MOHCCN Data Privacy Policy (17 November 2023)

Introduction

The Network recognizes that protecting the privacy and confidentiality of research participants and patients is an essential component of responsible data sharing and commit to respecting applicable personal (health) information protection laws and research ethics guidelines, as well as the expectations of research participants.

The federated approach of the Network places ultimate responsibility for privacy protection with individual Network Institutions (those Institutions made a party to the Network Master Agreement) together with researcher(s) who contribute data to the common Network resource. When conducting data collection and database population, they are responsible to obtain consent and to ensure data are coded such that all direct identifiers (such as name, DOB, MRN, SIN) are removed and replaced with a randomly generated number ("Coded Data") before being made available to the Network. Network committees and TFRI provide support in the form of policies, standards, and coordinated Data Access Committee (DAC) review processes.

This policy is developed and maintained by the Data Policy and Standards Committee with appropriate input from the regional Consortia and interdisciplinary perspectives (scientific, technical, legal, ethics).

Privacy Safeguards for Controlled Access Data

All individual-level Network data as described in **Table 1** will only be made available subject to participant consent through a controlled-access tier.[1] The following privacy safeguards apply to controlled-access data:

To obtain controlled-access data, a researcher must submit a Data Access Request to the Network Data Access Committee (DAC). This applies both to Network as well as (in the future) external researchers. (*Note: the Phase I Network Master Agreement only covers access by Investigators based at Network Institutions.)*

Note that researchers from each group will only be able to request access to data after the expiry (or waiver) of the respective embargo period. The Network has approved **embargo periods** giving privileged access to data-contributing Teams, and to the Network (6+ months after sequencing), before data are to be made available externally (18+ months after sequencing). Data-contributing Teams may forgo the embargo periods and are encouraged to make data available to the Network and beyond as early as

[1] A data access model whereby qualified researchers apply for data access and their research plans are reviewed by a data access committee to ensure an ethically permissible balance between data protection and accessibility. Global Alliance for Genomics and Health, Data Privacy and Security Policy (2019).

possible. During the Network access period, only users approved by the Network and authenticated with an institutional identity/password can access data.

The Network Data Access Committee (DAC) reviews and approves Data Access Requests according to transparent review criteria, including the following privacy safeguards:

- Lead applicant on a Data Access Request must be a Principal Investigator (PI).[2]
- Access is for the purpose of a Study with local Research Ethics Board (REB) approval (affiliate institution of the Principal Investigator).[3]
- Study aligns with MOHCCN's aim (and patient consents) to advance medical research.
- Research team members needing access to data must be listed on the Data Access Request; team members participate under the PI's responsibility.
- *Additional review criteria may be considered for access to data from vulnerable populations (e.g., pediatric).*

The Network DAC is composed of members with relevant expertise in cancer research, bioinformatics, and law/ethics, and potentially also patient representatives. *Note: specific scientific and patient representative members may be included for access to data from special populations (e.g., pediatric).*

The Network DAC will only grant researchers access for a specific duration of 2 years, with possibility of renewal. The DAC as well as the Sites implementing user permissions will track and enforce this.

Where a researcher is authorized to access data, their Network Institution will be bound by a standard Data Access Agreement, a contract including obligations to safeguard data privacy, confidentiality, and security consistent with Network policies and the Network Master Agreement, which includes data access and use terms. Any access to data outside of this framework would require an equivalent legal agreement.

The Network agrees that certain **aggregate data/results** will be made available within the Network without the need for DAC or REB approval. These may include statistical summaries of Network data resources or the outputs of data discovery queries (e.g., aggregate # of cases matching features listed in the query). (*TBD*)

In principle, data will generally only be available through the secure access mechanisms on the distributed data platform. (*TBD*)

[2] An individual with "responsibility for the conduct and supervision of the activities described in the Implementing Email and/or Research Project Grant Agreement, as applicable." (Network Master Agreement) For external researchers, a standard definition will need to be developed reflecting standard funder/institutional definitions.
[3] As is currently required by the Network Master Agreement.

Table 1. Individual Participant-level Data to be made available through Controlled Access
Molecular data types containing germline data, e.g., • raw reads from WGS, RNA-seq or any other molecular assays, • germline variant data, • somatic variant data in non-coding regions.
Clinical data fields listed in the MOHCCN Clinical Data Model v 2.1. • *Dates of clinical events are collected by the Network but will only be shared as intervals since the date of diagnosis.* • *In addition the year of diagnosis will be included.* • *Outliers such as extreme ages (e.g., >90) are binned/generalized before sharing.* • *Note: Submitters may choose to submit dates with year and month only (YYYY/MM) or year, month and day (YYYY/MM/DD). An additional clinical data model field will indicate whether intervals have day or month resolution.*
Molecular data types not containing germline data, e.g., • somatic variant data in coding regions, • copy number changes, • gene/isoform-expression levels.
Imaging data (appropriately de-identified) [preferred data types], e.g., • digital pathology slide images, • cross-sectional imaging (CT/MRI/PET).

Document revision history

Developed by	Reviewed by	Endorsed by	Effective Date	Policy Version	Summary of revisions
DPSC	Steering Committee	Network Council	December 7, 2023	1	

Authors

Name	Institution	Title
Steven Jones (Chair)	BCGSC	Director of Bioinformatics
Lincoln Stein (Chair)	OICR	Head, Adaptive Oncology
Guillaume Bourque	McGill	Professor
Jennifer Chan	U of Calgary	Director / Associate Professor
Sidney Croul	NS Health	Lead, Pathologist
Daniel Gaston	Dalhousie	Lead, Bioinformatician
Stephanie Grover	SickKids	Program Manager
Benjamin Haibe-Kains	U of Toronto	Associate Professor
Timothy Hanna	Kingston HSC	Clinician Scientist
Martin Hirst	UBC	Senior Scientist
Anne-Marie Mes-Masson	U of Montreal	Associate Scientific Director
Jessica Nelson	BCGSC	Projects Team Leader
Dominique Trudel	CHUM	Pathologist / Associate Clinical Professor
Tran Truong	UHN	Director of Data & Technology
Emily Van de Laar	UHN	Project Team Lead
Ian Watson	McGill	Associate Professor
Ma'n Zawati	McGill	Assistant Professor