, it is relevant to consider practical data stewardship approaches that can help ensure ethico-legal compliance while maximizing the potential to reuse rare disease data.

Part IV: Organisational strategies for ethico-legal compliance in rare disease data sharing

The previous section described the principal ethico-legal requirements that govern the use of data by health sector entities in Canada and the potential for perceived difficulties in ensuring ethico-legal compliance to preclude the transfer of Canadian-sourced rare disease data to international clinical care or international research initiatives. To ensure the regulatory compliance of widespread rare disease data sharing, two general approaches to data stewardship can be proposed. The first approach is organizational, and it relies on the appropriate control of data flows to ensure that ethical and legal obligations are discharged by each network participant receiving or transmitting data. The second approach is technological, and it relies on computational safeguards to permit parties to access and share the nonidentifiable outputs of data-analysis processes without directly accessing or sharing any underlying identifiable data.

Organisational strategies for rare disease data stewardship

Organisational strategies for rare disease data stewardship aim to ensure that data are accessed exclusively by individuals holding appropriate ethico-legal permissions, that the data are transmitted for ethically and legally sound purposes, that the data will be subject to requisite safeguards and timely destruction, and that the data will not be further disclosed to third parties absent appropriate authorisations. A combination of such approaches has been implemented by international rare disease data sharing consortia and health data sharing consortia more generally. The following mechanisms are proposed to enable such an approach, namely, the use of mandatory minimum consent elements, retrospective consent filters, data access oversight bodies, registered-access and controlled-access mechanisms for data access, and data use agreements establishing the responsibilities of downstream data recipients. Data can be made available using centralized access models which make data subject to compatible consent and ethico-legal reuse conditions widely available to participants within a common network. Alternatively, data sharing networks can be used to connect nodes irrespective of

common ethico-legal permissions applicable to the data sets shared. In the latter instance, network participants bear a greater burden in ensuring that their data releases and data access practices are compliant with their local ethical and legal requirements.

Minimum consent elements and retrospective consent filters

The first approach entails the creation of a common minimum of required permissions that data contributors (clinicians, researchers, etc.) must attest to before being able to submit data to the infrastructure. Self-assessment tools and retrospective consent filters can be used to compare the permissions the contributors have in the data to the minimum consent elements and ensure that the data conforms to the required minimum elements (Wallace et al. 2020). Such an approach would require a consortium to agree to the minimum required ethico-legal permissions to contribute data. The principal disadvantage of such an approach is that data sets that do not conform to the required minimum elements would be excluded from the infrastructure. Health care practitioners today differ considerably in their consent and data stewardship practices and so this approach could exclude much pre-existing data (Fowler et al. 2017). The Global Alliance for Genomics and Health and the International Rare Diseases Research Consortium have composed robust lists of recommended minimum consent elements that can serve as the basis of a common consent policy for a pan-Canadian rare disease research consortium, and it could be repurposed to create a retrospective consent filter for a consortium integrating retrospectively collected data (Nguyen et al. 2019; Global Alliance for Genomics and Health 2020). Such a filter would be intended to help prospective data contributors gauge whether the data they have collected complies with the minimum required permissions to contribute data to the consortium. The alternative approach is to include data sets governed by different permissions in the same infrastructure.

Matching compatible data sets using standardized representations of ethico-legal permissions and restrictions

If data sets subject to differing permissions are integrated to a common network, parties contributing and accessing data will bear the onus of ensuring that the anticipated uses of the data are compatible with the ethico-legal permissions applicable to the concerned data sets. This can be facilitated in creating standardized representations of the specific ethico-legal permissions in data using standard templates. It can also be facilitated by applying general “tags” to identify data sets bearing common ethico-legal permissions. Existing ontologies and templates such as Consent Codes, the Data Use Ontology, or the Automated Discovery and Access Matrix can be used for such purposes (Dyke et al. 2016b; Cabili et al. 2018; Woolley et al. 2018). Using these templates allows scientists from across different contributor nodes and even across separate consortia to ensure that the permissions in their health data can be meaningfully compared. I have considered two alternate methods for ensuring that the permissions enshrined in data integrated to a data sharing network are compatible with anticipated uses. Prospective oversight mechanisms will now be discussed.

Responsible access to data

Consortia can use standardized oversight bodies and contractual agreements to ensure that data users will make ethical, lawful, and responsible use of health data. Consortia can use a web-based portal to receive data access requests. Such data access requests can then be reviewed by a Data Access Committee (DAC), Data Access Compliance Office (DACO), or other equivalent body responsible for authorizing data access requests (Council of Canadian Academies 2015; Shabani et al. 2015, 2016; Wong et al. 2017; Villanueva et al. 2018). Generally, prospective data users and their institutions will jointly sign a data access agreement detailing their anticipated data uses, demonstrating their institutional affiliation, confirming their ability to perform the anticipated activities (i.e., research,

care, quality assurance), and agreeing to the conditions imposed by the data provider (Dyke et al. 2016a, 2016c; Dyke 2020). The DAC or DACO is responsible for reviewing the agreement and assessing if the proposed data use is reasonable and permissible in the circumstances.

DACs and DACOs

DACs and DACOs can be coordinated either by the institution that has contributed the data to the consortium or centrally by a singular data access committee. Using contributor-specific DACs can be beneficial if the different data sets bear differing permissions, as the contributors will be especially familiar with the permissions inherent in the data. Using a singular centralized DAC for an entire health data sharing network can help ensure that access to a consortium's data is equitable for all users and that application procedures are streamlined. Further, a centralized DAC can be outfitted with personnel specialized in ethics, medicine, science, and law that are particularly well-placed to understand the ethico-legal privacy risks inherent in a proposed data use and determine if the proposed data use is scientifically feasible (Shabani et al. 2016; Dyke 2020). Mechanisms for granting responsible access to data have been described. Contractual mechanisms for ensuring downstream accountability for parties accessing the data will now be considered.

Data access agreements and responsibility for downstream use

It is a recommended best practice that data access agreements be structured as enforceable contracts but be simple in structure so that data users can easily understand their responsibilities (Joly et al. 2011; Knoppers 2014; Dyke et al. 2016c; The ICGC Data Access Compliance Office and the ICGC International Data Access Committee 2016). Intelligibility for signatories may prove more conducive to encouraging compliance than obscurely written contracts that are not digestible for accessing researchers with no legal training (Knoppers et al. 2013; Saulnier et al. 2019).

Canadian ethics and law often require data access agreements to be binding and enforceable against data recipients and third-party service providers (Flaumenhaft and Ben-Assuli 2018; Inions et al. 2018). The Canadian contributor or consortium may remain responsible to Canadian authorities for the use of data in the hands of third-party processors outside Canada or when using cloud services to host or analyze data. For these reasons, data contributors are best protected if binding requirements are imposed on downstream recipients (Office of the Privacy Commissioner of Canada 2009, 2011). Canadian laws can require that these agreements be auditable by the entity providing the data to ensure that a standard of demonstrable accountability is upheld (Office of the Privacy Commissioner of Canada 2009, 2012).

The content of these agreements can vary depending on the structure of a specific consortium. For a rare disease consortium that draws together data from different sorts of entities, it could be difficult to create a harmonized data access agreement, as the legal requirements governing each of the participating institutions may prove too disparate. Harmonizing data access agreements across a consortium where possible is beneficial, as it creates consistency in application processes and subjects researchers accessing multiple data sets to harmonious obligations for all data sets accessed (Saulnier et al. 2019).

Requiring the acknowledgment of contributors for data used in research publications, mandating confidentiality, and requiring that necessary ethical and legal permissions be obtained as preconditions to data access are common contractual elements (Saulnier et al. 2019).

Part V: Tiered data access models and technological privacy safeguards

Having discussed general systems for ensuring that data contribution and data access is performed in compliance with the ethico-legal obligations of data contributors and parties accessing data, it is appropriate to consider more sophisticated governance arrangements that recognize the potential for differing data elements within a database to present dissimilar degrees of privacy risk and consequently requiring different stewardship practices. Technological mechanisms that can allow for the harmonization of individual privacy interests and secondary data analysis for research or clinical care purposes are also addressed.

Tiered access

The first approach considered is a “tiered access” approach, which establishes multiple tiers of data access permissions. Tiered access can impose increasingly restrictive tiers of data access to safeguard data of differing degrees of sensitivity. The least restrictive access tier, public access, can be used to house data that is not sensitive, such as certain aggregate or anonymised data, or data that are subject to the required ethico-legal approvals and sufficient consent to be made fully public (Joly et al. 2016; Dyke et al. 2018).

The second most restrictive tier, registered access, requires individuals to register as a prerequisite to submitting data access requests or accessing certain data sets. This can mean requiring simple registration (username and password) or a more rigorous demonstration of institutional credentials (Dyke et al. 2016a, 2018; Joly et al. 2016). Such an access model can be used to safeguard a genetic data search platform if the mere results of database-level search queries could reveal identifiable personal information in the database, as can be the case for databases containing rare genomic variants, wherein the mere presence of a rare genomic variant could reveal the presence of a specific individual or of one among a small number of individuals (Fiume et al. 2019). Registered access is also useful if data could be misused if it were made accessible to the broad public, but the ethico-legal permissions in the data allow for broad reuse by accredited researchers. Last, registered access can be used to prevent systematic data mining or data scraping techniques from being applied to a data set to attempt re-identification or engage in other privacy-invading practices (Goldfein and Keyte 2017).

The most restrictive tier of data access is managed, or controlled, access. Such an access tier is best suited to hold data if the ethico-legal permissions inherent in the data are sufficiently restrictive that the data must be released for specific prescribed purposes rather than for general reuse, or if there is a high risk that combining multiple data sets within a database would create heightened privacy risks due to data linkage. Controlled access may also be best suited to granting selective access to data sets if a database uses multiple decentralized oversight committees rather than a singular access committee—or if the permissions in the constituent records or data sets differ such that access to data sets must be granted individually (Dyke et al. 2016a; Joly et al. 2016).

Technological privacy safeguards

Technological safeguards can be implemented to allow for the meaningful analysis and use of sensitive health data without jeopardizing the privacy of the individuals to whom the analysed data pertains (Joly et al. 2016; Fiume et al. 2019).

Technological safeguards fall into two general families. First, there are disclosure controls that release health data to requesting parties in limited quantities or in modified form, ensuring that the data released does not allow individual identification or otherwise reveal sensitive information (Lin et al. 2016; Raisaro et al. 2017; Wang et al. 2017; Fiume et al. 2019; von Thenen et al. 2019; Ayday 2020).

Such mechanisms include the use of differential privacy methods, which can allow individuals to search among the aggregate records of a database but add noise to the results returned such that no single record can sufficiently skew the results to reveal that the data of said component record is comprised in the overall data set ([Dankar and El Emam 2013](#)).

Beacon systems are another form of disclosure control. Beacon systems allow researchers to query for the presence of specific genomic variants within a database of aggregate genomic records, but they ensure that only a limited number of results are returned from one individual genomic record in a data set. This mechanism prevents third parties from inferring that a specific individual's record is present in the aggregate data set using strategic queries of the concerned individual's rare genetic variants ([Raisaro et al. 2017](#); [Wang et al. 2017](#); [Fiume et al. 2019](#); [von Thenen et al. 2019](#); [Ayday 2020](#)).

Second, there are distributed learning mechanisms that perform computational data analysis directly using sensitive health data, but only release the nonsensitive outputs of such analysis to the human parties performing the analysis ([Deist et al. 2017](#); [Bogowicz et al. 2020](#); [Duan et al. 2020](#); [Zerka et al. 2020](#)). Such mechanisms can alleviate potential ethico-legal difficulties in sharing data outside of its site of collection by allowing local nodes to perform relevant data analysis alone and aggregating the nonsensitive results of such analysis across participating nodes.

Conclusion: Holistic principles for data flow management and health system design

In conclusion, three governing concepts are synthesized from the foregoing analysis that can ensure holistic ethical and legal data flow management—namely, the appropriate delegation of responsibilities within and across participating institutions, the efficient communication of information to stakeholders, and the integration of privacy to organisational and technological design ([Cavoukian 2011](#)).

The appropriate delegation of responsibilities can help create pan-Canadian rare disease data sharing infrastructures composed of institutions subject to unharmonized ethical and legal obligations. In requiring that each contributing node discharge its own obligations prior to data submission, and in ensuring that data recipients continue to uphold those obligations, a chain of responsible data transmission can be ensured. This can be achieved even if no single entity or person understands the total relationship between the responsibilities of all participants in a data exchange. Internally, data sharing institutions can use delegated responsibilities and efficient correspondence channels to ensure that staff members are trained to identify and address potential privacy difficulties or to report potential privacy difficulties to specialized personnel.

The transparent and efficient communication of information can require ensuring that the individuals concerned by health data are informed of their rights and the anticipated uses of their data at all stages of the data lifecycle. This requires designing consent materials and consent processes that provide meaningful and understandable information to participants. It also requires the use of accessible and secure technological platforms to disseminate relevant information to participants throughout the course of a consortium's existence. Transparent and efficient communication also requires that consortium partners receiving data are given a clear explanation of their rights and obligations, without requiring help from specialized ethico-legal personnel within their institutions. This can be achieved, for instance, by adopting consortium-wide policies that establish obligations regarding onward transfer of data, standards for data linkage, reidentification, prohibition of certain uses of data, and requisite security measures.

Designing for privacy first requires the integration of technological measures that grant meaningful data privacy guarantees to the data stored. It also requires ensuring not only that each component

of the data infrastructure complies with formal privacy requirements, but also that the ensuing composite of networked infrastructures and institutions offers coherent privacy protections across the ensemble of its activities. Last, designing for privacy requires conceptualizing privacy not as a burden to discharge but as value-added to the entire system. Designing for privacy can encourage public and institutional trust in a health system that can stimulate funding, increase population engagement, and entice external stakeholders to integrate their own infrastructures to the larger network.

Funding acknowledgements

One4ALL: “Sharing Big Data for Health Innovation Advancing the Objectives of the Global Alliance for Genomics and Health (GA4GH) Regulatory and Ethics Work Stream” Genome Canada/CIHR (2019–2022).

Acknowledgements

I wish to extend my sincere thanks to Bartha Maria Knoppers and Michael Beauvais for offering prescient substantive and stylistic input on multiple drafts of this manuscript.

Author contributions

AB conceived and designed the study. AB contributed resources. AB drafted or revised the manuscript.

Competing interests

The author has declared that no competing interests exist.

Data availability statement

All relevant data are within the paper.

References

Act Respecting Health and Social Services, Revised Statutes of Quebec. 1991a, chapter S-4.2 at ss.19, 19.1, 19.2.

Act Respecting Health Services and Social Services, Revised Statutes of Quebec. 1991b, chapter S-4.2, s. 27.1.

Act Respecting the Protection of Personal Information in the Private Sector, Statutes of Quebec. 1993, chapter P-39.1 at s. 17.

Article 29 Data Protection Working Party. 4 June 2014. Opinion 7/2014 on the protection of personal data in Quebec. 14/EN WP 2019.

Ayday E. 2020. Information leaks in genomic data: inference attacks. *In* Responsible genomic data sharing: challenges and approaches. Edited by X Jiang and H Tang. Academic Press. pp. 33–49. DOI: [10.1016/B978-0-12-816197-5.00003-6](https://doi.org/10.1016/B978-0-12-816197-5.00003-6)

Bastarache M. 2012. The constitutionality of PIPEDA: a re-consideration in the wake of the Supreme Court of Canada’s *Reference re Securities Act*. Heenan Blaikie LLP [online]: Available from [google.com/search?q=quebec+pipeda+challenge&ie=utf-8&oe=utf-8](https://www.google.com/search?q=quebec+pipeda+challenge&ie=utf-8&oe=utf-8).

Bayley RM, and Bennett CJ. 2012. Privacy impact assessments in Canada. *In* Privacy impact assessment. Edited by D Wright. Springer, Dordrecht, the Netherlands. pp. 161–185. DOI: [10.1007/978-94-007-2543-0](https://doi.org/10.1007/978-94-007-2543-0)

Bender JL, Cyr AB, Arbuckle L, and Ferris LE. 2017. Ethics and privacy implications of using the Internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. *Journal of Medical Internet Research*, 19(4): e104. PMID: [28385682](https://pubmed.ncbi.nlm.nih.gov/28385682/) DOI: [10.2196/jmir.7029](https://doi.org/10.2196/jmir.7029)

Bernier A, and Knoppers BM. 2020. Health data sharing in Canada. Manuscript accepted for publication in *Health Law in Canada Journal*.

Blais CM. 2016. Evaluating health policy and legal responses: how to reduce barriers and improve access to orphan drugs for rare diseases in Canada. Masters' thesis (LLM), Sherbrooke University, Sherbrooke, Quebec [online]: Available from savoirs.usherbrooke.ca/bitstream/handle/11143/9592/Blais_Catherine_Marie_LLM_2016.pdf?sequence=4&isAllowed=y.

Blay J, Coindre J, Ducimetière F, and Ray-Coquard I. 2016. The value of research collaborations and consortia in rare cancers. *The Lancet Oncology*, 17(2): e62–e69. PMID: [26868355](https://pubmed.ncbi.nlm.nih.gov/26868355/) DOI: [10.1016/S1470-2045\(15\)00388-5](https://doi.org/10.1016/S1470-2045(15)00388-5)

Bogowicz M, Jochems A, Deist TM, Tanadini-Lang S, Huang SH, Chan B, et al. 2020. Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer. *Scientific Reports*, 10: 4542. PMID: [32161279](https://pubmed.ncbi.nlm.nih.gov/32161279/) DOI: [10.1038/s41598-020-61297-4](https://doi.org/10.1038/s41598-020-61297-4)

Bolignano D, and Pisano N. 2016. Good-quality research in rare diseases: trials and tribulations. *Pediatric Nephrology*, 31(11): 2017–2023. PMID: [26817476](https://pubmed.ncbi.nlm.nih.gov/26817476/) DOI: [10.1007/s00467-016-3323-7](https://doi.org/10.1007/s00467-016-3323-7)

Bowdin S, Gilbert A, Bedoukian E, Carew C, Adam MP, Belmont J, et al. 2016. Recommendations for the integration of genomics into clinical practice. *Genetics in Medicine*, 18(11): 1075–1084. PMID: [27171546](https://pubmed.ncbi.nlm.nih.gov/27171546/) DOI: [10.1038/gim.2016.17](https://doi.org/10.1038/gim.2016.17)

Boycott KM, and Ardigo D. 2018. Addressing challenges in the diagnosis and treatment of rare diseases. *Nature Reviews Drug Discovery*, 17(3): 151–152. PMID: [29242613](https://pubmed.ncbi.nlm.nih.gov/29242613/) DOI: [10.1038/nrd.2017.246](https://doi.org/10.1038/nrd.2017.246)

Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. 2017. International cooperation to enable the diagnosis of all rare genetic diseases. *American Journal of Human Genetics*, 5(4): 695–705. PMID: [28475856](https://pubmed.ncbi.nlm.nih.gov/28475856/) DOI: [10.1016/j.ajhg.2017.04.003](https://doi.org/10.1016/j.ajhg.2017.04.003)

Boycott KM, Lau LP, Cutillo CM, and Austin CP. 2019. International collaborative actions and transparency to understand, diagnose, and develop therapies for rare diseases. *EMBO Molecular Medicine*, 11(5): e10486. PMID: [30979709](https://pubmed.ncbi.nlm.nih.gov/30979709/) DOI: [10.15252/emmm.201910486F](https://doi.org/10.15252/emmm.201910486F)

Boycott KM, Campeau PM, Howley HE, Pavlidis P, Rogic S, Oriel C, et al. 2020. The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: connecting understudied genes to model organisms. *The American Journal of Human Genetics*, 106(2): 143–152. PMID: [32032513](https://pubmed.ncbi.nlm.nih.gov/32032513/) DOI: [10.1016/j.ajhg.2020.01.009](https://doi.org/10.1016/j.ajhg.2020.01.009)

Breach of Security Safeguards Regulations. 2018. SOR/2018-64.

British Columbia Medical Association. 2017. British Columbia physician privacy toolkit: a guide for physicians in private practice. 3rd edition.

Bubela T, Genuis SK, Janjua N, Krajden M, Mittmann N, Podolak K, et al. 2019. Medical information commons to support learning healthcare systems: examples from Canada. *The Journal of Law, Medicine & Ethics*, 47(1): 97–105. PMID: [30994059](#) DOI: [10.1177/1073110519840488](#)

Burgun A, Bernal-Delgado E, Kuchinke W, van Staa T, Cunningham J, Lettieri E, et al. 2017. Health data for public health: towards new ways of combining data sources to support research efforts in Europe. *Yearbook of Medical Informatics*, 26(1): 235–240. PMID: [29063571](#) DOI: [10.15265/IY-2017-034](#)

Byrne N, Turner J, Marron R, Lambert DM, Murphy DN, O’Sullivan G, et al. 2020. The role of primary care in management of rare diseases in Ireland. *Irish Journal of Medical Science*, 189(3): 771–776. PMID: [31933130](#) DOI: [10.1007/s11845-019-02168-4](#)

Cabili MN, Carey K, Dyke SOM, Brookes AJ, Fiume M, Jeanson F, et al. 2018. Simplifying research access to genomics and health data with Library Cards. *Scientific Data*, 5(1): 180039. PMID: [29537396](#) DOI: [10.1038/sdata.2018.39](#)

Canada Health Infoway. 2017. Secondary use governance across Canada: common understandings of the Pan-Canadian Health Information Privacy Group [online]: Available from [infoway-inforoute.ca/en/component/edocman/resources/reports/privacy/3356-secondary-use-governance-across-canada-common-understandings](#).

Canadian Institutes of Health Research. 2005. CIHR best practices for protecting privacy in health research.

Canadian Tri-Council (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council). 2018. Tri-council policy statement: ethical conduct for research involving humans.

Cavoukian A. 2005. Fact sheet: safeguarding personal health information. Office of the Information and Privacy Commissioner of Ontario, Toronto, Ontario.

Cavoukian A. 2009. Privacy externalities, security breach notification, and the role of independent oversight. Paper presented at: the Eighth Workshop on the Economics of Information Security, University College London, London, England, 24–25 June 2009. Information and Privacy Commissioner, Toronto, Ontario.

Cavoukian A. 2011. Privacy by design: the 7 foundational principles, implementation and mapping of fair information practices. Information and Privacy Commissioner of Ontario, Toronto, Ontario.

Choudhury AR, Cheng T, Phan L, Bryant SH, and Wang Y. 2017. Supporting precision medicine by data mining across multi-disciplines: an integrative approach for generating comprehensive linkages between single nucleotide variants (SNVs) and drug-binding sites. *Bioinformatics*, 33(11): 1621–1629. PMID: [28158543](#) DOI: [10.1093/bioinformatics/btx031](#)

Chow S, and Huang Z. 2020. Innovative design and analysis for rare disease drug development. *Journal of Biopharmaceutical Statistics*, 30(3): 537–549. PMID: [32065047](#) DOI: [10.1080/10543406.2020.1726371](#)

Christofides E, Stroud K, Tullis DE, and O’Doherty KC. 2019. Improving dissemination of study results: perspectives of individuals with cystic fibrosis. *Research Ethics*, 15(3–4): 1–14. DOI: [10.1177/1747016119869847](#)

Clarke JTR, Coyle C, Evans G, Martin J, and Winquist E. 2014. Toward a functional definition of a “rare disease” for regulatory authorities and funding agencies. *Value in Health*, 17(8): 757–761. PMID: [25498770](#) DOI: [10.1016/j.jval.2014.08.2672](#)

Council of Canadian Academies. 2015. Accessing health and health-related data in Canada. The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, Council of Canadian Academies, Ottawa, Ontario.

Daley JM, Priebe J, and Zeller P. 2015. The impact of emerging Asia-Pacific data protection and data residency requirements on transnational information governance and cross-border discovery. *The Sedona Conference Journal*, 16: 201–256.

Dankar FK, and El Emam K. 2013. Practicing differential privacy in health care: a review. *Transactions on Data Privacy*, 5: 35–67.

Das S, Huang S, and Lo AW. 2018. Acceleration of rare disease therapeutic development: a case study of AGIL-AADC. *Drug Discovery Today*, 24(3): 678–684. PMID: [30610920](#) DOI: [10.1016/j.drudis.2018.12.006](#)

Dawkins JS, Draghia-Akali R, Lasko P, Lau LP, Jonker AH, Cutillo CM, et al. 2018. Progress in rare disease research 2010–2016: an IRDiRC perspective. *Clinical and Translational Science*, 11(1): 11–20. PMID: [28796411](#) DOI: [10.1111/cts.12501](#)

Décret 1368-2003. 17 December 2003. Concernant un renvoi à la Cour d’appel relatif à la Loi sur la protection des renseignements personnels et les documents électroniques.

Deist TM, Jochems A, van Soest J, Nallbantov G, Oberije C, Eble M, et al. 2017. Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. *Clinical and Translational Radiation Oncology*, 4: 24–31. PMID: [29594204](#) DOI: [10.1016/j.ctro.2016.12.004](#)

Dove ES, and Phillips M. 2015. Privacy law, data sharing policies, and medical data: a comparative perspective. In *Medical data privacy handbook*. Edited by A Gkoulalas-Divanis and G Loukides. Springer, Cham, Switzerland. pp. 639–678. DOI: [10.1007/978-3-319-23633-9_24](#)

Dove ES, Joly Y, and Knoppers BM. 2015. International genomic cloud computing: ‘mining’ the terms of service. In *Privacy and legal issues in cloud computing*. Edited by ASY Cheung and RH Weber. Elgar Law, Technology and Society. Edward Elgar Publishing, Cheltenham, UK; Northampton, Massachusetts. pp. 245–246.

Duan R, Boland MR, Liu Z, Liu Y, Chang HH, Xu H, et al. 2020. Learning from electronic health records across multiple sites: a communication-efficient and privacy-preserving distributed algorithm. *Journal of the American Informatics Association*, 27(3): 376–385. PMID: [31816040](#) DOI: [10.1093/jamia/ocz199](#)

Dyke SOM. 2020. Genomic data access policy models. In *Responsible genomic data sharing: challenges and approaches*. Edited by X Jiang and H Tang. Academic Press. pp. 19–33.

Dyke SOM, Kirby E, Shabani M, Thorogood A, Kato K, and Knoppers BM. 2016a. Registered access: a ‘Triple-A’ approach. *European Journal of Human Genetics*, 24(12): 1676–1680. PMID: [27677416](#) DOI: [10.1038/ejhg.2016.115](#)

Dyke SOM, Philippakis AA, Rambla De Argila J, Paltoo DN, Luetkemmer ES, Knoppers BM, et al. 2016b. Consent codes: upholding standard data use conditions. *PLoS Genetics*, 12(1): e1005772. PMID: 26796797 DOI: [10.1371/journal.pgen.1005772](https://doi.org/10.1371/journal.pgen.1005772)

Dyke SOM, Saulnier KM, Pastinen T, Bourque G, and Joly Y. 2016c. Evolving data access policy: the Canadian context. *FACETS*, 1(1): 138–147. PMID: 27990475 DOI: [10.1139/facets-2016-0002](https://doi.org/10.1139/facets-2016-0002)

Dyke SOM, Linden M, Lappalainen I, Rambla De Argila J, Knox C, Lloyd DJ, et al. 2018. Registered access: authorizing data access. *European Journal of Human Genetics*, 26(12): 1721–1731. PMID: 30069064 DOI: [10.1038/s41431-018-0219-y](https://doi.org/10.1038/s41431-018-0219-y)

Fiume M, Cupak M, Keenan S, Rambla J, de la Torre S, Dyke SOM, et al. 2019. Federated discovery and sharing of genomic data using Beacons. *Nature Biotechnology*, 37(3): 220–224. PMID: 30833764 DOI: [10.1038/s41587-019-0046-x](https://doi.org/10.1038/s41587-019-0046-x)

Flaumenhaft Y, and Ben-Assuli O. 2018. Personal health records, global policy and regulation review. *Health Policy*, 122(8): 815–826. PMID: 29884294 DOI: [10.1016/j.healthpol.2018.05.002](https://doi.org/10.1016/j.healthpol.2018.05.002)

Fowler SA, Saunders CJ, and Hoffman MA. 2017. Variation among consent forms for clinical whole exome sequencing. *Journal of Genetic Counseling*, 27(1): 104–114. PMID: 28689263 DOI: [10.1007/s10897-017-0127-2](https://doi.org/10.1007/s10897-017-0127-2)

Freedom of Information and Protection of Privacy Act, Revised Statutes of British Columbia. 1996a, chapter 165, ss. 33–33.3.

Freedom of Information and Protection of Privacy Act, Revised Statutes of British Columbia. 1996b, chapter 165, s. 33.1, 35 (1) (d).

Frieden TR. 2017. Evidence for health decision making—beyond randomized, controlled trials. *The New England Journal of Medicine*, 377: 465–475. PMID: 28767357 DOI: [10.1056/NEJMr1614394](https://doi.org/10.1056/NEJMr1614394)

Gainotti S, Turner C, Woods S, Kole A, McCormack P, Lochmüller H, et al. 2016. Improving the informed consent process in international collaborative rare disease research: effective consent for effective research. *European Journal of Human Genetics*, 24(9): 1248–1254. PMID: 26860059 DOI: [10.1038/ejhg.2016.2](https://doi.org/10.1038/ejhg.2016.2)

Gainotti S, Torreri P, Wang CM, Reihls R, Mueiller H, Heslop E, et al. 2018. The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers. *European Journal of Human Genetics*, 26(5): 631–643. PMID: 29396563 DOI: [10.1038/s41431-017-0085-z](https://doi.org/10.1038/s41431-017-0085-z)

Garcia M, Downs J, Russel A, and Wang W. 2018. Impact of biobanks on research outcomes in rare diseases: a systematic review. *Orphanet Journal of Rare Diseases*, 13(1): 202. PMID: 30419920 DOI: [10.1186/s13023-018-0942-z](https://doi.org/10.1186/s13023-018-0942-z)

Genome Canada. 2018. Precision health initiative—rare diseases: clinical implementation projects [online]: Available from genomecanada.ca/en/programs/translation/funding-opportunities/precision-health-initiative-rare-diseases-clinical.

Genome Canada. 2019. Canadian Genomics Partnership for Rare Diseases (CGP4-RD): mission statement [online]: Available from genomecanada.ca/en/programs/precision-health-strategy/canadian-genomics-partnership-rare-diseases-cgp4-rd-mission.

Global Alliance for Genomics and Health. 2020. Global Alliance for Genomics and Health: consent codes (forthcoming).

Goldfein S, and Keyte J. 2017. Big data, web ‘scraping’ and competition law: the debate continues. *New York Law Journal*, 258(49): 1–3.

Government of Canada. 2016. Tri-agency statement of principles on digital data management.

Government of Canada. 2018a. Draft tri-agency statement of principles on digital data management.

Government of Canada. 2018b. Draft tri-agency research data management policy for consultation.

Gratton E, and Hoffman PC. 2014. Privacy, trusts and cross-border transfers of personal information: the Quebec perspective in the Canadian context. *Dalhousie Law Journal*, 37(1): 255–300.

Grebe TA, Khushf G, Chen M, Bailey D, Manace Brenman L, Williams MS, et al. 2020. The interface of genomic information with the electronic health record: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 22(9): 1431–1436. PMID: [32475985](#) DOI: [10.1038/s41436-020-0841-2](#)

Groft SC, and Posada de la Paz M. 2017. Rare diseases: joining mainstream research and treatment based on reliable epidemiological data. *In* Rare diseases epidemiology: update and overview. *Edited by* M Posada de la Paz, D Taruscio, and SC Groft. Springer, Cham, Switzerland. pp. 3–21.

Haendel M, Vasilevsky N, Unni D, Bologna C, Harris N, Rehm H, et al. 2020. How many rare diseases are there? *Nature Reviews Drug Discovery*, 19: 77–78. PMID: [32020066](#) DOI: [10.1038/d41573-019-00180-y](#)

Halfmann SS, Mählmann L, Leyens L, Reumann M, and Brand A. 2017. Personalized medicine: what’s in it for rare diseases? *In* Rare diseases epidemiology: update and overview. *Edited by* M Posada de la Paz, D Taruscio, and SC Groft. Springer, Cham, Switzerland. pp. 387–404.

Hannemann-Weber H. 2011. Influence of team composition and knowledge sharing on the ability to innovate in patient-centred healthcare teams for rare diseases. *Journal of Management & Marketing in Healthcare*, 4(4): 265–272. DOI: [10.1179/1753304X11Y.0000000014](#)

Hannemann-Weber H, Kessel M, Budych K, and Schultz C. 2011. Shared communication processes within healthcare teams for rare diseases and their influence on healthcare professionals’ innovative behavior and patient satisfaction. *Implementation Science*, 6: 40. PMID: [21510848](#) DOI: [10.1186/1748-5908-6-40](#)

Health Canada. 2019. Guidance document on public release of clinical information. Government of Canada, Ottawa, Ontario.

Health Information Act, Revised Statutes of Alberta. 2000a, chapter H-5, at s. 34 (2), 54.

Health Information Act, Revised Statutes of Alberta. 2000b, chapter H-5, at s. 27 (1), 35 (1), 37.3, 46 (1), 47 (1), 49–56.

Health Information Protection Act, Statutes of Saskatchewan. 1999, c H-0.021, ss. 29–30.

Hee SW, Willis A, Smith CT, Day S, Miller F, Madan J, et al. 2017. Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate

analysis of ClinicalTrials.gov. Orphanet Journal of Rare Diseases, 12: 44. PMID: [28253932](#) DOI: [10.1186/s13023-017-0597-1](#)

Hilgers R-D, König F, Molenberghs G, and Senn S. 2016. Design and analysis of clinical trials for small rare disease populations. Journal of Rare Disease Research & Treatment, 1(3): 53–60. DOI: [10.29245/2572-9411/2016/3.1054](#)

Hintze M. 2019. Science and privacy: data protection laws and their impact on research. Washington Journal of Law, Technology, and Arts, 14(2): 103–137.

Information and Privacy Commissioner of Ontario. 2010. Sensors and in-home collection of health data: a privacy by design approach. Information and Privacy Commissioner of Ontario, Toronto, Ontario.

Inions NJ, Tran LE, and Rozovsky LE. 2018. Canadian health information: a legal and risk management guide. 4th edition. LexisNexis Canada, Toronto, Ontario.

Institute for Clinical Evaluative Sciences. 2019. Building Ontario's next-generation smart cities through data governance. Part 1: health data safe haven.

Jansen-van der Weide MC, Gaasterland CMW, Roes KCB, Pontes C, Vives R, Sancho A, et al. 2018. Rare disease registries: potential applications towards impact on development of new drug treatments. Orphanet Journal of Rare Diseases, 13: 154. PMID: [30185208](#) DOI: [10.1186/s13023-018-0836-0](#)

Jia J, Wang R, An Z, Guo Y, Ni X, and Shi T. 2018. RDAD: a machine learning system to support phenotype-based rare disease diagnosis. Frontiers in Genetics, 9: 587. PMID: [30564269](#) DOI: [10.3389/fgene.2018.00587](#)

Joly Y, Zeps N, and Knoppers BM. 2011. Genomic databases access agreements: legal validity and possible sanctions. Journal of Human Genetics, 130(3): 441–449. PMID: [21706183](#) DOI: [10.1007/s00439-011-1044-3](#)

Joly Y, Dyke SOM, Knoppers BM, and Pastinen T. 2016. Are data sharing and privacy protection mutually exclusive? Cell, 167(5): 1150–1154. PMID: [27863233](#) DOI: [10.1016/j.cell.2016.11.004](#)

Kim H, Kim SY, and Joly Y. 2018. South Korea: in the midst of a privacy reform centered on data sharing. Human Genetics, 137(8): 627–635. PMID: [30121900](#) DOI: [10.1007/s00439-018-1920-1](#)

Kinkorová J. 2016. Biobanks in the era of precision medicine: objectives, challenges, and innovation. EPMA Journal, 7: 4. PMID: [26904153](#) DOI: [10.1186/s13167-016-0053-7](#)

Knoppers BM. 2014. International ethics harmonization and the global alliance for genomics and health. Genome Medicine, 6: 13. PMID: [25031613](#) DOI: [10.1186/gm530](#)

Knoppers BM, Chrisholm RL, Kaye J, and Cox D. 2013. A P3G generic access agreement for population genomic studies. Nature Biotechnology, 31(5): 384–385. PMID: [23657386](#) DOI: [10.1038/nbt.2567](#)

Kosseim P, Dove ES, Baggaley C, Meslin EM, Cate FH, Kaye J, et al. 2014. Building a data sharing model for global genomic research. Genome Biology, 15: 430. PMID: [25221857](#) DOI: [10.1186/s13059-014-0430-2](#)

Legislative Assembly of Ontario. 2020. Bill 188, the Economic and Fiscal Update Act.

- Li Q, Zhao K, Bustamante CD, Ma X, and Wong WH. 2019. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genetics in Medicine*, 21: 2126–2134. PMID: [30675030](#) DOI: [10.1038/s41436-019-0439-8](#)
- Li X, and Song Y. 2020. Target population statistical inference with data integration across multiple sources—an approach to mitigate information shortage in rare disease clinical trials. *Statistics in Biopharmaceutical Research*, 12(3): 322–333. DOI: [10.1080/19466315.2019.1654913](#)
- Lin C, Song Z, Song H, Zhou Y, Wang Y, and Wu G. 2016. Differential privacy preserving in big data analytics for connected health. *Journal of Medical Systems*, 40: 97. PMID: [26872779](#) DOI: [10.1007/s10916-016-0446-0](#)
- Liu Z, Fang H, Slikker W, and Tong W. 2016. Potential reuse of oncology drugs in the treatment of rare diseases. *Trends in Pharmacological Studies*, 37(10): 843–857. PMID: [27461952](#) DOI: [10.1016/j.tips.2016.06.010](#)
- Liu Z, Zhu L, Roberts R, and Tong W. 2019. Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: where are we? *Trends in Genetics*, 35(11): 852–867. PMID: [31623871](#) DOI: [10.1016/j.tig.2019.08.006](#)
- Logviss K, Krievens D, and Purvina S. 2018. Characteristics of clinical trials in rare vs. common diseases: a register-based Latvian study. *PLoS ONE*, 13(4): e0194494. PMID: [29614103](#) DOI: [10.1371/journal.pone.0194494](#)
- Luo J, Wu M, Gopukumar D, and Zhao Y. 2016. Big data application in biomedical research and health care: a literature review. *Biomedical Informatics Insights*, 8: 1–10. PMID: [26843812](#) DOI: [10.4137/BII.S31559](#)
- Mears AJ, Schock SC, Hadwen J, Putos S, Dymant KM, Boycott KM, et al. 2017. Mining the transcriptome for rare disease therapies: a comparison of the efficiencies of two data mining approaches and a targeted cell-based drug screen. *npj Genomic Medicine*, 2: 14. PMID: [29263829](#) DOI: [10.1038/s41525-017-0018-3](#)
- Mukherjee K. 2019. Care for rare: spotlight on rare diseases. *Trends in Pharmacological Sciences*, 40(4): 227–228. PMID: [30905358](#) DOI: [10.1016/j.tips.2019.02.008](#)
- Mulberg AE, Bucci-Rechtweg C, Giuliano J, Jacoby D, Johnson FK, Liu K, et al. 2019. Regulatory strategies for rare diseases under current global regulatory statutes: a discussion with stakeholders. *Orphanet Journal of Rare Diseases*, 14(1): 36. PMID: [30736861](#) DOI: [10.1186/s13023-019-1017-5](#)
- Nguyen MT, Goldblatt J, Isasi R, Jagut M, Jonker AH, Kaufmann P, et al. 2019. Model consent clauses for rare disease research. *BMC Medical Ethics*, 20(1): 55. PMID: [31370847](#) DOI: [10.1186/s12910-019-0390-x](#)
- Nikser J. 2006. PIPEDA: a constitutional analysis. *The Canadian Bar Review*, 85: 317–343.
- Office of the Privacy Commissioner of Canada. 2003. Company's collection of medical information unnecessary; safeguards are inappropriate. PIPEDA Case Summary #2003-226.
- Office of the Privacy Commissioner of Canada. 2004. Questions and Answers regarding the application of PIPEDA, Alberta and British Columbia's *Personal Information Protection Acts* [online]: Available from [priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/r_o_p/02_05_d_26/](#).

Office of the Privacy Commissioner of Canada. 2009. Guidelines for processing personal data across borders.

Office of the Privacy Commissioner of Canada. 2011. Cloud computing and privacy.

Office of the Privacy Commissioner of Canada. 2012. Cloud computing for small and medium-size enterprises.

Office of the Privacy Commissioner of Canada. 2015a. The application of PIPEDA to municipalities, universities, schools, and hospitals [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/r_o_p/02_05_d_25/.

Office of the Privacy Commissioner of Canada. 2015b. PIPEDA interpretation bulletin: safeguards [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda-compliance-help/pipeda-interpretation-bulletins/interpretations_08_sg/.

Office of the Privacy Commissioner of Canada. 2017. PIPEDA interpretation bulletin, commercial activity [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda-compliance-help/pipeda-interpretation-bulletins/interpretations_03_ca/.

Office of the Privacy Commissioner of Canada. 2018a. Summary of privacy laws in Canada [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/.

Office of the Privacy Commissioner of Canada. 2018b. Breach of the world anti-doping database. PIPEDA Report of Findings #2018-006.

Office of the Privacy Commissioner of Canada. 2019a. How PIPEDA applies to charitable and non-profit organizations.

Office of the Privacy Commissioner of Canada. 2019b. PIPEDA fair information principle 2—identifying purposes.

Office of the Privacy Commissioner of Canada. 2019c. PIPEDA fair information principle 1—accountability [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/principles/p_accountability/.

Office of the Privacy Commissioner of Canada. 2019d. Investigation into Equifax Inc. and Equifax Canada Co.'s compliance with PIPEDA in light of the 2017 breach of personal information. PIPEDA Report of Findings #2019-001.

Office of the Privacy Commissioner of Canada. 2020a. Provincial laws that may apply instead of PIPEDA [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/r_o_p/prov-pipeda/.

Office of the Privacy Commissioner of Canada. 2020b. Expectations: OPC's guide to the privacy impact assessment process. Government of Canada.

Office of the Privacy Commissioner of Canada, Office of the Privacy Commissioner of Alberta, and Office of the Information and Privacy Commissioner for British Columbia. 2012. Getting

accountability right with a privacy management program [joint statement of multiple privacy commissioners] [online]: Available from priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda-compliance-help/pipeda-compliance-and-training-tools/gl_acc_201204/.

Office of the Privacy Commissioner of Ontario. 2017. Big data guidelines [online]: Available from ipc.on.ca/new-big-data-guidelines/.

Ormondroyd E, Mackley MP, Blair E, Craft J, Knight JC, Taylor J, et al. 2017. Insights from the early experience of a rare disease genomic medicine multidisciplinary team: a qualitative study. *European Journal of Human Genetics*, 25: 680–686. PMID: [28327571](#) DOI: [10.1038/ejhg.2017.37](#)

Peloquin D, DiMaio M, Bierer B, and Barnes M. 2020. Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics*, 28: 697–705. PMID: [32123329](#) DOI: [10.1038/s41431-020-0596-x](#)

Personal Health Information Act, Continuing Consolidation of the Statutes of Manitoba. 1997, c. P33.5 ss. 22–23 (3).

Personal Health Information Act, Statutes of Newfoundland. 2008a, c. P-7.01, s. 12 (1).

Personal Health Information Act, Statutes of Newfoundland. 2008b, c. P-7.01, ss. 37, 39, 40, 41, 42, 45, 47.

Personal Health Information Privacy and Access Act, Statutes of New Brunswick. 2009, c. P-7.05, s. 6 (1).

Personal Health Information Protection Act, 2004, Statutes of Ontario. 2004a, c. 3, Sched. A, s. 8 (1).

Personal Health Information Protection Act, 2004, Statutes of Ontario. 2004b, c. 3, Sched. A, ss. 55–55.14.

Personal Health Information Protection Act, 2004, Statutes of Ontario. 2004c, c. 3, Sched. A., ss.42 (1), 44 (5).

Personal Information Protection Act, Statutes of British Columbia. 2003, ss. 17–23.

Personal Information Protection and Electronic Documents Act, Statutes of Canada. 2000, c. 5.

Pogue RE, Cavalcanti DP, Shanker S, Andrade RV, Aguiar LR, de Carvalho JL, et al. 2017. Rare disease genetic diseases: update on diagnosis, treatment, and online resources. *Drug Discovery Today*, 23(1): 187–195. PMID: [29129805](#) DOI: [10.1016/j.drudis.2017.11.002](#)

Power M. 2017. The law of privacy. 2nd edition. LexisNexis Canada, Toronto, Ontario.

Privacy Act, Revised Statutes of Canada. 1985, c. P-21.

Raisaro JL, Tramèr F, Ji Z, Bu D, Zhao Y, Carey K, et al. 2017. Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. *Journal of the Medical Informatics Association*, 24(4): 799–805. PMID: [28339683](#) DOI: [10.1093/jamia/ocw167](#)

Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al. 2015. Rare disease terminology and definitions—a systematic global review: report of the ISPOR Rare Disease Special Interest Group. *Value in Health*, 18(6): 906–914. PMID: [26409619](#) DOI: [10.1016/j.jval.2015.05.008](#)

Rodgers v Calvert, 244 DLR (4th) 479 49 BLR (3d) 53, paras 44–56 [2004 ON SC].

Rothstein MA, Knoppers BM, and Harrell HL. 2016. Comparative approaches to biobanks and privacy. *The Journal of Law, Medicine & Ethics*, 44(1): 161–172. PMID: [27256132](#) DOI: [10.1177/1073110516644207](#)

Saulnier KM, and Joly Y. 2016. Locating biobanks in the Canadian privacy maze. *The Journal of Law, Medicine & Ethics*, 44(1): 7–19. PMID: [27256119](#) DOI: [10.1177/1073110516644185](#)

Saulnier KM, Bujold D, Dyke SOM, Dupras C, Beck S, Bourque G, et al. 2019. Benefits and barriers in the design of harmonized access agreements for international data sharing. *Scientific Data*, 6(1): 297. PMID: [31792275](#) DOI: [10.1038/s41597-019-0310-4](#)

Scassa T, and Conroy A. 2016. Strategies for protecting privacy in open data and proactive disclosure. *Canadian Journal of Law and Technology*, 14: 215–262.

Shabani M, Dyke SOM, Joly Y, and Borry P. 2015. Controlled access under review: improving the governance of genomic data access. *PLoS Biology*, 13(12): e1002339. PMID: [26720729](#) DOI: [10.1371/journal.pbio.1002339](#)

Shabani M, Thorogood A, and Borry P. 2016. Who should have access to genomic data and how should they be held accountable? Perspectives of Data Access Committee members and experts. *European Journal of Human Genetics*, 24(12): 1671–1675. PMID: [27552966](#) DOI: [10.1038/ejhg.2016.111](#)

Shen F, Liu S, Wang Y, Wen A, Wang L, and Liu H. 2018. Utilization of electronic medical records and biomedical literature to support the diagnosis of rare diseases using data fusion and collaborative filtering approaches. *JMIR Medical Informatics*, 6(4): e11301. PMID: [30305261](#) DOI: [10.2196/11301](#)

Shen F, Zhao Y, Wang L, Mojarad MR, Wang Y, Liu S, et al. 2019. Rare disease knowledge enrichment through a data-driven approach. *BMC Medical Informatics and Decision Making*, 19: 32. PMID: [30764825](#) DOI: [10.1186/s12911-019-0752-9](#)

Taroni JN, Grayson PC, Hu Q, Eddy S, Kretzler M, Merkel PA, et al. 2019. MultiPLIER: a transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Systems*, 8(5): 380–394. PMID: [31121115](#) DOI: [10.1016/j.cels.2019.04.003](#)

The Access to Information and Privacy Unit. 2008. Access to information and privacy process and compliance manual for the OPC. Office of the Privacy Commissioner of Canada, The ATIP Unit, Gatineau, Quebec.

The ICGC Data Access Compliance Office and the ICGC International Data Access Committee. 2016. Analysis of five years of controlled access and data sharing compliance at the International Cancer Genome Consortium. *Nature Genetics*, 48(3): 224–225. PMID: [26906679](#) DOI: [10.1038/ng.3499](#)

Thompson R, Johnston L, Taruscio D, Monaco L, Bérout C, Gut IG, et al. 2014. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *Journal of General Internal Medicine*, 29(Suppl. 3): S780–S787. PMID: [25029978](#) DOI: [10.1007/s11606-014-2908-8](#)

Thorogood A. 2018. Canada: will privacy rules continue to favour open science? *Human Genetics*, 137: 595–602. PMID: [30014188](#) DOI: [10.1007/s00439-018-1905-0](#)

Thorogood A. 2020. International data sharing and rare disease: the importance of ethics and patient involvement. *In* Rare diseases. *Edited by* ZH Wu. IntechOpen, London, UK. DOI: [10.5772/intechopen.91237](https://doi.org/10.5772/intechopen.91237)

Thorogood A, Simkevitz H, Phillips M, Dove ES, and Joly Y. 2016. Protecting the privacy of Canadians' health information in the cloud. *Canadian Journal of Law and Technology*, 14(1): 173–213.

Thorogood A, Dalpé G, and Knoppers BM. 2019. Return of individual genomic research results: are laws and policies keeping step? *European Journal of Human Genetics*, 27: 535–546. PMID: [30622328](https://pubmed.ncbi.nlm.nih.gov/30622328/) DOI: [10.1038/s41431-018-0311-3](https://doi.org/10.1038/s41431-018-0311-3)

Townsend v Sun Life Financial. 2012 FC 550 408 FTR 279.

Valdez R, Ouyang L, and Bolen J. 2016. Public health and rare diseases: oxymoron no more. *Preventing Chronic Disease*, 13: 150491. PMID: [26766846](https://pubmed.ncbi.nlm.nih.gov/26766846/) DOI: [10.5888/pcd13.150491](https://doi.org/10.5888/pcd13.150491)

Valdez R, Grosse SD, and Khoury MJ. 2017. The need for a next-generation public health response to rare diseases. *Genetics in Medicine*, 19(5): 489–490. PMID: [27787501](https://pubmed.ncbi.nlm.nih.gov/27787501/) DOI: [10.1038/gim.2016.166](https://doi.org/10.1038/gim.2016.166)

Villanueva AG, Cook-Deegan R, Robinson JO, McGuire AL, and Majumder MA. 2018. Genomic data-sharing practices. *The Journal of Law, Medicine & Ethics*, 47: 31–40. PMID: [30994063](https://pubmed.ncbi.nlm.nih.gov/30994063/) DOI: [10.1177/1073110519840482](https://doi.org/10.1177/1073110519840482)

Volmar C-H, Wahlestedt C, and Brothers CP. 2017. Orphan diseases: state of the drug discovery art. *Wiener Medizinische Wochenschrift*, 167: 197–204. PMID: [26819216](https://pubmed.ncbi.nlm.nih.gov/26819216/) DOI: [10.1007/s10354-015-0423-0](https://doi.org/10.1007/s10354-015-0423-0)

von Thenen N, Ayday E, and Cicek AE. 2019. Re-identification of individuals in genomic data-sharing beacons via allele inference. *Bioinformatics*, 35(3): 365–371. PMID: [30052749](https://pubmed.ncbi.nlm.nih.gov/30052749/) DOI: [10.1093/bioinformatics/bty643](https://doi.org/10.1093/bioinformatics/bty643)

von Tigerstrom B, and Ries NM. 2009. Cancer surveillance in Canada: analysis of legal and policy frameworks and tools for reform. *Health Law Journal*, 17: 1–49. PMID: [20608314](https://pubmed.ncbi.nlm.nih.gov/20608314/)

von Tigerstrom B, Deschenes M, Knoppers BM, and Caulfield T. 2000. Legal regulation of cancer surveillance: Canadian and international perspectives. *Health Law Journal*, 8(1): 1–94. PMID: [11398220](https://pubmed.ncbi.nlm.nih.gov/11398220/)

Wakap SN, Lambert DM, and Olry A. 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, 28: 165–173. PMID: [31527858](https://pubmed.ncbi.nlm.nih.gov/31527858/) DOI: [10.1038/s41431-019-0508-0](https://doi.org/10.1038/s41431-019-0508-0)

Wakulowsky L. 2011. *Personal Health Information Protection Act: implementing best privacy practices*. LexisNexis Canada, Markham, Ontario.

Walkley SU, Davidson CD, Jacoby J, Marella PD, Ottinger EA, Austin CP, et al. 2016. Fostering collaborative research for rare genetic disease: the example of Niemann-Pick type C disease. *Orphanet Journal of Rare Diseases*, 11: 161. PMID: [27903269](https://pubmed.ncbi.nlm.nih.gov/27903269/) DOI: [10.1186/s13023-016-0540-x](https://doi.org/10.1186/s13023-016-0540-x)

Wallace SE, Kirby E, and Knoppers BM. 2020. How can we not waste legacy genomic research data? *Frontiers in Genetics*, 11: 446. PMID: [32457803](https://pubmed.ncbi.nlm.nih.gov/32457803/) DOI: [10.3389/fgene.2020.00446](https://doi.org/10.3389/fgene.2020.00446)

Wang S, Jiang X, Tang H, Wang X, Bu D, Carey K, et al. 2017. A community effort to protect genomic data sharing, collaboration and outsourcing. *npj Genomic Medicine*, 2: 33. PMID: [29263842](#) DOI: [10.1038/s41525-017-0036-1](#)

Wilkinson MA. 2010. Genealogy and the law in Canada. Ontario Genealogical Society.

Williams J, and Weber-Jahnke JH. 2010. The regulation of personal health record systems in Canada. *Canadian Journal of Law and Technology*, 8: 243–276.

Willison DJ. 2009. Use of data from the electronic health record for health research—current governance challenges and potential approaches. Office of the Privacy Commissioner of Canada, Gatineau, Quebec [online]: Available from [priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2009/ehr_200903/](#).

Wong KM, Langlais K, Tobias GS, Fletcher-Hoppe C, Krasnewich D, Leeds HS, et al. 2017. The dbGaP data browser: a new tool for browsing dbGaP controlled-access genomic data. *Nucleic Acids Research*, 45(D1): D819–D826. PMID: [27899644](#) DOI: [10.1093/nar/gkw1139](#)

Woolley JP, Kirby E, Leslie J, Cabili MN, Rushton G, Hazard JG, et al. 2018. Responsible sharing of biomedical data and biospecimens via the “Automatable Discovery and Access Matrix” (ADA-M). *npj Genomic Medicine*, 3(1): 17. PMID: [30062047](#) DOI: [10.1038/s41525-018-0057-4](#)

Wright CF, McRae JF, Clayton S, Gallone G, Aitken S, Fitzgerald TW, et al. 2018. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1133 families with developmental disorders. *Genetics in Medicine*, 20(10): 1216–1223. PMID: [29323667](#) DOI: [10.1038/gim.2017.246](#)

Wyndowe v Rousseau. 2008 FCA 39 373 NR 301, paras 34–39.

Zakrezewska JM, Palmer J, Bendsten L, Stefano GD, Ettlin DA, Maarbjerger S, et al. 2018. Challenges recruiting to a proof-of-concept pharmaceutical trial for a rare disease: the trigeminal neuralgia experience. *Trials*, 19: 704. PMID: [30587219](#) DOI: [10.1186/s13063-018-3045-1](#)

Zerka F, Barakat S, Walsh S, Bogowicz M, Leijenaar RTH, Jochems A, et al. 2020. Systematic review of privacy-preserving distributed machine learning from federated databases in health care. *JCO Clinical Cancer Informatics*, 4: 184–200. PMID: [32134684](#) DOI: [10.1200/CCI.19.00047](#)