MOHCCN DATA ACCESS AND USE POLICY V1.1

Table of Contents

1. Introduction

The Marathon of Hope Cancer Centres Network (MOHCCN; the Network) Data Access and Use Policy defines the procedures and timelines for access to and use of data shared with the Network (Network Data) in research and publication endeavors. Network Data includes all data that is generated with the use of MOHCCN funding. Data generated with matched funding is encouraged to be made available by the Network and its Partners. Network Data may include genomic and other molecular data types, along with clinical and other health-related information.

Note: Within the initial agreement phase, MOHCCN covers MOHCCN Network members. The next phase of the agreement will include users outside of the Network such as external scientists and commercial "for profit" users. While this policy is inclusive of different user types, certain access provisions may need to be reconsidered as subsequent phases of the MOHCCN agreement are established.

2. Data Sharing

The MOHCCN is comprised of individual cancer research projects from across Canada; where clinical information is usually collected at the project site and data, which includes whole genome and transcriptome data, is generated at one of the sequencing centres. Genomic and other molecular data is generated for each specimen, and clinical and other health-related data is collected throughout the timeline of the project. The timelines for sharing should be appropriately set at project set-up to ensure there is enough time for proper quality control and to check the data integrity (Figure 1).

This section covers the use of Network Data from data generation to access for MOHCCN and other users. These policies are in place to ensure the fair and equitable access to Network Data for all researchers to enable the maximal possible insight and discovery and facilitate federated learning.

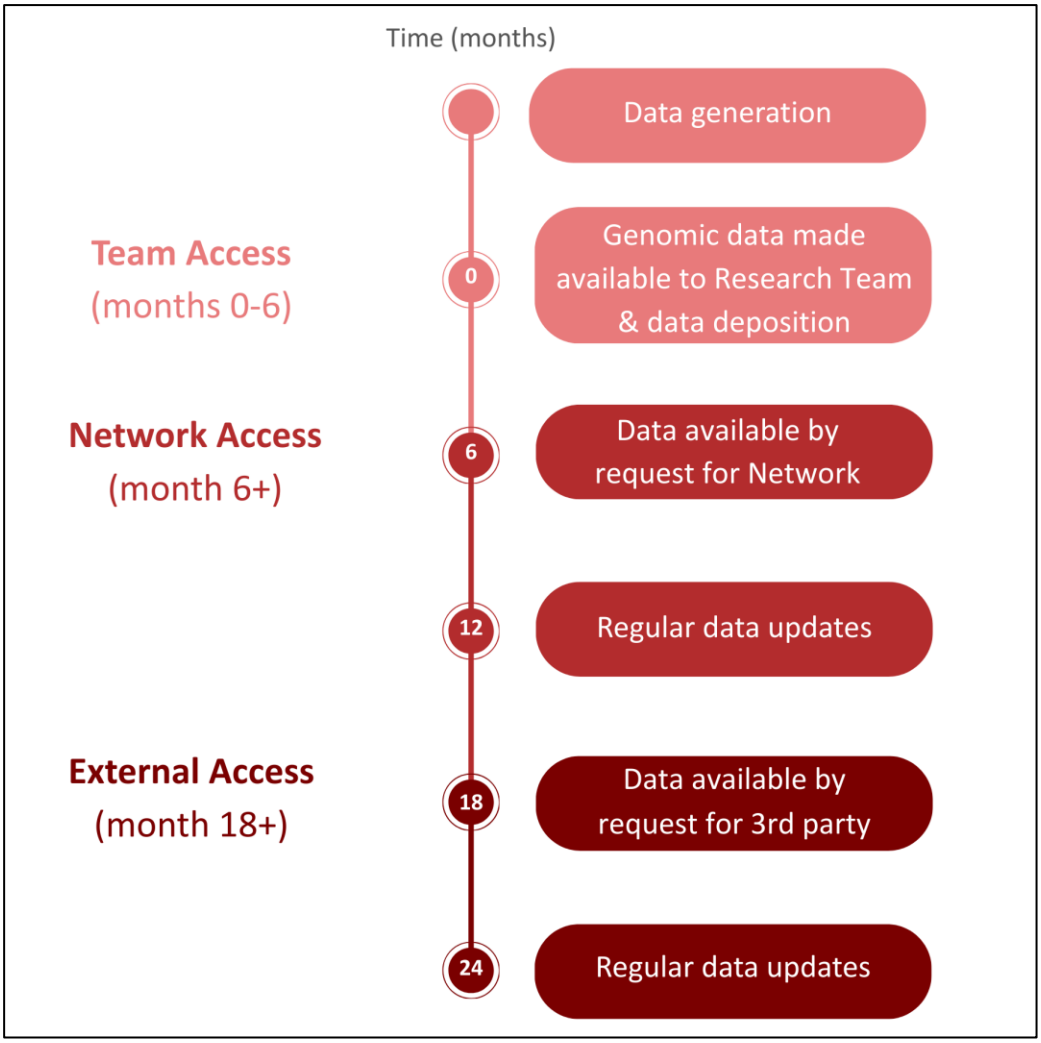

Figure 1

3. Data Generation & Deposition (0-6 Months)

Network Data are made available to be shared (only under an REB-approved study) on a patient-by-patient basis as the data is generated; the data unit outlined in this policy is for individual patients, not datasets or full cohorts.

Material Derived Data

Material Derived Data means research subject-level or patient-level data generated from the analysis of Biospecimens. For example, molecular data such as generated from assays such as whole genome sequencing (WGS) and whole exome sequencing (WES), including but not limited to aligned reads (FASTQ/BAM/CRAM) and variant calls (VCF/MAF).

The sequencing centres should share genomic data with the Network in a timely fashion. As a general rule, genomic data for each sample should be made available by the sequencing centre to the research team within 1 month of completing quality control (QC).

Other Molecular Data

Molecular Data other than Genomic Data may include a wide range of proteomic, metabolomic, cellular, and immuno-histochemistry data (timelines to be determined).

Clinical Data

Clinical Data is any medical data that is collected from the patient's medical chart or from the patient and may include patient status, diagnosis, treatment, outcomes, imaging, medical notes, reports, and laboratory values. MOHCCN recognizes that the collection of the clinical data is a long-term endeavor, that project sites will need to share data with the Network at multiple time points, and the resources for collection may vary from site to site. Harmonization of data collection will be a priority for the Network.

The mandatory MOHCCN clinical data should be made available to the Network as soon as possible, ideally at the time as genomic data is available, and must be complete by the end of the Team Access period (6 months after data are made available to the research team - see section below). Thereafter, clinical information should be updated for purposes of making such data available, at minimum yearly for each case or more frequently if possible. MOHCCN recognizes that some fields will not be applicable to certain cancer types and some historical information, such as prior response to therapies or diagnostic reports, might not be available for collection. The most recent Clinical Data Model can be found on the MOHCCN website here.

Patients on Clinical Trials

For patients enrolled in ongoing clinical trials, it is acknowledged that certain clinical outcome information pertaining to trial will not be able to be shared until the trial has completed. However, it is expected that clinical information up until the time of enrollment into a trial will be shared. Therefore, some of the clinical fields may be restricted, such as I^{ry} and II^{ry} endpoints, and therapeutic treatment resulting from the trial may therefore be withheld beyond 6 months until trial completion or termination. All other clinical data available at the time of biopsy must be shared within the standard timeframe.

The participation in a clinical trial will therefore not exclude the genomic and transcriptomic data from being shared. It should also not preclude the collection and sharing of any clinical data obtained prior

to the trial, or any data fields not explicitly restricted by the trial protocol. The participation of the patient in a clinical trial may be omitted from the provided clinical information until trial completion if preferred by the submitting group. The submitting group should also confirm that they will provide the relevant trial and post-trial clinical information once the trial is completed in a timely manner.

In cases where the trial sponsor agreement and/or pre-existing study requirements precludes or limits the ability to share the genomic data and/or non-trial related clinical information or where intellectual property encumbrances might exist, the submitting group should seek clarification and approval from the TFRI MOHCCN executive group that the generation of genomic data is beneficial to the Network or whether the generation of such data should be delayed until trial completion.

Other Health-Related Data

Other Health-Related Data may include patient-related administrative data and patient-reported outcomes data (timelines to be determined).

4. Team Access (Months 0-6)

Initially, case data is available only to the team having "contributed" the data for their research purposes for six months (the "Team Access" period). For clarity, the embargo period start date for each case is the earliest date when any sequencing data is made available to the research team by the genome sequencing centre.

5. Network Access (After Month 6)

Six months after the data are made available to the research team, they must be made available to the rest of the Network (start of the Network Access period). Sites may exceptionally request up to four extra weeks from TFRI where justified. During the Network access period, Network members may use the data for approved publications and IP development. Network members are expected to work collaboratively and must be respectful and inclusive of their colleagues. By bringing together researchers and clinicians with shared research interests and questions, the Network aims to foster an open and collaborative scientific environment that all participants benefit from.

Any publications resulting from the use of Network resources must adhere to the MOHCCN Publication Policy and Code of Conduct (Policies and Guidelines).

Data Updates

Projects are required to follow up with their patients regarding any return of individual research results as appropriate, and to provide regular data updates to the Network (e.g., longitudinal clinical data).

Analysis & Publications Data

Publications resulting from the use of Network resources must follow the MOHCCN Publication Policy (Policies and Guidelines).

6. External Access (After Month 18)

At 18 months after data have been made available to the research team, non-MOHCCN member researchers in Canada and abroad may request access to restricted MOHCCN data through the platform via a Registered Access or Controlled Access data model. Requests for controlled access data will be made to the Data Access Committee (DAC) and will require a Data Access Agreement to be in place between the MOHCCN and the third-party researcher and their organization. Data permissions will be uniform within a group and individuals may register by showing proof of membership to the requested group.

Note: In the initial phase, MOHCCN only covers sharing with MOHCCN Network members. The next phase will include users outside of the Network and commercial "for profit" users. As these users are onboarded, access provisions may need to be considered.

7. Data Access Model

All Network Data will be shared and available for access by other Network Members, as well as by the wider Canadian and international research communities, following the phases and timelines described above. Furthermore, access procedures and approvals will be contingent on the level of privacy risk engendered by data sharing and use outlined in the Privacy Policy. Unrestricted Access Data will be publicly available for use without any access controls. Network Data requiring access controls will be available to the research community *via* Registered Access and Controlled Access procedures.

It is important to reiterate that genomic data that contains information that could potentially identify a participant would not be made publicly available. This includes sequence information and analysis that contain germline variants. Such identifying data would include all sequences deriving from tumour, normal tissue control and transcriptome. Potentially identifying data would always be considered to be protected and restricted. The sharing of protected data would be under the purview and control of the MOHCCN DAC.

8. Addressing Regional Inequity

MOHCCN recognizes that data sharing can be disadvantageous for investigators and centres who do not have the capacity or resources for rapid analysis of their data. However, data sharing can redress an inequity ensuring that a patient’s contribution can be studied to its maximal benefit regardless of the capacity of scientists to study it in their local region.

The collaborative Network research process is additionally meant to encourage smaller centres to take advantage of resources that may only be available at larger centres, such as expertise in novel bioinformatic or AI tools. In particular, the first scheduled publication for MOHCCN resources will either be led by, or include, the teams having contributed those materials if they wish to participate in that research. A strong community of collaboration will provide new training and learning opportunities for members of all the Network labs.

Document Revision History

Developed by	Reviewed by	Endorsed by	Effective Date	Policy Version	Summary of revisions
DPSC	Steering Committee	Network Council	June 2,2022	V1	n/a
DPSC	Steering Committee	Network Council	February 27, 2026	V1.1	Clarification of embargo start date. Update of outdated information.

Authors

Name	Institution	Title
Steven Jones (Chair)	BC GSC	Director, Bioinformatics
Lincoln Stein (Chair)	OICR	Head, Adaptive Oncology
Guillaume Bourque	McGill	Professor
Jeffrey Bruce	UHN	Bioinformatics/Data Science Lead
Christine Caron	CHUM	Project Manager
Jennifer Chan	U of C	Director / Associate Professor
Karen Cranston	UHN / CanDIG	Technical Project Manager
Daniel Gaston	DAL	Lead, Bioinformatician
Benjamin Haibe Kains	U of T	Associate Professor
Martin Hirst	BCC RI / UBC	Senior Scientist / Professor
Anne-Marie Mes-Masson	CHUM	Associate Scientific Director
Hardeep Nahal-Bose	OICR	Senior Manager
Jessica Nelson	BCC	Projects Team Leader
Carolyn-Ann Robinson	U of C	Senior Research Associate
Enrique Sanz-Garcia	UHN	Assistant Professor / Staff Medical Oncologist
Tran Truong	UHN	Director
Ian Watson	McGill	Associate Professor
Shira Yair Sabag	BC GSC	Projects Coordinator
Ma’n Zawati	McGill	Assistant Professor