Policy Brief: Protecting Privacy in Cloud-Based Genomic Research

Version 1.0 July 21st, 2015

Suggested Citation: Adrian Thorogood, Howard Simkevitz, Mark Phillips, Edward S Dove & Yann Joly, "Policy Brief: Protecting Privacy in Cloud-Based Genomic Research" (21 July 2015).

Authors: Adrian Thorogood[*], Howard Simkevitz[†], Mark Phillips[‡], Edward S. Dove[§] & Yann Joly[**]

This Policy Brief recommends ways to protect against privacy risks related to the storage and analysis of genomic and associated health data in the cloud. It was prepared as part of the Office of the Privacy Commissioner Contributions Program 2014-2015. The Policy Brief is informed by comparative analysis of Canadian, US and European privacy/data protection laws, and a case study of the Terms of Service of six cloud service providers. Part I introduces the sensitive nature of genomic information and the challenges to the privacy of participants associated with cloud-based genomic research. Part II analyzes the legal and ethical obligations of researchers using the cloud. Part III identifies gaps in the privacy protections offered by cloud providers, in their Terms of Service, and recommends how they can be improved. A *Cloud Contract Checklist* (see Annex 1) lists important clauses for researchers to consider in their contracts with cloud providers. Part IV identifies legal and regulatory barriers to the uptake of cloud services and gaps in privacy protection.

I. Context

Genomes and Privacy: Genomic research involves comparisons of individuals' genomes and associated health data across large groups to determine how genetic, medical, and environmental factors interact to cause disease. The genome is the entirety of an individual's genetic information, which is largely inherited from biologically related kin, especially one's parents. Genomic data are sensitive as they contain information about one's health and various personal

[*] Academic Associate, Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Canada.
[†] General Counsel and Privacy Officer, Ontario Institute for Cancer Research, MaRS Centre Toronto, Canada.
[‡] Research Assistant, Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Canada.
[§] PhD Candidate, School of Law, University of Edinburgh, United Kingdom.
[**] Associate Professor, Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Canada.

traits. Unauthorized access to genomic and associated health data can lead to individual and societal harms. These harms include discrimination by third parties such as employers and insurers, harm to dignity, psychological damage, and a loss of public confidence in health research and health care.

The Changing Face of Genomic Research: Globalization and increased connectivity allow researchers to collaborate and share research data around the world. Advances in sequencing technologies have made deciphering an individual's genome more rapid and affordable. Enhanced computing power and new bioinformatics techniques allow researchers to assemble individual genomes and compare genomic data across many individuals in order to identify variations that cause disease. Researchers are, therefore, collecting increasing amounts of genomic and associated health data about participants or accessing these data from existing repositories. They are also storing data for longer periods of time (if not indefinitely), exploiting the same data to answer multiple and varied research questions. Accelerating discovery and improvements in research and health care will require a powerful computing infrastructure capable of storing and analyzing massive research data-sets, and sharing them around the world.

Cloud-Based Genomics: Many gigabytes of genomic data may be associated with a single participant; projects involving tens of thousands of participants can quickly grow to several petabytes in size. Significant computing and network resources are required to store, analyze and share these massive datasets. Cloud computing provides a cost-effective means for the health research community to meet these challenges.

Cloud computing is a complex combination of computing platforms, often integrated with business services, which allows for scalable, on-demand, and remote access to computing

resources over a network. A cloud may be deployed by a private organization or community (such as a university), and can be made accessible on an internal network. Commercial companies also offer cloud services to the public, allowing customers to store and analyze data in the cloud remotely over the internet. The cloud ecosystem is varied and complex. Cloud companies, or layered composites of companies, may, for instance, provide access to raw computing hardware, to a platform on which researchers can develop software applications, or to pre-packaged software applications that researchers can run on the cloud.

Privacy Challenges in the Cloud: Cloud-based genomic research presents some risk to privacy. Wide sharing by researchers through the cloud raises the risk of unauthorized access to and use of sensitive genomic and associated health data about participants and their relatives. Participants' data transferred out of Canada may be subject to compelled disclosure under foreign laws, and privacy rights granted under Canadian law may be difficult to enforce. In addition, the complex storage and networking architecture of cloud computing creates new and potentially unknown cyber security risks, especially where cloud users are inexperienced. Researchers using the cloud may have to give up some control over their research data and how they are protected. This raises concerns of accountability.

II. Legal and Ethical Regulation of Cloud-based Genomic Research

Privacy laws in Canada generally allow researchers to store and analyze genomic and health-related data in the cloud, with some conditions and exceptions. However, certain approaches to enhancing privacy, such as de-identification techniques, may not be sufficient to guarantee the anonymity of genomic data. Genomic data is inherently identifiable and should be treated as "personal health information". Personal health information is protected by federal and provincial

privacy laws, as well as the ethical requirements of the *Tri-Council Policy Statement*, which must be fulfilled as a condition of ethics approval for most research projects and federal funding.

Researchers Remain Accountable for Information Transferred to 3rd Parties: The legal principle of accountability generally requires researchers, as custodians of personal health information, to fulfill the following legal obligations, even after transfer to a cloud provider:

1. Limit use and disclosure to those authorized by the participant (or by a research ethics board);
2. Take reasonable steps to secure personal health information against error, loss, theft, and unauthorized use or disclosure;
3. Designate a contact person to respond to privacy related inquiries and complaints;
4. Be transparent about information handling and security practices; and
5. Notify privacy authorities, and potentially participants, of privacy and/or security breaches.

At a minimum, researchers must use contracts to ensure cloud providers only use participants' data for specified purposes, and employ robust privacy and security safeguards.

Conditions for Cross-border Transfer: Participant data transferred to the cloud or between distributed data centres within the cloud may be moved into foreign jurisdictions. This is permitted by Canadian law with the explicit consent of the participant. Without explicit consent, researchers are generally only permitted to transfer information to a foreign cloud provider if they notify participants of where the data are being transferred, and of the risk of compelled disclosure under foreign law. Compliance in the cloud context is suspect when the cloud provider does not even inform the researcher where data may be transferred. In some provinces,

however, public sector researchers are prohibited from transferring research data to a foreign cloud provider without explicit consent.[††]

Ethical Obligations: The *Tri-Council Policy Statement* requires full disclosure of privacy risks to participants, including information on who will have access to data; how confidentiality will be protected; who may have a duty to disclose data, and to whom such disclosures could be made. It follows that researchers should be transparent with participants about the limits of de-identification, and about the risks associated with cloud storage. If some of this information is initially unknown, consent processes may need to be supplemented with ongoing ethical oversight, public engagement, and a means for participants to withdraw.

III. Using Contracts to Protect Privacy in the Cloud

In order to offer affordable, on-demand computing services to many customers, most commercial cloud providers offer standard Terms of Service. Our review of the terms offered by six cloud providers revealed a number of issues with regard to transparency and security. For example, terms are typically non-negotiable and are subject to unilateral modification by the cloud provider; the location of data storage, security measures and subcontracting practices are not clearly specified; cloud providers tend to waive all liability relating to customer data; and breach notification may be discretionary or completely absent. In light of this, researchers should carefully assess if the cloud is suitable for a particular project, and compare the benefits and risks of different cloud services and providers. In order to facilitate compliance with Canadian laws and ethical standards, cloud providers should be contractually bound to:

[††] Explicit consent or authorization by statute or a government minister. See e.g., British Columbia *Freedom of Information and Protection of Privacy Act* [RSBC 1996] Chapter 165, s. 33.1; Nova Scotia *Personal Information International Disclosure Protection Act* Chapter 3 of the Acts of 2006, s. 9.

1. Employ detailed privacy and security measures, including comprehensive technical, administrative, and physical safeguards meeting industry standards and subject to periodic audits by an independent IT security expert.
2. Maintain monitoring systems such that it will be able to demonstrate compliance with policies, procedures, controls, and recognized best practices. Ensure all individuals involved in the provision of services (including subcontractors) are bound by enforceable nondisclosure or confidentiality agreements during and after the service relationship.
3. Employ a protocol to respond to any privacy and/or security breaches, which includes notifying the researcher.
4. Establish policies and procedures to ensure the integrity of research data and availability of service during the full life-cycle of storage, analysis and access (e.g., backup procedures, disaster recovery/business continuity plans, return of data upon termination).
5. Provide adequate notice periods whenever Terms of Service change, in case researchers need to migrate their data, particularly changes to the location of data storage, security and privacy safeguards, or suspension of service.
6. Assume liability for damages resulting from the cloud provider's own negligence or misconduct, including privacy and/or security breaches and data misuse.

IV. Improving Regulation of Cloud-Based Genomics in Canada

Regulators and policy makers in Canada can also improve privacy protections, and promote the responsible use of cloud-based genomics, by addressing the following issues:

1. Definitions of legal terms such as "identifiable", "coded", "anonymized", "de-identified", "personal health information", "consent", "research", and "custodian" differ between

provinces. Similarly, obligations imposed on researchers, e.g. to notify privacy authorities of a data breach, or to impose specific contractual terms when transferring personal health information to a cloud provider, may also vary. Legal discrepancies across provinces and countries are a barrier to collaborative research in the cloud. Canadian privacy authorities should cooperate to help develop uniform guidelines for privacy compliant cloud computing applicable across provinces and sectors, and internationally.

2. Ethics boards charged with approving and monitoring research may lack the expertise and resources to oversee data-intensive science involving international collaborations and technologies like cloud computing. Separate expert boards should be created to review privacy and security issues, or at a minimum such expertise should be present on ethics boards.
3. Participation in research, particularly international cloud-based research, is discouraged by a perceived lack of transparency and control over government surveillance, both within Canada and abroad. Governments should clarify the purpose and extent of health information surveillance so Canadians can make informed decisions about research participation.

Primary References

1. E.S. Dove et al., "Genomic Cloud Computing: Legal and Ethical Points to Consider" *European Journal of Human Genetics*, 2014.

2. E.S. Dove, Y. Joly, B.M. Knoppers, "International Genomic Cloud Computing: 'Mining' the Terms of Service", in A.S.Y. Cheung, R. Weber, eds. *Privacy and Legal Issues in Cloud Computing*. Cheltenham: Edward Elgar, 2015: 237-260.

3. M. Naveed et al., "Privacy and Security in the Genomic Era", arXiv.org (2014).

4. "Accessing Health and Social 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, 2015.

5. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (2014).

6. M. Power, *The Law of Privacy*, (Ontario: LexisNexis: 2013).

7. C. Kuner, *Transborder Data Flow Regulation and Data Privacy Law* (Oxford: Oxford University Press, 2013).

8. W. Jansen, T. Grance, "Guidelines on Security and Privacy in Public Cloud Computing" National Institute of Standards and Technology, Special Publication 800-144 (2001).

9. D. Krebs, "Regulating the Cloud: A Comparative Analysis of the Current and Proposed Privacy Frameworks in Canada and the European Union" (2012) 10 *Canadian Journal of Law and Technology* 29.

10. G. Gunasekara, "Paddling in unison or just paddling? International trends in reforming information privacy law" *Int J Law Info Tech* (Summer 2014) 22 (2).

11. NIH, "NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy Security Best Practices" 2015, available at <http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf>.

12. K. El Emam, *Guide to the De-identification of Personal Health Information*, (Taylor & Francis: New York, 2013)

13. Office of the Privacy Commissioner of Canada, "Reaching for the Cloud(s): Privacy Issues Related to Cloud Computing" (2010).