COVID-19 IMMUNITY TASK FORCE

Quick Guide on Data Sharing Funded Studies

Version 2.0 | October 5, 2020

Table of Contents

1. Background

Understanding the disease epidemiology of SARS-CoV-2 (COVID-19) is paramount to slow the spread of infection, minimize mortality, and provide robust evidence to inform policy making. In response to this unprecedented public health crisis, the COVID-19 Immunity Task Force (CITF) has been given the mandate to fund studies to investigate the epidemiology of COVID-19 in Canada. These funded studies will contain a variety of data including clinical, serological, administrative, demographic, socioeconomic, cultural practices and lifestyle, and environmental data, among others.

In order to make timely use and reuse of these data the CITF is coordinating a system for collecting, documenting, and disseminating COVID-19 related data and metadata to ensure data arising from these studies is well documented, interoperable, complete, methodologically heterogeneous, and of high quality. The following guide is intended to describe and facilitate the process of data centralization and data sharing with the CITF (Figure 1).

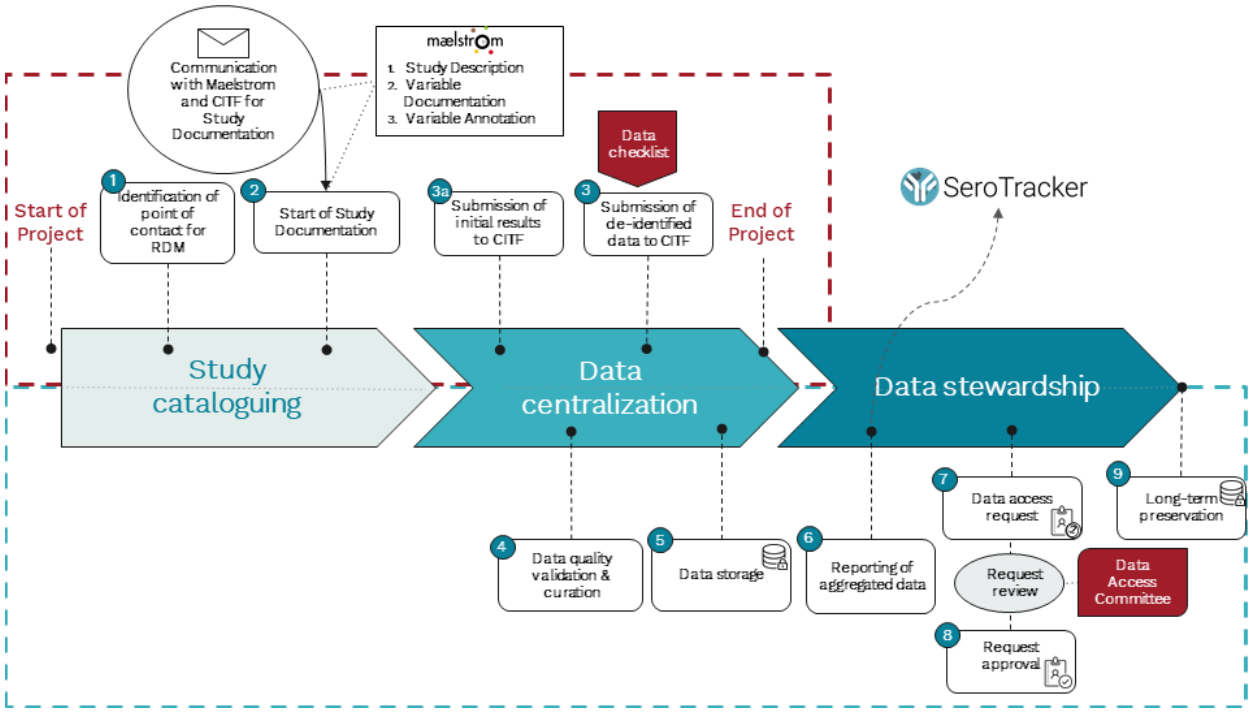

Figure 1. This diagram represents the general data life-cycle for CITF-funded studies, including data sharing after the end of the project. Actions from studies are described in the upper half diagram (red dashed square), and actions from the CITF Secretariat are described in the lower half (aqua dashed square).

2. Study documentation

To optimize the usage of data produced from CITF-funded studies, and to promote scientific discovery about COVID-19 immunity in Canada, the CITF plans to partner with the Maelstrom Research Group, led by Prof. Isabel Fortier from the Research Institute of the McGill University Health Centre, to implement a high-quality metadata catalogue across these research project. Maelstrom's cataloguing toolkit has been extensively used by international initiatives to document observational cohort studies, and to disseminate their metadata with the scientific community.

The cataloguing process used by the Maelstrom Research team to document individual studies is broadly divided in 3 steps: study description, variables documentation, and variables annotation, which are described below. CITF-funded studies are expected to interact with Maelstrom before data collection to begin the study documentation process—although this may be impossible for studies currently on underway. It is encouraged to appoint a point of contact for such communications.

2.1 Study description

The first step is for study design, targeted population(s), and data collection events to be documented. The Maelstrom team will begin by describing the study using different sources including published papers and study website, and subsequently validate this information with the PI.

2.2 Variables documentation

During the early stages of project initiation, standardized variable dictionaries will be selected or prepared. The Maelstrom team will obtain the questionnaires and data dictionaries from projects for study documentation. Variables should at least have a name, a label, and if applicable, codes and labels for categories. All study variables (not only Core Data Elements) should be catalogued as they are determined.

2.3 Variable annotation

The last step will be for each variable in the study to be classified in at least one domain and subdomain of the Maelstrom Research classification. The Maelstrom team will attribute each variable to one or more subdomains of the areas of information with the help of the questionnaires and the information documented in the previous cataloguing steps. These attributions will be validated with PI to avoid any discrepancies.

General study characteristics, targeted populations, data collection events, and data dictionaries will be made available on Maelstrom's website and linked to study data in the CITF database as it becomes available.

The complete list of Maelstrom's metadata fields is provided in Appendix 1.

Note: Studies that have already began their study-documentation in a context other than the CITF, should contact the CITF Secretariat to arrange how to best harmonize their current documentation.

3. Data elements

Given the nature of the data collected by CITF-funded studies, data centralization can be greatly optimized by the use of shared data elements (i.e., Core Data Elements). The CITF has defined these core data elements, which all studies are expected to measure and make available to the CITF Secretariat, as well as the serology or laboratory results for individual subjects. These data elements are broadly defined in the following CITF Domains: Demographic, COVID-19, Symptoms, Exposure, Health and Health Behaviours, and Serology.

The Core Data Elements do not constitute the whole suite of data elements individual studies will measure. Nonetheless, they are minimum data required by the CITF to provide the Public Health Agency of Canada regular scientific updates on the state of serologic testing and the evolving understanding of immunity related to SARSCoV-2 in Canada.

For the list of Core Data Elements, please visit the **CITF website.**

Note: Studies that have already began their data collection before these Core Data Elements are disseminated should contact the CITF Secretariat to arrange how to best harmonize their data elements with those identified by the CITF.

CCITF Domain	Visit	Number	Element	Question	Type	Technical specs
DEMOGRAPHIC	Baseline	1	Date of interview	Date of interview	Date	DD/MO/YR
DEMOGRAPHIC	Baseline	2	Age	What is your age	Numerical	YEARS
DEMOGRAPHIC	Baseline	3	Sex	What was your assigned sex at birth	Categorical	MALE/FEMALE/PREFER TO SELF-DESCRIBE/PREFER NOT TO ANSWER
DEMOGRAPHIC	Baseline	4	Sex	What is your sex now	Categorical	MALE/FEMALE/PREFER TO SELF-DESCRIBE/PREFER NOT TO ANSWER

4. Legal and ethical approvals

Data sharing reduces duplication of effort, facilitates efficiency and transparency, and maximizes the benefit from funding from the CITF. Despite the pressing need for the rapid sharing of COVID-19 related data, this cannot be done at the expense of individuals' privacy.

4.1 Consent and approvals

To ensure adherence with local, provincial, and federal laws, CITF expects that each study will obtain all the required permits to conduct their research in accordance with provincial and federal laws, as well as their local institutional requirements.

CITF will assist projects with consent guidance and "templates", and will also support studies for the inclusion of past data collections.

4.2 Data anonymization

All data submitted to CITF must be anonymized or at least de-identified (removal of identifying information such as names, dates of birth, etc.).

Data beyond the Core Data Elements that could present a higher risk of potential re-identification will be managed with extra precautions in CITF sharing mechanisms. Projects' data sharing plans will be developed in collaboration with CITF Secretariat and on a case by case basis.

Confirmation that approvals are in place and that data have been anonymized will be sought upon data submission to the CITF database.

Note: Studies that need support with any of these aspects should contact the CITF Secretariat for advice.

5. Sharing study data with CITF for centralization

In order to securely receive data from CITF-funded studies, the CITF has developed a secure web portal, where studies will have access to private and public spaces.

Public spaces will be used by the CITF to host guiding documents and templates for general use by the studies.

Private spaces will be used to upload data and documents, and will be shared between a given study and the CITF. Data exports, and uploads of executed agreements and forms will be done in these private spaces. Data exports are expected to be provided in a standard-common-open format such as "csv".

The data upload portal can be found at **https://pydio.mchi.mcgill.ca** and should be accessible from anywhere. This platform is hosted privately, in a secure environment maintained by the CITF at McGill University. All interactions about data products from or to studies will be done through this portal.

Appendix 1

Metadata fields in the Maelstrom Research metadata catalogue.

STUDY	
Field	**Definition**
Name	Official name of the study.
Acronym	Study acronym.
Website	Study website URL.
Investigators	Name, affiliated institution and contact information of the principal investigators.
Contacts	Name, affiliated institution and contact information of the person to be contacted to have more information about the study.
Objectives	Main objectives of the study.
Study timeline	Date when first participants were recruited and study end date if the study is completed.
Study design	Information on specific study design. 1. Cohort 2. Case-control 3. Case only 4. Cross-sectional 5. Clinical trial 6. Other
General information on follow-up	Profile and frequency of participants' follow-up (e.g. Participants are followed-up every 5 years).
Supplementary information about study design	Additional information about study design (e.g. Subgroups of the population were intentionally over-sampled).
Recruitment target	Type of participant units targeted by the study. 1. Individuals 2. Families 3. Other
Number of participants	Number of participants planned to be recruited. If the study is completed, the final number of participants.
Number of participants with biological samples	If the study is collecting biological samples, number of participants that should provide samples. If the study is completed, the final number of participants that provided biological samples.
Supplementary information about number of participants	Additional information about target number of participants (e.g. Additional biological samples will be collected for population 2).

Access	Whether access to study data, biological samples or other study material by external researchers or third parties is allowed or foreseen.
Marker paper(s)	Bibliographic citation(s) which should be used to refer to the study and, if applicable, the paper's Pubmed ID.
Logo	Logo used by the study.
Documents	Relevant documents about the study (e.g. Questionnaires, standard operating procedures, codebooks).

POPULATION	
Field	**Definition**
Name	Name of the study population.
Description	A brief description of the population.
Sources of recruitment	Specification of the sources of recruitment. 4. General population (volunteer enrolment, selected sample, random digit dialing) 5. Specific population (clinic patients, members of specific association, other specific population) 6. Participants from existing studies 7. Other source
Supplementary information about sources of recruitment	Additional information about recruitment procedures (e.g. Participants were identified from the electoral register and general practice lists).
Selection criteria	If relevant, specification for the following selection criteria of the participants. 8. Gender (women or men) 9. Age (minimum age and maximum age) 10. Residence (country, territory or city) 11. Pregnant women (first trimester, second trimester, third trimester) 12. Newborns 13. Twins 14. Ethnic origin 15. Health status 16. Other
Supplementary information about selection criteria	Additional information about selection criteria of the population (e.g. All subjects identified at baseline as affected by cognitive impairment without dementia were eligible for the longitudinal phase conducted after one year).
Number of participants	Number of participants planned to be recruited for the population. If the study is completed, the final number of participants.

Number of participants with biological samples	If the study is collecting biological samples, number of participants that should provide samples for the population. If the study is completed, the final number of participants that provided biological samples.
Supplementary information about number of participants	Additional information about number of participants. Usually the number of participants for each wave of the study (e.g. Number of participants for each data collection event Wave 1: 7175 participants Wave 2: 3145 participants Wave 3: 1733 participants).

DATA COLLECTION EVENT	
Field	**Definition**
Name	Name of the data collection event.
Description	A brief description of the data collection event.
Data collection event date	Data collection start date and end date.
Data sources	Data sources from which the information is obtained. 17. Questionnaires 18. Physical measures 19. Cognitive measures 20. Biological samples (blood, cord blood, buccal cells, tissues, saliva, urine, hair, nail, other) 21. Administrative databases (health databases, vital statistics databases, socioeconomic databases, environmental databases) 22. Others (e.g. medical files)

DATASET	
Field	**Definition**
Name	Name of the dataset.
Acronym	Dataset acronym.
Description	Short description of the dataset specifying its content.
Entity type	What the data are about (usually the participant).