DATA ACCESS AND SHARING BY RESEARCHERS IN GENOMICS

– POLICY BRIEF –

E. KIRBY, A.M. TASSÉ, M. ZAWATI AND B.M. KNOPPERS

CENTRE OF GENOMICS AND POLICY (CGP) AND

PUBLIC POPULATION PROJECT IN GENOMICS AND SOCIETY (P^3G)

MONTRÉAL, MARCH 2018

GenomeQuébec

CGP Centre of Genomics and Policy
Centre de génomique et politiques

McGill

Génome Québec

Established in 2000, Génome Québec is a private, non-profit organization with its headquarters in Montréal. Its mission is to catalyze the development and excellence of genomics research and promote its integration and democratization. Génome Québec is recognized for its assertive leadership in promoting an optimal environment conducive to the advancement of genomics research and the integration of its benefits into priority sectors for Québec. A strong culture of ethics drives its mission, providing assurance that research will be conducted within ethical guidelines acceptable to society at large.

To promote a better understanding and support decision making regarding the complex issues raised by data sharing and access, Génome Québec asked the Centre of Genomics and Policy to produce a Policy Brief on the subject. This document is the result of analysis and research conducted by the authors of the CGP. The views expressed herein do not necessarily reflect those of Génome Québec.

Centre of Genomics and Policy (CGP)

Located within the Génome Québec Innovation Centre at McGill University, the CGP works at the crossroads of law, medicine, and public policy. Applying a multidisciplinary perspective and collaborating with national and international partners, the CGP analyzes the socio-ethical and legal norms influencing the promotion, prevention and protection of human health.

Currently, the CGP conducts research on the ethical and legal issues involved in several areas of genomics research, including personalized health, pediatrics, cancer research, gene therapy, biobanks (population genetics) and the impact of new technologies on privacy.

Public Population Project in Genomics and Society (P^3G)

The Public Population Project in Genomics and Society (P^3G) is a non-profit organization located in Montréal. Dedicated to the ethical and policy management of health research infrastructures, P^3 G supports local and international research and meets ethical requirements and standards, pertaining to the integration of genomics research in clinical settings and personalized medicine.

More specifically, its mission is to lead, catalyze, and coordinate international efforts and expertise to optimize the use of studies, biobanks, research databases and other similar health and social research infrastructures. P^3G's international contributions include the development of the Framework for Responsible Sharing of Genomic and Health-related Data, of the Global Alliance for Genomics and Health (GA4GH) (Knoppers, 2014)[1].

[1] Knoppers, B.M., Framework for Responsible Sharing of Genomic and Health-Related Data, (2014) HUGO Journal 8:3.

About the Authors

Emily Kirby (BSc; LL. B.; M. Env.) is a lawyer and IPAC Coordinator at P^{3}G (Public Population Project in Genomics and Society). In her role as IPAC Coordinator, she develops legal and ethical tools to facilitate the access and sharing of international research in genomics.

Anne Marie Tassé (LL.B., LL.M., M.A., LL.D.) is a lawyer specialized in health law, bioethics and international comparative law. She is the Executive Director of P^{3}G (Public Population Project in Genomics and Society) and an Academic Associate at the Centre of Genomics and Policy at McGill University. As such, she coordinates the legal and ethical aspects of more than ten Canadian and international research projects involving genetic and genomic data.

Ma'n Zawati (LL.B., LL.M.) is a lawyer and the Executive Director of the Centre of Genomics and Policy at McGill University. He is also an Associate Member of the University's Biomedical Ethics Unit. His research focuses on the legal and ethical aspects of biobanking as well as the legal duties and liability of health care professionals in both the clinical and research settings.

Bartha Maria Knoppers (PhD (Comparative Medical Law), Ad. E., O.C., O.Q.), is Full Professor and Director of the Centre of Genomics and Policy (CGP), Faculty of Medicine, Human Genetics, McGill University. She also holds the Canada Research Chair in Law and Medicine (Tier 1, since 2001). She is the founder of the Population Project in Genomics and Society (P^{3}G) and CARTaGENE, Québec's population biobank.

Table of Contents

Executive Summary

The collection, analysis, use and sharing of data for genomics research promise major breakthroughs in health research, more specifically in personalized medicine and for population-based studies. Genomics research generates large, rich datasets that are generally costly to produce and store. Given that a significant share of genomics research funding comes from public sources, it is vital to ensure the optimal use and sustainability of these datasets as a research resource. The access and sharing of data for future use is a simple, efficient and cost-effective way to make the most of existing genomic and health-related data.

That being said, genomics research raises a number of ethical and legal issues pertaining to the protection of privacy, participant consent and identifiability of genetic information. Moreover, researchers encounter hurdles when accessing certain types of information required for genomics research (e.g., information from medical records, administrative records, etc.).

Several governance processes and models are used by government bodies and research consortia to address the ethical and legal challenges stemming from the access and sharing of genomic data. Some jurisdictions, for example, have implemented one-stop access points for administrative data. In addition, research consortia generally adopt genomic data sharing models that take into account the sensitive nature and level of identifiability of the information involved. A number of different tools, policies, regulations and mechanisms (technological or governance) are used to propose a balance between protecting the privacy of participants and facilitating innovative research.

Recommendations to Policymakers

1. Recognize the validity and primacy of broad consent. (p. 10).
2. Eliminate jurisdictional restrictions on the use of cloud computing technologies. (p. 10)
3. Adjust the access to information request process according to the level of risk and anticipated social benefits. (p. 14)
4. Ensure that general or broad consent is recognized by both research ethics committees and administrative bodies/committees in charge of access to information held by government agencies. (p. 15)
5. Ensure that rules governing research without consent do not override a duly obtained consent. (p. 15)
6. Eliminate restrictions on the storage mode and retention period of information obtained from public bodies. (p. 15)
7. Develop mechanisms to streamline access to de-identified information held by public bodies. (p. 15)
8. Set up an organization (“one-stop access point”) whose mission is to provide access to health-related data for research. (p. 15)
9. Draw on innovative information technologies approaches adopted by other jurisdictions for the sharing of sensitive information. (p. 20)
10. Launch a discussion forum among Québec research ethics committees to promote a harmonized approach to the ethical review of research involving genomic data. (p. 20)
11. Develop a Québec strategy on the use of big data from genomics research. (p. 20)

Scientific, Ethical and Legal Context

Accessing and sharing data: The cornerstones of genetic research

Rooted in a complex ecosystem of data collection, analysis, use and sharing, genomics research[2] promises to deliver significant breakthroughs in health. Given the intrinsic individual nature of human genome sequencing, this research paves the way for progress in what is referred to as "personalized medicine." Personalize medicine proposes to make use of the genetic, environmental or metabolic information of an individual by way of preventing, diagnosing or treating his or her diseases (Tremblay & Hamet, 2013). Personalized medicine research relies on more than just data generated by genome sequencing; it also entails the study of a patient's overall health, thus the need to link (or match) genomic data with rich phenotypic data[3], such as environmental data, information in medical records and administrative data (e.g., cancer registries, data on disbursements by the public health system [RAMQ], etc.).

Population-based studies in genomics also require such large datasets (if not larger). While some medical conditions are due to purely genetic mutations (e.g., certain rare diseases or certain single-gene diseases caused by a mutation in one gene, etc.), a significant number of non-communicable common diseases (e.g., Alzheimer, diabetes, cancer, heart disease, etc.) are believed to result from a combination of genetic, lifestyle and environmental factors. To study these factors and their interactions, population-based research requires the linkage of data from a variety of sources.

Moreover, data sharing is important for clinical purposes, since the genetic information collected in the context of research studies can, in certain cases, inform clinical decision-making[4]. For example, genetic information can be used to evaluate the statistical significance of a patient's genetic mutations or to link two patients with the same rare disease living in different countries[5].

[2] The terms "genetic" and "genomic" are often used interchangeably. However, "genomic" is the preferred term when referring to the study of the function and interaction of an individual's genes, whereas "genetic" primarily refers to the study of hereditary characteristics associated with specific genes.
[3] Phenotypic data include data on an individual's observable traits or characteristics, such as medical information (e.g., clinical signs, drug intake, laboratory results, etc.), demographic information (e.g., age, gender, weight, height) and other similar information.
[4] It should be noted that clinical sequencing is not always available in clinical care settings (except, for example, in cases involving the diagnosis of rare diseases, etc.). A large share of genome sequencing is conducted in the research context, although clinical applications may eventually result from such research.
[5] Matchmaker Exchange : http://www.matchmakerexchange.org/ (accessed on October 19, 2017).

Given the size of the human genome and related phenotypic data, the field of genomics is part of what is known as “big data” (i.e., datasets so large that they cannot be collected, analyzed, stored or managed using conventional methods) (Stephens et al., 2015). Genomic datasets are large, rich and generally costly to produce and store. Together, big data research, population-based studies and advances in personalized medicine are helping to stratify risk of diseases in the population, target the necessary interventions and ensure the allocation of appropriate resources .

Since a significant share of genomics research funding comes from the public sector, it is vital to ensure the optimal and sustainable use of these resources. Access and sharing of data for future use is a simple, efficient and cost-effective way to leverage existing genomic and health-related data. Indeed, the cost of analyzing genomic data that has already been collected is far lower than the cost of collecting new data or sequencing new biological samples (Kohane, 2011).

As is the case with many innovative research sectors, genomics research raises certain ethical and legal issues. To start, free and informed consent is the cornerstone of research with human subjects. The notion of consent is increasingly important when planning to use rich and potentially sensitive genetic and medical data for future research. Furthermore, issues related to the privacy and protection of personal information are important challenges in genomics due to the potential identifiability of certain types of genetic data, and of information generated by linking datasets from different sources.

These issues are generally reviewed by research ethics committees, whose mission is to oversee the ethical and legal aspects of research and ensure the protection of the rights and interests of participants. However, a significant share of administrative data is held by different public bodies and their access for research purposes is governed by differing legal frameworks and access mechanisms, thus creating confusion within the research community (The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, 2015).

Currently, the obstacles and challenges associated with accessing information held by public bodies hamper the activities of the research community in Québec (Quirion, 2016). On the one hand, to ensure the sustainable use of public funding dedicated to research, public bodies need to encourage and support a culture of sharing and openness. On the other, given the complexity of datasets required for genomics research, the policies, mechanisms and guidelines governing

access to information and data sharing must take into account the specific needs of the biomedical research community. Against this backdrop, support for innovation in genomics research in Québec requires deliberation by provincial policymakers in order to promote a safe and efficient research environment and facilitate participation in international research initiatives through data sharing.

This Policy Brief first sets out the ethical, legal and social issues related to the access and sharing of genomic and personal data for research purposes. Second, it presents the legal and policy barriers related to the exchange of information for research here in Québec and abroad. Third, it describes governance models pertaining to the access and sharing of data as examples of solutions intended to promote the development of a research sector respectful of the ethical issues and social considerations it raises. At the end of each section, recommendations are made to shed light on the discussion and suggest a few practical solutions that could support a more thorough examination of potential measures to be adopted.

Ethical and Social Issues related to Data Sharing in Research

As in many innovative research sectors, genomics raises a number of ethical, legal and organizational questions and concerns to consider. Consent, privacy and data sharing are among these issues.

Consent

A first issue related to the access and sharing of data for genomics research pertains to obtaining free and informed consent from participants. Consent is the cornerstone of participation in research and generally involves participants signing a consent form, as governed by research ethics standards[6]. In all cases, the content and format of the consent form must be approved by a competent research ethics committee whose mandate is to authorize or prohibit the conduct of research projects.[7]

In terms of interventional research, such as clinical trials, consent involves, amongst others, disclosing the risks related to the intervention (e.g., possible physical harms) and the potential benefits of participating in the research, to participants. Genomics research, however, entails

[6] In this respect, see Chapter 3 of the Tri-Council Policy Statement (TCPS 2) by the Social Sciences and Humanities Research Council of Canada, Natural Sciences and Engineering Research Council of Canada & Canadian Institutes of Health Research entitled: Ethical Conduct for Research Involving Humans, December 2014, 220 p.

[7] See section 24 of the Civil Code of Québec, CQLR c CCQ-1991 (the CCQ) and Chapter 3 of the TCPS 2, among others.

little physical risk, with the primary risk being informational in nature (i.e., related to the potential invasion of a person's privacy or breach of personal information). Consequently, the nature of consent differs slightly. For example, consent must include the purpose for which data will be used and the manner in which they will be protected. If research involves the linkage of data to other sources of information (e.g., medical records or administrative data), this should also be disclosed as part of the consent process. Likewise, participants need to be told if their data is to be subsequently shared or used for future research (e.g., creation of a database accessible to third parties, sharing of genetic data with international consortia) (Sénécal et al., 2016)[8]. In other words, consent mechanisms and communication with participants should provide information on the access and/or linkage of personal data, as well as details on any subsequent sharing of the data for genomics research (Gainotti et al., 2016). The explanation must be as precise as possible, even though broad consent may also be obtained, in certain circumstances[9]. Broad consent allows participants to explicitly grant researchers permission to use or share their data as part of future research projects whose specific objectives and conditions have yet to be determined (Sénécal et al., 2016). Moreover, broad consent, combined with the right of withdrawal and with ongoing consent (e.g., informing participants of research projects using the data collected), provides a way to meet the needs of sample and data collection for genomics research (Grady, et al., 2015). Following approval from a research ethics committee, this type of consent is generally adopted when developing biobanks or genetic databases which will involve data sharing or the subsequent secondary use of data. Historically, the ethical acceptability of broad consent has been debated at length, but an international consensus now authorizes its use when circumstances warrant, if approved by a competent research ethics committee (Council for International Organizations of Medical Sciences [CIOMS] in collaboration with the World Health Organization [WHO], 2016; International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use [ICH], 2017; World Medical Association [WMA], 2016).

In Québec, research ethics standards[10] and the *Act respecting access to documents held by public bodies and the protection of personal information* [11] provide for the sharing of data and their secondary use without prior consent under certain circumstances. Other than these

[8] See article 12.2 of the TCPS 2.
[9] See article 12.2 of the TCPS 2.
[10] See article 12.3A of the TCPS 2.
[11] Section 125 of the *Act respecting access to documents held by public bodies and the protection of personal information*, R.S.Q., Chapter A-2.1.

exceptions, the wishes of participants must always be respected when stipulated in a consent form approved by a competent research ethics committee. Under no circumstances should the standards governing research without consent trump duly obtained consent.

Confidentiality and Privacy Protection

Access and sharing of genetic and health-related data also pertains raise the issues of data confidentiality and the risk of invasion of privacy. The potential impact of a breach in privacy varies according to the identifiability, and sensitivity of the data collected. Related risks include the accidental disclosure of information, unauthorized access, inadvertent access and the re-identification of previously de-identified data (The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, 2015).

In this respect, it should be noted that the level of de-identification of personal and genetic information is set on a continuum (see Table 1). In other words, between completely anonymous data (non-identifying) on the one hand and directly identifiable information on the other, there are varying degrees of identifiability of personal information based on the nature of the data and the manner in which it is processed. In fact, many techniques used in information technology today can reduce the level of identifiability of data. For instance, it is possible to replace directly identifiable information (e.g., name, address, health insurance number) with a code, add a random “noise” to datasets, or use masking techniques, such as data swapping, encryption, tokenization, aggregation or other methods (El Emam, 2007; El Emam, 2011; Phillips & Knoppers, 2016). As a result, information that was initially identifiable to an individual can become de-identified.

Genomics research raises a specific concern in that a complete DNA sequence is generally unique to a single individual. Given the high cost of whole-genome sequencing, many research projects do not sequence the entire genome, but use gene panels (predetermined selection of genes) or exomes (limited collection of genes containing protein-coding sequences), thus raising fewer concerns regarding identification.

In the case of research using whole genomes, the risk of privacy breaches could theoretically be greater due to a hypothetical intentional re-identification of individuals, using their genetic sequence and associated data, in combination with identifying information found in third-party sources (e.g., by triangulation). Although this theoretical risk does pose a challenge, in practice, there are very few documented cases, in the research, sector involving the re-identification of individuals using genetic information (Laurie, Stevens, Jones, & Dobbs, 2015).

Lastly, by taking into account the nature of data, their degree of sensitivity, their level of identifiability and the manner in which they are processed, it is possible to develop data access and sharing systems that reflect the level of risk involved (Dyke, Dove, & Knoppers, 2016). Examples of innovative models adopting this approach will be presented in a subsequent section.

As previously mentioned, if research participants have been adequately informed of the risks and means to minimize such risks through a consent form duly approved by a competent research ethics committee, and if they have given their free and informed consent to the use and/or sharing of their genomic data for research purposes, the rules on their right to autonomy[12] dictate that their consent be respected.

National and International Data Sharing

A third challenge specific to the genomics research ecosystem involves the national and international sharing of data. As previously mentioned, in order to reach statistically significant results regarding the clinical impact of major genetic mutations, leading-edge genomics research depends on the analysis of data from a large number of participants (Kosseim et al., 2014). Many international consortia were created in response to this need (e.g., Canadian Partnership for Tomorrow Project [CPTP][13], International Cancer Genome Consortium [ICGC][14], European Genome Archive [EGA][15], The Cancer Genome Atlas [TCGA][16], etc.). These consortia bring together large volumes of genetic and health-related data from participants all over the world. However, interprovincial and international sharing of data is complex due to legal restrictions on the storage period and location of information, the differences in terminology used (Litton, 2017) and the mechanisms governing access to certain personal data (for example, information in medical records and administrative data, etc.). Moreover, some jurisdictions limit the use of cloud computing technologies, which are often required to transfer big data among researchers (Dove, Joly, Tassé, & Knoppers, 2015).

[12] Article 10 CCQ
[13] CPTP, https://www.partnershipfortomorrow.ca/ (accessed on October 19, 2017).
[14] International Cancer Genome Consortium (ICGC): http://icgc.org/ (accessed on October 19, 2017).
[15] European Genome Archive (EGA): https://www.ebi.ac.uk/ega/home (accessed on October 19, 2017).
[16] The Cancer Genome Atlas (TCGA): https://cancergenome.nih.gov/ (accessed on October 19, 2017).

Recommendation 1: Recognize the validity and primacy of broad consent when approved by a research ethics committee.

Recommendation 2: Eliminate the jurisdictional restrictions on the use of cloud computing technologies needed for the transfer of big data among national and international researchers.

Existing Ethical and Legal Barriers to Data Access and Sharing

Obstacles to national and international data sharing

As a general rule, the collection of biological samples for genetic sequencing is governed by standards and regulations that apply to human health research. However, access to other types of information used in research (e.g., data in medical records, administrative data) and the linkage of this information with sequencing data, are covered by access to information and privacy law regimes. Internationally, however, the legal standards that apply to the protection of personal information vary significantly both in terms of processes to request access to data, as well as potential use of data for research purposes.

Some jurisdictions, such as Australia, China and Germany (Information Technology and Innovation Foundation, 2017), have adopted restrictive approaches to the sharing of personal information outside of their jurisdictions, and provide few exceptions when it comes to data sharing for research purposes (Kosseim et al., 2014). There are also jurisdictional variations regarding data linkage processes, thus limiting the possibly of creating enriched datasets (The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, 2015). Indeed, the multitude of administrative approval processes in different jurisdictions often constitute an operational obstacle that makes it difficult to share personal data internationally.

Moreover, as illustrated in Table 1, there are significant terminological differences when referring to research data (Phillips & Knoppers, 2016). For example, the terms anonymous, anonymized, non-identifying, de-identified, coded and pseudonymized are often used in genomics, but their meaning is not equivalent from one jurisdiction to another. As a result, depending on the type of data (e.g., genome sequencing, clinical or medical data), it is difficult if the method of processing the data is sufficient to allow for the information to be shared for research purposes, across borders.

Table 1: Table of Concordance of Data Privacy and Security Terms and Spectrum of Identifiability (adapted from the Privacy and Security Policy of the Global Alliance for Genomics and Health) (Global Alliance for Genomics and Health, 2015).

1 (Most Identifiable)	2	3	4 (Least Identifiable)
"Is or can be fully identifiable to everyone"	"Is unidentifiable to most, but remains re-identifiable to those with access to the key(s)"	"Is likely no longer identifiable to anyone"	"Never was identifiable"
• Identifiable • Personal • Nominative	• Coded • Pseudonymized • Reversibly de-identified • Masked • Encrypted • Linked anonymized	• Anonymized • De-identified • Irreversibly de-identified • Non-identifiable • Unlinked anonymized	• Anonymous

Yet, differences in terms of accessing and sharing data globally are not limited to variations in terminology or jurisdictional restrictions on the sharing of personal information. At the interprovincial and international level, another significant obstacle involves the diverging and inconsistent review criteria used by research ethics committees (Dove et al., 2016). Due to applicable regulatory frameworks, research ethics committees are often required to review and approve research projects that have previously been assessed and authorized in other jurisdictions. At times, this results in contradictory decisions which place restrictions on the sharing of genomic data for research purposes (Townend, Dove, Nicol, Bovenberg, & Knoppers, 2016). In response, several authors have suggested the implementation of a framework for the mutual recognition of international ethics reviews, under certain conditions (Dove et al., 2016; Townend et al., 2016).

Overview of Québec-specific obstacles

In addition to the national and international barriers previously discussed, researchers in Québec face a number of other challenges regarding access to information held by government bodies, and linkage of this information with genomic data.

As stated by the President of the Commission d'accès à l'information (CAI) in his introductory remarks to the CAI's 2016 five-year report (Rapport quinquennal 2016: Rétablir l'équilibre): « […] while the 1982 *Act respecting access to documents held by public bodies and the*

protection of personal information was one of the most progressive at the time of its enactment, today many experts agree that it ranks at the lower end of the scale." (Translation) (Commission d'accès à l'information du Québec, 2016).

Although Québec is often described as a technological hub in terms of big data, artificial intelligence, genomics research and pharmaceutical development (Gouvernement du Québec, 2017), unfortunately, certain regulatory and institutional obstacles threaten the growth of this innovative ecosystem.

The creation of rich datasets useful for genomics research generally requires several levels of institutional and administrative approval. This process is considered to be too slow (Montréal InVivo, 2017). For instance, a researcher wishing to conduct research involving genome sequencing of cancer patients, and to link these data with clinical and administrative information must minimally obtain:

- **Scientific approval** of his project by a committee of his peers (review of the project's scientific feasibility and relevance as determined by the funding agency's mandate);
- **Ethics approval** of his project by a competent research ethics committee;
- **Feasibility approval** of the project when conducted in part or in whole in a Québec health and social services institution (review of its feasibility and relevance for each institution involved); and, in certain cases, **approval by an access committee or governmental body responsible for access to information** (e.g., CAI) when part of the information needed for the research are held in third-party databases. (Public Population Project in Genomics and Society [P^3G] & Centre of Genomics and Policy [CGP])

Unfortunately, these committees and organizations do not all adopt the same review criteria nor do they apply the same laws and policies. Moreover, the data needed for genomics research are often held in administrative silos, which requires several levels of approval, leading to increasing delays and costs (Quirion, 2016).

Although a significant share of genomics research requires only de-identified data (e.g., coded), the Québec legislation governing access to information "does not contain exceptions which would allow requests to access personal information to be scaled based on the requested information's sensitivity or on the person's reasonable expectation of privacy […]" (Translation) (Commission d'accès à l'information du Québec, 2016). Given the reality of genomics research,

the data protection techniques available and the low level of risk of privacy breaches, it would be appropriate to adjust the access request review process based on the level of identifiability and sensitivity of the data requested, thus complying with the principle of proportionality (Public Population Project in Genomics and Society (P^3G) & Centre of Genomics and Policy [CGP]).

In addition, even in cases where participant consent has been granted, the current wording and interpretation of the *Act respecting access to documents held by public bodies and the protection of personal information* make it difficult to access medico-administrative data held by government bodies, for research purposes, and to link this data with genetic data.

In this respect, the CAI stipulates that for consent to be valid, it must be explicit, free, informed, specific and time limited (Commission d'accès à l'information du Québec, 2016). However, the CAI does not recognize broad consent as valid even when approved by a competent research ethics committee (Public Population Project in Genomics and Society (P^3G) & Centre of Genomics and Policy [CGP], 2017). Failure to recognize broad consent can have a drastic impact on the operation and sustainability of biobanks and genomic databases. For instance, as part of Québec biobanking project CARTaGENE, even though participants had given their free and informed consent to the regular access of their medico-administrative data held by the CAI and the Québec public health system (Régie de l'assurance maladie du Québec), the CAI still reviews each access request, in accordance with the provisions of the *Act respecting access to documents held by public bodies and the protection of personal information*, applicable to access to information without the consent of the individuals involved (CARTaGENE [CHU Sainte-Justine], 2017). This results in significant delays, reaching more than one year per access request (CARTaGENE [CHU Sainte-Justine], 2017).

The obstacles mentioned previously may be due to the lack, in Québec, of an organization whose specific mission it is to promote research which uses data held by public bodies. Given the wording of the *Act respecting access to documents held by public bodies and the protection of personal information*, the CAI's mission is to protect personal information held in the private and public sectors, but not to optimize use of this information for research purposes. In order to re-establish an equilibrium, it could be interesting to consider the creation of a neutral organization, having the resources and technical expertise needed to protect, process and manage data (Quirion, 2016), and whose mission would be not only to review access to information requests, but also to facilitate and optimize the access and use of public data, for research purposes (Montréal InVivo, 2017). This would promote the development of research in

Québec, foster a culture of data sharing for innovative research purposes and, more generally, support the creation of rich and secure databases. More broadly, this one-stop access point could also drive pan-Canadian efforts related to administrative data linkage (Doiron, Raina, & Fortier, 2013).

Recommendation 3: In keeping with the principle of proportionality, when updating the *Act respecting access to documents held by public bodies and the protection of personal information*, scale the access to information review process according to the level of risk and anticipated social benefits.

Recommendation 4: In the context of genomics research, ensure that the principle of general or broad consent is recognized by both research ethics committees and by administrative bodies and committees in charge of access to information held by government agencies. This could be accomplished by revising the relevant provisions of the *Act respecting access to documents held by public bodies and the protection of personal information*.

Recommendation 5: With respect to access to information held by public bodies, standards governing research without consent should not override duly obtained consent.

Recommendation 6: Eliminate restrictions on the storage mode and retention period of information obtained from public bodies.

Recommendation 7: Develop mechanisms to streamline access to de-identified information held by public bodies.

Recommendation 8: To foster a more efficient research ecosystem in Québec, set up an organization ("one-stop access point") whose mission is to provide access to health-related data for research purposes.

Possible Solutions: Promote Data Sharing and Access

Innovative mechanisms for accessing data held by public bodies

In Canada and around the world, researchers use a variety of systems to access medico-administrative data, which are then linked with genomic information in order to create rich datasets which are useful for research. The role of these systems is to protect the confidential

and sensitive nature of personal data and simultaneously stimulate the development of clinical applications stemming from the research being conducted using such data. P^3G was recently tasked by the Fonds de recherche du Québec to carry out an international comparative analysis on the standards governing access to personal information held by government bodies, including the implementation of regional or provincial models of direct access to data (e.g., one-stop access point). Some findings from the study were also summarized in a brief presented by P^3G and the CGP to the Commission des institutions on the 2016 five-year plan of the Commission d'accès à l'information du Québec (Public Population Project in Genomics and Society [P^3G] & Centre of Genomics and Policy [CGP], 2017). The creation of a centralized access process, or one-stop access point, is the preferred course of action in many jurisdictions as a way to facilitate access to information held by public bodies. In fact, many jurisdictions have opted to implement this type of streamlined access process, which can involve direct, remote network or access via satellite sites.

With the **direct access** model, researchers are required to visit, in person, the access centre and consult the data directly on site. In some cases, they can also access information via satellite locations or institutions affiliated with the access centres. This model is a preferred option for providing access to sensitive information, such as health or financial data (Organisation for Economic Co-operation and Development [OECD], 2015).

With **remote access**, users can access and analyze data held in an access centre, from their own location, through a highly secure computer network connected to the access centre's network. This access model generates significant savings since it does not require dedicated workspaces and provides researchers with direct and virtual access to data. One example of the remote access model is the SD Box, developed in France, which addresses the reluctance of some researchers to install the software and card readers traditionally required to access data (Gadouche, 2013).

Access via **satellite sites** involves the set-up of secure locations where data can be accessed, while protecting their confidentiality. In Manitoba, for example, a library can constitute an appropriate access location (Manitoba Centre for Health Policy, n.d.). Statistics Netherlands, however, requires that the room used by researchers to access administrative data be locked (Organisation for Economic Co-operation and Development [OECD], 2015). Satellite sites are

also generally required to have a private virtual network with a firewall.[17] Lastly, to keep data secure and confidential, access via satellite sites typically impose obligations upon users, including the signing of data transfer and confidentiality agreements.

Furthermore, researchers who breach their obligations may be subject to major sanctions. For instance, on July 6, 2016, the Department of Health of the United Kingdom implemented stricter criminal sanctions for individuals who use de-identified data in a manner that permits the identification of individuals involved (Department of Health and George Freeman MP, 2016). Stricter penalties for violating data protection laws have also been introduced in Germany. (Rehder & Paez, 2010). Moreover, when accessing data from Statistics Denmark, imprisonment can be imposed if privacy protection regulations are breached (Thaulow, 2015).

[17]For example: Secure Anonymised Information Linkage (SAIL), online: http://www.saildatabank.com (accessed on September 27, 2017); Population Data BC, online: https://www.popdata.bc.ca/dataaccess (accessed on September 27, 2017).

Table 2. Examples of Jurisdictions Providing Centralized Access to Data Held by Government Bodies

Name	Jurisdiction
Institute for Clinical Evaluative Sciences (ICES)	Ontario, Canada
Population Data BC (PopData BC)	British Columbia, Canada
Manitoba Centre for Health Policy (MCHP)	Manitoba, Canada
Centre for Data Linkage (CDL)	Australia
Statistics Netherlands	Netherlands
Statistics Denmark	Denmark
Centre d'accès sécurisé distant aux données (CASD)	France
Scottish Informatics and Linkage Collaboration (SILC)	Scotland, United Kingdom
Secure Anonymised Information Linkage (SAIL)	Wales, United Kingdom

In Canada, the Ontario-based Institute for Clinical Evaluative Sciences (ICES)[18], a non-profit organization tasked with providing secure access to an array of Ontarians' health-related data, is considered a particularly interesting model. ICES provides approved researchers with remote access to certain types of data. It has been reported that many researchers from Québec have been using ICES as a way to access health-related data, as it is perceived to be more reliable than Québec's access to information channels (Montréal InVivo, 2017).

Genomic data-sharing mechanisms among researchers

While some research projects involve the manual - or piecemeal - linkage of genomic data, other projects propose to create genomic databases or biobanks, where genomic data are systematically linked with other health-related information. These databases are primarily developed to serve as resources for research. Generally, the resource itself directly oversees the sharing of data. As a result, this model substantially reduces the costs associated with conducting exploratory research and prevents the need to generate new sequencing data. However, the value of these initiatives greatly depends on the capacity to transfer data between jurisdictions, since, on the one hand, the size of the data typically requires the use of cloud computing technologies (Litton, 2017) and, on the other, the statistical power needed to analyze certain research questions requires a number of cases that often exceeds the amount of data available in a single country (Mascalzoni et al., 2015).

[18] Institute for Clinical Evaluative Sciences (ICES) : https://www.ices.on.ca/ (Accessed on September 28, 2017).

Several solutions have been adopted to address issues associated with the sharing of genomic data through databases. For instance, multilateral contracts, such as Consortium Agreements (which govern the creation, access modes and transfer of data) or Data Transfer Agreements (which stipulate the rights and obligations of researchers using the data), are legally binding tools, which dictate, for example, authorized uses for the data (Kosseim et al., 2014), applicable data protection standards, sanctions in the event of breaches and other mechanisms aiming to protect the shared genetic and clinical data (Knoppers et al., 2013). Furthermore, in order to meet ethical and legal requirements, these databases implement various governance mechanisms that aim to ensure a responsible sharing of data among genomics researchers (Kosseim et al., 2014).

In the context of biobanking and genomic databases, international guidelines have generally established the ethical acceptability of broad consent, based on the existence and implementation of appropriate governance mechanisms. For example, the World Medical Association stipulates that governance arrangements for health databases and biobanks should include the following elements pertaining to data sharing:

- "Arrangement for how the data and material will be documented and traceable in accordance with the consent of the concerned persons; […]
- Arrangement for obtaining appropriate consent or other legal basis for data or material collection; […]
- Criteria and procedures concerning the access to and the sharing of the health data or biological material including the systematic use of Material Transfer Agreement (MTA) when necessary; […]
- The security measures to prevent unauthorized access or inappropriate sharing […]." (World Medical Association [WMA], 2016).

In practice, data access committees are typically tasked with ensuring the oversight of ethical and regulatory aspects of data access and, more specifically, of ensuring compliance with the governance mechanisms of the resource (Shabani & Borry, 2016). In addition, to facilitate the sharing of less sensitive types of genomic data, some access models, including the "Registered Access" model, propose an approach modulated in accordance with the level of identifiability and sensitivity of the data and the qualifications of the researcher requesting access (Dyke, et al., 2016). Finally, to ensure an efficient, transparent and systematic sharing of data, it is

important to support centralized access models, which involve the review of access requests by a common entity.

In summary, the Québec innovation ecosystem in genetics and genomics depends on the seamless interaction among institutions, their research ethics committees, custodians of data held by public bodies, biobanks, research databases and researchers. This collaboration can be achieved through the use of tools, policies, regulations and technological or governance mechanisms that propose a balance, on the one hand, between protecting the privacy of participants who have consented to contribute their genetic information to research and, on the other hand, fostering innovative research.

Recommendation 9: When implementing a one-stop access point for the access of data held by public bodies, draw on innovative information technologies approaches adopted by other jurisdictions for the sharing of sensitive information.

Recommendation 10: Launch a discussion forum among Québec research ethics committees to promote a harmonized approach to the ethical review of multi-site and international research involving big data in genomics.

Recommendation 11: Develop a Québec strategy on the use and sharing of big data in genomics as part of international, collaborative research initiatives.

References

CARTaGENE (CHU Sainte-Justine). (2017) Mémoire sur le Plan quinquennal 2016 de la Commission d'accès à l'information. Retrieved from: https://www.cartagene.qc.ca/sites/default/files/documents/other/Memoire_CARTaGENE_Mai2017.pdf (Accessed on October 19, 2017).

Commission d'accès à l'information du Québec. (2016) Rétablir l'équilibre : Rapport sur l'application de la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels et de la Loi sur la protection des renseignements personnels dans le secteur privé*.

Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO). (2016) International Ethical Guidelines for Health-related Research Involving Humans. Geneva. Retrieved from: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf (Accessed on October 19, 2017).

Department of Health and George Freeman MP. (2016) The Care Quality Commission and National Data Guardian's recommendations for data security and a consent/opt-out model for data sharing. Retrieved from: https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent (Accessed on October 19, 2017).

Doiron, D., Raina, P., & Fortier, I. (2013). Linking Canadian Population Health Data: Maximizing the Potential of Cohort and Administrative Data. *Canadian Journal of Public Health = Revue Canadienne de Sante Publique*, *104*(3), e258–e261. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3880355/

Dove, E. S., Joly, Y., Tasse, A.-M., & Knoppers, B. M. (2015). Genomic cloud computing: legal and ethical points to consider. *European Journal of Human Genetics : EJHG*, *23* (10), 1271–1278. doi:10.1038/ejhg.2014.196

Dove, E. S., Townend, D., Meslin, E. M., Bobrow, M., Littler, K., Nicol, D., … Knoppers, B. M. (2016). Ethics review for international data-intensive research. *Science*, *351* (6280), 1399 LP – 1400. http://science.sciencemag.org/content/351/6280/1399.abstract

Dyke, S. O. M., Dove, E. S., & Knoppers, B. M. (2016). Sharing health-related data: a privacy test? *NPJ Genomic Medicine*, *1* (1), 16024–16026. doi:10.1038/npjgenmed.2016.24

Dyke, S. O. M., Kirby, E., Shabani, M., Thorogood, A., Kato, K., & Knoppers, B. M. (2016). Registered access: a "Triple-A" approach. *European Journal of Human Genetics*, *24* (12), 1676–1680. doi:10.1038/ejhg.2016.115

El Emam, K. (2007). Pan-Canadian de-identification Guidelines for Personal Health Information. Ottawa, Ontario: CHEO Research Institute.

El Emam, K. (2011). Methods for the de-identification of electronic health records for genomic research. *Genome Medicine*, *3* (4), 25. doi:10.1186/gm239

Gadouche, K. (2013). Le centre d'accès sécurisé aux données (CASD). Orsay. Retrieved from: http://www.centre-dalembert.u-psud.fr/wp-content/uploads/2013/03/K.Gadouche.pdf (Accessed on October 19, 2017).

Gainotti, S., Turner, C., Woods, S., Kole, A., McCormack, P., Lochmüller, H., … Mascalzoni, D. (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. doi:10.1038/ejhg.2016.2

Global Alliance for Genomics and Health. (2015) Privacy and Security Policy. Retrieved from: https://www.ga4gh.org/docs/ga4ghtoolkit/data-security/Privacy-and-Security-Policy.pdf (Accessed on October 19, 2017).

Gouvernement du Québec. (2017) Stratégie québécoise des sciences de la vie 2017-2027. Retrieved from: https://www.economie.gouv.qc.ca/fileadmin/contenu/publications/administratives/strategies/sciences_vie/strategie_sciences_vie.pdf (Accessed on October 19, 2017).

Grady, C., Eckstein, L., Berkman, B., Brock, D., Cook-Deegan, R., Fullerton, S. M., … Wendler, D. (2015). Broad Consent for Research With Biological Samples: Workshop Conclusions. *The American Journal of Bioethics*, *15* (9), 34–42. doi:10.1080/15265161.2015.1062162

Information Technology and Innovation Foundation. (2017) Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost? Retrieved from: https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost (Accessed on October 19, 2017).

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH). (2017) Guidelines on Genomic Sampling and Management of Genomic Data (E18) DRAFT (Step 3). Retrieved from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E18/E18EWG_Step4_Guideline_2017_0803.pdf (Accessed on October 19, 2017).

Knoppers, B., Chisholm, R., Kaye, J., Cox, D., Thorogood, A., Burton, P., … P Stolk, R. (2013). A P3G Generic Access Agreement for Population Genomic Studies. *Nature biotechnology* (Vol. 31). doi:10.1038/nbt.2567

Kohane, I. S. (2011). Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*, *12* (6), 417–428. http://dx.doi.org/10.1038/nrg2999

Kosseim, P., Dove, E. S., Baggaley, C., Meslin, E. M., Cate, F. H., Kaye, J., … Knoppers, B. M. (2014). Building a data sharing model for global genomic research. *Genome Biology*, *15* (8), 430. doi:10.1186/s13059-014-0430-2

Laurie, G., Stevens, L., Jones, K. H., & Dobbs, C. (2015). A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data. Nuffield Council on Bioethics.

Litton, J.-E. (2017). We must urgently clarify data-sharing rules. *Nature*, *541* (7638), 437. doi:10.1038/541437a

Manitoba Centre for Health Policy. (n.d.). Accreditation 2016. Retrieved from: http://umanitoba.ca/faculties/health_sciences/medicine/units/community_health_sciences/departmental_units/mchp/education/media/4_privacy_2016.pdf (Accessed on October 19, 2017).

Mascalzoni, D., Dove, E. S., Rubinstein, Y., Dawkins, H. J. S., Kole, A., McCormack, P., … Hansson, M. (2015). International Charter of principles for sharing bio-specimens and data.

Eur J Hum Genet, *23* (6), 721–728. Recovered from: http://dx.doi.org/10.1038/ejhg.2014.197

Montréal InVivo. (2017) Les sciences de la vie et technologies de la santé (SVTS) et l'accès aux données de santé : Propositions quant au rapport sur l'application de la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* et de la *Loi sur la protection des renseignements personnels dans le secteur privé.*

Organisation for Economic Co-operation and Development (OECD). (2015) *Health Policy Study* - Health Data Governance: Privacy Monitoring and Research. Paris. Retrieved from: http://www.oecd.org/publications/health-data-governance-9789264244566-en.htm (Accessed on October 19, 2017).

Phillips, M., & Knoppers, B. M. (2016). The discombobulation of de-identification. *Nat Biotech*, *34*(11), 1102–1103. http://dx.doi.org/10.1038/nbt.3696

Public Population Project in Genomics and Society (P^3G) & Centre of Genomics and Policy (CGP) (2017) Rétablir l'équilibre entre protection et développement des connaissances : Pour une approche contemporaine de l'utilisation des renseignements personnels à des fins de recherche biomédicale - Mémoire sur le Plan quinquennal 2016 de la Commission d'accès à l'information. Retrieved from: http://www.assnat.qc.ca/Media/Process.aspx?MediaId=ANQ.Vigie.Bll.DocumentGenerique_131361&process=Default&token=ZyMoxNwUn8ikQ+TRKYwPCjWrKwg+vIv9rjij7p3xLGTZDmLVSmJLoqe/vG7/YWzz (Accessed on October 19, 2017).

Quirion, R. (2016). Pour un meilleur accès aux données. Montréal : Colloque Faciliter l'accès aux données du Québec : Comment et à quelles fins? Retrieved from: https://www.ciqss.org/sites/default/files/documents/2016-11-30_Remi-Quirion.pdf (Accessed on October 19, 2017).

Rehder, J., & Paez, M. F. (2010). Germany strengthens its Data Protection Act and introduces data breach notification requirement. *World Data Report*, *16* (1).

Sénécal, K., Thorogood, A., Parry, D., Joly, Y., & Knoppers, B. M. (2016). The Québec Network of Applied Genetic Medicine (RMGA) Consolidated Statement of Principles Retrieved from: https://www.rmga.qc.ca/en/documents/Enonceconsolide_ENprint_001.pdf (Accessed on October 19, 2017).

Shabani, M., & Borry, P. (2016, September). You want the right amount of oversight: interviews with data access committee members and experts on genomic data access. *Genet Med*. American College of Medical Genetics and Genomics. http://dx.doi.org/10.1038/gim.2015.189

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., … Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLOS Biology*, *13* (7), 1–11. doi:10.1371/journal.pbio.1002195

Thaulow, I. (2015). Microdata for Research Purposes at Statistics Denmark. Copenhague. Retrieved from: https://www.innokyla.fi/documents/1598478/86804e78-8d35-4b55-9f63-296dd7d1d7b2 (Accessed on October 19, 2017).

The Expert Panel on Timely Access to Health and Social Data for Health Research and Health

System Innovation. (2015) Accessing Health and Health-Related Data in Canada. Retrieved from: http://www.scienceadvice.ca/uploads/eng/assessments and publications and news releases/Health-data/HealthDataFullReportEn.pdf (Accessed on October 19, 2017).

Townend, D., Dove, E. S., Nicol, D., Bovenberg, J., & Knoppers, B. M. (2016). Streamlining ethical review of data intensive research. *BMJ*, *354*. http://www.bmj.com/content/354/bmj.i4181.abstract

Tremblay, J., & Hamet, P. (2013). Role of genomics on the path to personalized medicine. *Metabolism*, *62* (Supplement 1), S2–S5. doi:https://doi.org/10.1016/j.metabol.2012.08.023

World Medical Association (WMA). (2016) WMA Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. Retrieved from: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/ (Accessed on October 19, 2017).