

Proposal Summary

The CITF Databank is the only accessible repository of individual-level epidemiological and serological data on COVID-19 in Canada. Predicated on internationally aligned best practices of data governance and sharing, the CITF Databank enables researchers to maximize the return on publicly funded SARS-CoV-2 research by sharing results and individual-level data from CITF-supported studies. When completed, the Databank is expected to host individual-level data on 105 variables collected on nearly 200,000 participants in over 60 epidemiological studies of immunity to COVID-19 from diverse populations across Canada. The variables measure demographics, COVID-19 history and vaccination, symptoms, exposure, health and health behaviours, and SARS-CoV-2 serology. Individual-level records are available to Canadian and international researchers through a restricted access mechanism, where data can be requested either in the format submitted by individual studies or as a single harmonized dataset across multiple studies. The Databank also provides open access to statistical aggregates and study metadata through the CITF Open Access Dataverse [1], which has already had over 3,700 downloads.

The goal of the proposed research is to realize the full potential of the Databank by undertaking the six activities listed below, which will complete the development of the CITF Databank, expand the potential for research to be conducted with its data, and disseminate widely the resulting knowledge.

Activity 1: Curate, Organize, and Harmonize Incoming Data. All CITF-sponsored studies that have not yet deposited their data will be incorporated into the Databank, documented, and harmonized.

Activity 2. Enable Secondary Analyses and Ensure Open Access. The data access infrastructure built under the CITF will be adapted for long-term sustainability and in support of open science.

Activity 3. Manage Data Requests and Provision Data to Researchers and Policy Makers. The current data management processes will be optimized to ensure ongoing governance of data access including rapid, unbiased application review and fulfilment of data access requests, while maintaining data security and integrity.

Activity 4. Encourage and Promote the Use of Data and the Mobilization of Knowledge. Use of the Databank will be promoted by eliciting user needs and developing or adapting processes and tools to meet these needs. We will also mobilize the knowledge resulting from analyses of the data and from work on data standards and innovative approaches to interacting with the Databank to ensure that lessons learned can help Canada to prepare for future pandemics.

Activity 5. Develop and Maintain Data Dashboards to Visualize Data. Dashboards and other visualization tools will be developed and maintained in response to input and feedback from researchers.

Activity 6. Provide Analytical Consulting and Develop Methods. Analytical and methodological consulting will be provided to researchers and public health agencies using the Databank and methodological research to support serosurveillance and data management will be conducted.

The proposed activities will be led by a multidisciplinary team of investigators and completed by a core team of experienced staff with expertise in biomedical informatics, epidemiology, mathematics, statistics, and immunology. This team is qualified to achieve our goal of completing the Databank, maximizing the use of the data, and disseminating widely the resulting knowledge and lessons learned.

Introduction

In April 2020, the Government of Canada launched the COVID-19 Immunity Task Force (CITF) in response to the public health emergency caused by SARS-CoV-2. CITF coordinated national research on COVID-19 spread and immunity, funding studies, communicating findings, and enhancing data collection. A central aspect of the CITF mandate was to centralize and make available data from supported studies. To realize this aspect of the CITF mandate, we established the CITF Databank. Development of the Databank included the creation of a Data Governance Framework, common consent elements, and Core Data Elements to ensure that study data could be centralized and shared for research in Canada and internationally. After receiving several study data sets and establishing the Data Access Committee, the Databank officially opened for data requests in December 2022.

Since its launch, the number of datasets available for secondary analysis within the Databank has steadily increased, and multiple data access requests have been received and fulfilled. The data contained in the Databank hold enormous potential for different types of research, including meta-analyses, modelling studies, and immunological research. There is no other Canadian or international repository with similar content and accessibility. However, despite the novelty and potential of the Databank, when the CITF mandate officially ends on March 31st, 2024, the integration of all eligible data into the Databank and their harmonization will not be complete. Therefore, to reach the full potential value of the Databank and to maximize the returns from the considerable investment in CITF, it is critical to complete the development of the Databank and to increase the use of this monumental data resource by researchers.

Goal

Our goal is to realize the full potential of the Databank by completing the development of the Databank, expanding the potential for research to be conducted with its data, and disseminating widely the resulting knowledge. We will accomplish this goal by transitioning to a sustainable databank governance and management structure that is independent of the CITF and by conducting a coordinated set of activities, including finalizing data integration, encouraging data usage through outreach and the Databank web pages (including data dashboards), and sharing methodological insights.

Databank Governance and Management

The existing Databank assets and procedures are predicated on the CITF Data Governance Framework (Framework), which defines the legal and ethical foundations for data sharing from CITF-supported studies in alignment with international and evidence-based standards and principles. The Framework establishes a comprehensive approach to data sharing, including defining mechanisms for open access to statistical aggregates and for restricted access to centralized individual-level records. For sharing individual-level data, the Framework defines a set of consent elements that were required of any study depositing data into the Databank; the consent elements facilitate further sharing of data by ensuring that all data can be shared under a common set of conditions. A Data Sharing Agreement (DSA) template was developed from the Framework and used as the basis for agreements between McGill University and the institutions sharing data through the CITF. The Data Access Policies flow from the Framework and further define the conditions under which data may be used by a Data recipient. Applications for data access are managed by the Data Access Office (DAO) and are reviewed by an independent Data Access Committee (DAC), comprising experts in data governance, immune science, epidemiology, and indigenous health. Approved applicants must enter into a Data Access Agreement with McGill University prior to receiving data for research. These established data sharing

policies and practices will continue to ensure appropriate governance, as described in the Core Data Activities below.

After the CITF ceases to exist on March 31, 2024, the governance and management of the Databank will remain like the current structure described above, but some roles will be filled differently (Figure 1). Most notably, we will replace the strategic leadership and oversight that was provided by the Public Health Agency of Canada and the operational support that was provided by the CITF Secretariat. The CITF Databank will continue to be hosted at McGill University and will carry forward the Framework and Data Access Policies, both of which were designed to survive the CITF. The strategic oversight of the Databank governance will be provided by the investigators with input from an Advisory Board of scientists and stakeholders. The NPA (Buckeridge) and former Executive Director of the CITF Secretariat (Evans) will co-chair the Advisory Board, which will provide input regarding strategy and operations related to data sharing. McGill University will continue to support governance and management through the Faculty of Medicine and Health Sciences Institutional Review Board and the Office of Vice-President, Research & Innovation (e.g., Conflict of Interest), including the Office of Sponsored Research (e.g., contract support, financial support). Day-to-day Databank management will continue under the direction of the NPA, with advisors and staff from the CITF moving to become investigators or research staff under the current funding opportunity. These roles include Scientific Advisors, and the Data Management Team: Databank Manager, Data Analysts, and Software Developer, Medical Informatician, Data Access Lead, Data Governance Lead, Harmonization Lead. The majority of the roles are already staffed.

Indigenous Peoples and Communities Data Governance: In accordance with the Framework, the CITF Databank will not include data from studies that have focused exclusively on Indigenous populations. However, some participants in CITF supported studies have self-identified as Indigenous, and in facilitating the sharing of data with researchers, some manipulations and interpretations of the data could have implications for Indigenous Peoples and their communities. Given the historically inequitable and harmful nature of research involving Indigenous Peoples and their communities, the CITF has worked with organizations, such as the First Nations Information Governance Centre (FNIGC) [2], to draft a set of principles, found in the Framework, which are used to protect data assets related to Indigenous Peoples and their communities. These principles are guided by The First Nations Principles of OCAP® (ownership, control, access, and possession) [3], which assert that First Nations have control over data collection processes and that they own and control how this information can be used. In addition, an expert in Indigenous health research was appointed as a DAC Member to help oversee responsible use of data about First Nations, Inuit, and Métis peoples.

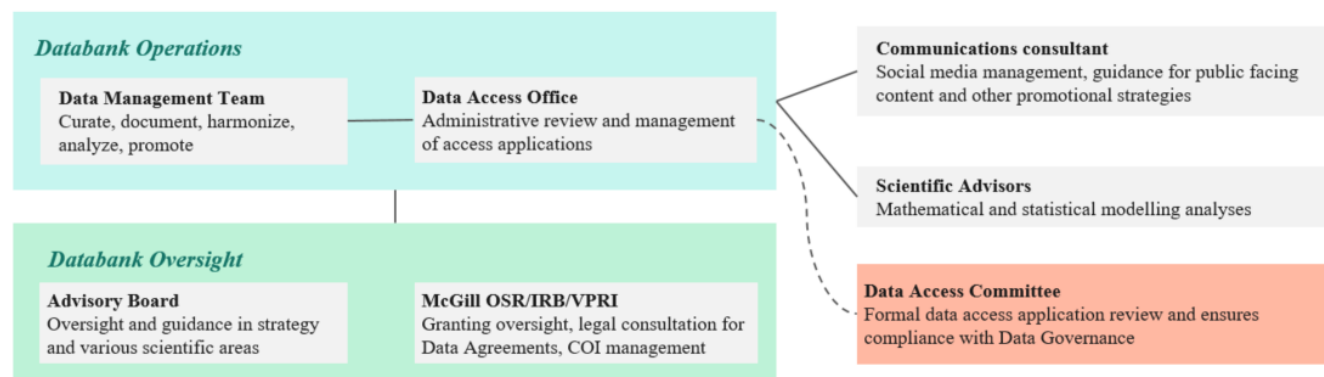


Figure 1. CITF Databank governance (OSR – Office of Sponsored Research; IRB – Institutional Review Board; VPRI – Office of Vice-President, Research & Innovation; COI – Conflict of Interest)

Risk Mitigation

The main risks we have identified are the risk of study participant identity disclosure, stigmatization risks, and the sustainability of the Databank:

- The risk of participant identity disclosure is managed through the data management and access policies under the Data Governance Framework. These include individual consent, IRB oversight, secure IT infrastructure, data sharing agreements, and obligations of all McGill employees working with data. The Data Dashboard also uses various confidentiality techniques to ensure individuals cannot be identified.
- Stigmatization risks arising from the misinterpretation or misuse of research findings may have negative impacts especially on marginalized or vulnerable populations, such as but not limited to Indigenous Peoples and Communities. Strategies adopted to mitigate these risks include IRB oversight, data access policies requiring appropriate engagement with communities and adequate justification of any requests for sensitive data in research protocols.
- A final risk to consider is the ongoing maintenance and support of the Databank in order to ensure sustained data access under proper conditions. For example, maintaining proper data conservation personnel and infrastructure and conducting audits for data destruction and security when necessary. We will work with the Advisory Board to identify potential sources of long-term funding to maintain necessary core services for the Databank beyond the life of this funding opportunity.

To take accountability of individual identity protection and long-term sustainability, strategies will be further defined through a Privacy Impact Assessment (**Activity 2**).

Proposed Activities

Core Data Activities

The first set of proposed activities support the core functions of bringing data into the CITF Databank and managing those data (**Activity 1**), enabling requests by researchers to access those data (**Activity 2**), and providing data to researchers for further study (**Activity 3**). We provide an overview of these activities in the paragraph below and in Figure 2 and describe the activities in detail in the following sections.

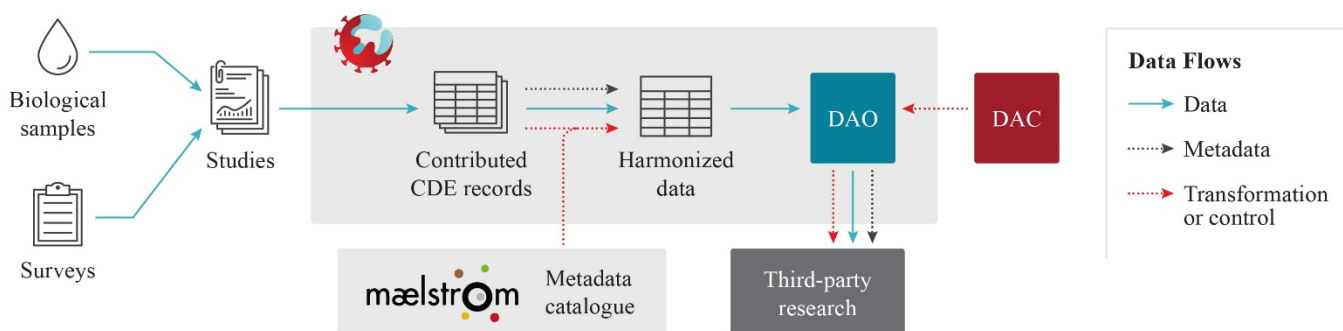


Figure 2. Overview of CITF Databank data flows (CDE – Core Data Elements; DAO – Data Access Office; DAC – Data Access Committee).

The CITF Databank holds individual-level records deposited by CITF-supported studies that have met the ethical and legal requirements to deposit data, including a DSA between the institution where the study was performed and McGill University. Individual-level records in the Databank are available through a restricted access mechanism in two formats, as data submitted by individual studies and as data harmonized across multiple studies. Access to both formats of individual-level records by Canadian and international researchers requires an application to the DAO. Metadata about the original studies and contributed microdata are documented in standardized formats on Maelstrom Research [4] [5] and Borealis McGill [1], respectively. The microdata include common data variables, termed Core Data Elements (CDE), which were designed through a rigorous consultation process with committees of scientific experts. The CDE cover demographics, COVID-19 history and vaccination, symptoms, exposure risks, health and health behaviours, and SARS-CoV-2 serology.

Activity 1. Curate, Organize, and Harmonize Incoming Data

Of the over 60 studies that have signed or are currently negotiating a data sharing agreement, 56 have initiated the data centralization process, 43 have uploaded data, and data have been harmonized across 20 studies. The Data Management Team developed processes to curate, standardize, centralize, and harmonize individual-level data following international data standards. We will continue these processes until data sharing agreements are signed for all studies and all data have been centralized and harmonized, as described below.

All studies were supported in developing consent materials so that they could ethically and legally deposit data into the Databank. Standardized questionnaire items and data schemas for collecting and storing were also provided to facilitate data harmonization. All attempts were made when defining the CDE to include Canadian and international recommended terminology in the questions and response choices [6, 7]. Studies were supported in the adoption of these CDE into their data collection procedures. Following the successful negotiation of a DSA, each study is asked to share their data dictionary with data variable definitions, including standardized reporting of serology and other assay data using ‘CITF assay IDs’, which characterize assays using a standard schema. The serology data are further coded and organized using existing standards (e.g. LOINC, SNOMED [8]).

To centralize data from a study, CDE variables are described using the study’s data dictionary and questions and comments about the CDE are communicated to study teams for documentation and data preparation purposes. The study team then uploads their data to a web-accessible, password-protected data storage platform housed at McGill. Our team reviews the uploaded data and clarifies with the study team any changes in data variables and/or definitions, documents study related details, and if needed, assists in further data preparation steps. Additional data, from any updates or follow-up visits, are shared with us and are uploaded as they become available.

Once a study’s data are uploaded, they are moved to a secure password-protected PostgreSQL v14 (PostgreSQL Global Development Group) database at McGill, where access to them is controlled and monitored. An R Server [9] utilized by our team connects directly to the database for further data quality improvement, to perform the data harmonization, and to conduct any analyses. Harmonization of the data is an iterative and ongoing process, which is applied periodically to create a pooled dataset across the data from all available studies [10]. Questions arising during the harmonization process, regarding variable definitions, missing or incomplete data, and invalid values, are addressed with the study team and their resolutions are documented. Currently, only the data harmonization schemas for baseline questionnaire data and repeated measures serology are developed, and we will continue incorporating new study datasets into the current harmonized dataset as they become available. Furthermore, we will work to extend the scope of the harmonization schemas to include COVID-19 vaccination information, data from follow-up visits, and data from pediatric participants.

Activity 2. Enable Secondary Analyses and Ensure Open Access

The CITF Databank access infrastructure will continue to operate in accordance with the FAIR and TRUST principles for scientific data management and data repositories [11, 12]. To increase findability and transparency, the Databank has been registered with the re3data directory of repositories [13] and all study metadata will be added to the CITF Dataverse on Borealis [1], which assigns a universal Document Object Identifier (DOI) code for each submission. This platform provides open access to the metadata but does not accept restricted-access individual-level study data, so only the metadata for study data sets will be deposited into the CITF Dataverse and the study data will be maintained in a secure server hosted at McGill University.

The data in the Databank are available for access by any researcher affiliated with a public or private institution, including institutions outside of Canada. The DAO and DAC will continue to operate and review applications in a transparent and unbiased manner. To ensure data integrity and quality, the incoming data to the Databank will continue to be cleaned, curated, and organized as described in **Activity 1**. The technological infrastructure that houses the Databank allows for secure transfers (inbound and outbound) and is built for long-term sustainability. Updates or changes to the infrastructure will be subject to approval by the arms-length DAC.

The Data Management team will continue to develop internal documentation and public facing content outlining the Databank's processes and controls to protect participants' privacy and strategies for risk mitigation. For example, we have initiated a Privacy Impact Assessment, and under the guidance of the Advisory Board, we will pursue other forms of certification to maintain legitimacy and reliability as a trusted data repository.

Consistent with the Framework, we will continue to make aggregate-level data available through open access. The CITF has previously curated aggregated seroprevalence estimates from individual studies and collaborators such as the Canadian Blood Services. These estimates were reported with interactive plots on the CITF website [14], and are openly accessible through Borealis [1]. All open datasets and supporting metadata are deposited into the CITF Open Data Dataverse and assigned a DOI. As of December 2023, there were over 3,700 downloads of open data. To build on the success of the open-data tier, additional open-access datasets derived from statistical aggregates of individual-level records will be added. Further development of open datasets will be guided by the user needs assessment as described in **Activity 4**. In **Activity 5**, these data will also be made available for visualization through the Dashboard.

Activity 3. Manage Data Requests and Provision Data to Researchers and Policy Makers

Applications to access data held in the CITF Databank must be supported by an REB-approved research protocol that is scientifically sound and of appropriate scope. Applicants are required to provide clear justification for why Databank data are required for their research. Access applications must be submitted to the DAO through the Data Access Portal, an online interface where researchers can select harmonized variables of interest and submit their application and relevant documentation (e.g., study protocol, ethics certificate, CV). The Portal will be continuously updated to reflect the current harmonization protocols and data schema as per **Activity 1**.

Applications will undergo an administrative review by the DAO, followed by a review by the independent DAC. The DAC, which includes five members with expertise in data governance, immune science, epidemiology, and indigenous health, will meet as necessary to review applications, with the goal of responding to requests within 2-4 weeks. Requests may be deemed invalid by the DAO or rejected by the DAC should the scientific research objectives not follow the guiding principles, or if insufficient information or documentation is given. However, in most cases, this feedback will be used by researchers to modify and resubmit their request.

Successful applicants will be bound by the terms of a *Data Access Agreement* with McGill University. To ensure the security of the Databank data and the protection of privacy of individual participants for whom data have been collected, all approved users must comply with the application procedure and publication policy.

Once the *Data Access Agreement* is signed, the requested dataset(s) will be prepared and transferred to the applicant through a secure, web-based data storage platform hosted at McGill. As the applicant must submit an ethics approval certificate, the expiry date of the approval will be logged and tracked with reminders sent out indicating that ethics must be renewed, or data access will be revoked, and researchers will be required to submit evidence that the copy of data in their possession has been appropriately destroyed.

Activities Supporting Use of the Databank

The Core Data Activities described above focus on the curation, management and sharing of data. To complement these activities, we propose activities to maximize the value of the CITF Databank. These activities include encouraging and promoting its use (**Activity 4**), developing, and maintaining data dashboards to allow researchers to visualize data in the Databank (**Activity 5**), and providing analytical and methodological consulting to researchers and networks (**Activity 6**). We describe the activities in detail in the subsequent sections.

Activity 4. Encourage and Promote the Use of Data and the Mobilization of Knowledge

To maximize the scientific return on the original investment of \$300M made by the federal government to the ‘Sero Surveillance Research Program’, which included funding to the CITF, it is important to understand the needs of potential users of the Databank. We will conduct focus groups with potential Databank users to better characterize them and how they could use the Databank. This work will be supported by a MSc student. The knowledge acquired from these focus groups will inform our promotional strategy and any modifications that need to be made to the Databank, such as refining our Core Data Activities, and will help us ensure that the Data Dashboard supports the overall Databank user experience. To increase engagement and facilitate the flow of discussion in our focus groups, we will segment our potential user population into groups that include graduate students, academic researchers, industry researchers, and public health practitioners. We have allocated funds for conducting up to seven focus groups and will aim to recruit 6 – 12 participants per focus group. When possible, we will hold focus groups at conferences attended by our targeted participants, where we will concurrently promote the Databank. Prior consultation has established that graduate students could greatly benefit from access to data. With input from our focus groups with students, we will create use cases in a variety of fields (epidemiology, statistics, mathematical modelling, immunology) to highlight graduate level projects for which the data are suited. We will present webinars to graduate departments, throughout Canada and internationally, to showcase the available data, explain the application process, and introduce these use cases.

In Summer 2023, the Data Management team promoted the Databank at several North American epidemiologic conferences. However, at that point in time, the Databank had not yet received the data from many of its participating studies, and harmonization of the available data was still in its infancy. Since then, the size of the Databank (both the number of individual studies and the harmonized data) has grown tremendously, and we anticipate this increase will stimulate interest in the Databank. In addition to presenting information about the available data and the application process at future conferences, we will also host virtual workshops and consider other activities to make the data accessible, including hackathons and the creation of synthetic data sets.

Another important medium to promote the use of data will be the CITF Databank website. While the CITF was still active, we worked closely with the CITF Communications team to create the public facing Databank page on the CITF website. This served as an important tool to describe the Databank and its progress and present procedural information and resources. As the official CITF website will no longer be updated, all Databank related content will be transferred to its own new web domain [15]. This site will serve as the “central hub” for the Databank; from here, researchers will be directed to other pages such as the Dashboard, Dataverse and Access Portal. In addition, promotional content or announcements, such as upcoming webinars, presentations, and workshops.

Throughout its mandate, the CITF communications team promoted CITF-funded research through multiple channels, including newsletters, on its website and on social media platforms. CITF-funded researchers were encouraged to notify the communications team of any published results, and automated searches were regularly conducted to keep up to date on publications, which were subsequently summarized by the communications team to highlight the studies’ key findings. We will build on this strategy to promote research and findings from successful Databank applications. Successful data access applicants are listed on the Data Access Portal’s Approved Projects page [16] and will be encouraged to notify the team when they publish their study results. We will then promote lay summaries of these publications on the website, through social media, and other channels so that scientists and policy makers are made aware of the value of secondary research using the Databank. In addition to promoting data users’ research findings, in **Activity 6** we describe how we will document lessons learned about data sharing and management during the development of the Databank and conduct methodological research. We will mobilize this knowledge through professional networks and publications.

Activity 5. Develop and Maintain Data Dashboards to Visualize Data

The CITF Databank Dashboard [17] was developed to provide transparency about the Databank’s progress and to improve public knowledge about the data. As the Databank’s contents are extensive and evolving, it is important to have an online tool that researchers can use to navigate the Databank data and assess its potential and feasibility as a data source.

The Dashboard currently has three tabs. The “Study Overview dashboard” introduces all studies that have been deposited into the Databank, displaying descriptive information and univariate statistics. The “Harmonized data overview dashboard” introduces the harmonized datasets, including harmonization progress and plotted univariate statistics for the total harmonized cohort. The “Harmonized cohort search dashboard” is intended for researchers with a research question in mind who want to assess the feasibility of harmonized datasets as a data source. Researchers can filter the harmonized data by their envisioned study selection criteria and see sample size and univariate statistics for the selected participants. Random rounding is used on all descriptive statistics to protect participants from disclosure [18]. All of the dashboards will be updated as individual studies are deposited to the Databank through **Activity 1**. New features will be added to reflect new harmonized datasets; for example, additional choices and visualizations will be added for follow-up visit variables.

As part of the development process, the Dashboard underwent User Acceptance Testing by targeted users, such as researchers and graduate students, to optimize the interface. The dashboard also accepts ongoing feedback from users through its Feedback tab and associated user satisfaction survey. Quantitative web analytics are collected through Google Analytics to monitor user activity and engagement trends in real time. We will use this qualitative and quantitative feedback, and input from the focus groups described in **Activity 4**, to work in an agile fashion to refine and extend the Dashboard.

Activity 6. Provide Analytic Consulting and Develop Methods

Our research team includes expertise in epidemiology, biomedical informatics, immunology, mathematics, and statistics and our established networks expand upon this expertise. Over the last four years, we have performed a range of analyses that have contributed to COVID-19 public health decision making. Examples include: the “Seroprevalence in Canada” monthly reports that were provided to decision makers in the Government of Canada throughout the pandemic; a scientific manuscript on the evolution of the pandemic [19]; descriptive statistics available through the CITF Databank Dashboard; and targeted analyses to support CITF-sponsored researchers and public health agencies, such as the Chief Public Health Officer for the Northwest Territories. We will continue activities that leverage our experience and expertise to understand the COVID-19 pandemic and pandemic preparedness.

To support researchers and public health practitioners, we will provide *ad hoc* statistical consulting as appropriate for using the Databank microdata or requiring custom data tabulations. We will continue to provide expertise in serosurveillance and immunology to mathematical modellers and biostatistics researchers. For example, health equity, multilevel, or spatial modelling is possible using data from the many studies that included sociodemographic and 3-digit postal code variables for participants, which can be augmented with census-based indicators and epidemiologic indicators from open public health COVID-19 data into or spatial analyses. In all cases, we will share analysis code with researchers and make it openly available through mechanisms such as Borealis and GitHub.

Our research team will also generate knowledge on databank management, serosurveillance methods, and mathematical modelling. For example, we will:

- Document lessons learned over the course of development of this Databank;
- Build on analyses to understand the representativeness of study participants [20];
- With support from a PhD student, exploit the wide range of study designs and serology assays in the Databank to evaluate serosurveillance strategies;
- With support from a PhD student, develop analytical methods in statistics and mathematical modelling (for our datasets) to enable the estimation of population-level immunity from in-host infection and vaccination data; and,
- Extend open data models and standards for wastewater surveillance data (e.g. PHES-ODM [21]) to accommodate serosurveillance data, thereby facilitating the broader use of these data.

The resulting knowledge will be disseminated as described in **Activity 4** and by building on our linkage with the re3data registry of data repositories and collaborating with other open access platforms, such as the Global Dataverse Community Consortium.

Applicant Information

An interdisciplinary approach is needed due to the intersection of immunity, public health surveillance, statistics, mathematical modelling, data governance, and biomedical informatics. Our team brings multiple disciplinary perspectives and diversity in gender, career stage, and ethnicity to this research project.

Role	Name	Affiliation	Expertise and Activities
NPA	David Buckeridge	McGill University, SPGH	Public health and biomedical informatics. Oversight of all grant activities and co-chair of Advisory Board. CITF Databank Custodian.
Co-Applicant	Tim Evans	McGill University, SPGH	Public health and economics. Co-chair Advisory Board
Co-Applicant	Olivia Oxlade	McGill University, SPGH	Epidemiology
Co-Applicant	Bruce Mazer	McGill University, Medicine	Immunology
Co-Applicant	Jane Heffernan	York University, Mathematics and Statistics	Mathematical Modelling
Co-Applicant	David Stephens	McGill University, Mathematics and Statistics	Statistics
Co-Applicant	Alisa Beth Rod	McGill University, Research Data Management	Data management
Co-Applicant	Doug Manuel	University of Ottawa, Medicine	Data standards

References

1. CITF Databank. *CITF: COVID-19 Immunity Task Force*. 2024 [cited 2024 03/28]; Available from: <https://borealisdata.ca/dataverse/CITF>.
2. First Nations Information Governance Centre. *First Nations Information Governance Centre (FNIGC)*. 2024 [cited 2024 03-26]; Available from: <https://fnigc.ca/>.
3. First Nations Information Governance Centre. *The First Nations Principles of OCAP (R)*. 2024 [cited 2024 03/26]; Available from: <https://fnigc.ca/ocap-training/>.
4. Maelstrom Research. *COVID-19 Immunity Task Force Study Catalogue*. 2024 [cited 2024 03/26]; Available from: <https://www.maelstrom-research.org/network/citf>.
5. Bergeron, J., et al., *Fostering population-based cohort data discovery: The Maelstrom Research cataloguing toolkit*. PLoS One, 2018. **13**(7): p. e0200926.
6. Canadian Institute for Health Information, *Measuring Health Inequalities: A Toolkit Area-Level Equity Stratifiers Using PCCF and PCCF+*. 2018, Canadian Institute for Health Information: Ottawa, Canada.
7. Canadian Institute for Health Information. *Equity Stratifiers*. 2022 [cited 2024 03/14]; Available from: <https://www.cihi.ca/en/equity-stratifiers>.
8. Regenstrief Institute Inc., S.I. *Introducing the LOINC Ontology: A LOINC and SNOMED CT Interoperability Solution*. 2024 [cited 2024 03/26]; Available from: <https://loincsnomed.org/>.
9. R Core Team, *R: A language and environment for statistical computing*. 2024, R Foundation for Statistical Computing: Vienna, Austria.
10. Fortier, I., et al., *Maelstrom Research guidelines for rigorous retrospective data harmonization*. Int J Epidemiol, 2017. **46**(1): p. 103-105.
11. Go Fair. *FAIR Principles*. [cited 2024 03/18]; Available from: <https://www.go-fair.org/fair-principles/>.
12. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data, 2016. **3**: p. 160018.
13. re3data.org - Registry of Research Data Repositories. *Covid-19 Immunity Task Force Databank*. 2024 [cited 2024 03/26]; Available from: <https://www.re3data.org/repository/r3d100014306>.
14. CITF. *The Covid-19 Immunity Task Force*. 2021 [cited 2024 03/26]; Available from: <https://www.covid19immunitytaskforce.ca/>.
15. CITF Databank. *CITF Databank / Banque de données du GTIC*. 2024 [cited 2024 03/28]; Available from: <https://citfdatabank.ca/>.
16. CITF Databank. *Approved Projects*. CITF Data Portal 2022 [cited 2024 03/26]; Available from: <https://portal.citf.mcgill.ca/projects>.
17. CITF Databank. *CITF Databank Dashboard*. 2023 [cited 2024 03/26]; Available from: <https://dashboard.citf.mcgill.ca/>.
18. Statistics Canada. *Appendix C: Data quality, confidentiality and random rounding, 2011*. 2011 [cited 2024 03/26]; Available from: <https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/app-ann003-eng.cfm>.
19. Murphy, T.J., et al., *The evolution of SARS-CoV-2 seroprevalence in Canada: a time-series study, 2020-2023*. CMAJ, 2023. **195**(31): p. E1030-E1037.
20. Yu, Y., et al., *Comparing statistical adjustment methods for representative bias across SARS-CoV-2 serosurveillance data sources*. 2023.

21. Public Health Environmental Surveillance (PHES ODM). *The Public Health Environmental Surveillance Open Data Model*. 2024 [cited 2024 03/26]; Available from: <https://www.phes-odm.org>.