



**GROUPE DE TRAVAIL
SUR L'IMMUNITÉ
FACE À LA COVID-19**

Guide abrégé sur le partage de données Études financées

Version 2.0 | 5 octobre 2020

Table des matières

1. CONTEXTE	3
2. CATALOGAGE DES ÉTUDES	4
2.1 Description de l'étude	4
2.2 Consignation des variables	4
2.3 Annotation des variables	4
3. ÉLÉMENTS DE DONNÉES.....	5
4. APPROBATIONS JURIDIQUES ET ÉTHIQUES.....	6
4.1 Consentements et approbations	6
4.2 Anonymisation des données.....	6
5. PARTAGE DE DONNÉES AVEC LE GTIC AUX FINS DE CENTRALISATION.....	7
ANNEXE 1	8

1. Contexte

Il est primordial de comprendre l'épidémiologie de la COVID-19 (causée par le SRAS-CoV-2) pour ralentir la propagation de la maladie, réduire au minimum les décès et générer des données probantes fiables pour l'élaboration de politiques. En réponse à cette crise de santé publique sans précédent, le Groupe de travail sur l'immunité face à la COVID-19 (GTIC) a reçu le mandat de financer des études visant à étudier l'épidémiologie de la COVID-19 au Canada. Les études financées s'appuieront sur divers types de données – entre autres, des données cliniques, sérologiques, administratives, démographiques, socioéconomiques, environnementales et sur les pratiques culturelles et le mode de vie.

Afin de permettre une utilisation et une réutilisation rapides de ces données, le GTIC coordonne la mise en place d'un système permettant de recueillir, consigner et diffuser des données et métadonnées relatives à la COVID-19. Ainsi, les données émanant des études financées seront consignées correctement, interexploitables, complètes, hétérogènes du point de vue méthodologique et de qualité élevée. Le présent guide a pour but de décrire et faciliter le processus de centralisation des données et de partage de données avec le GTIC (figure 1).

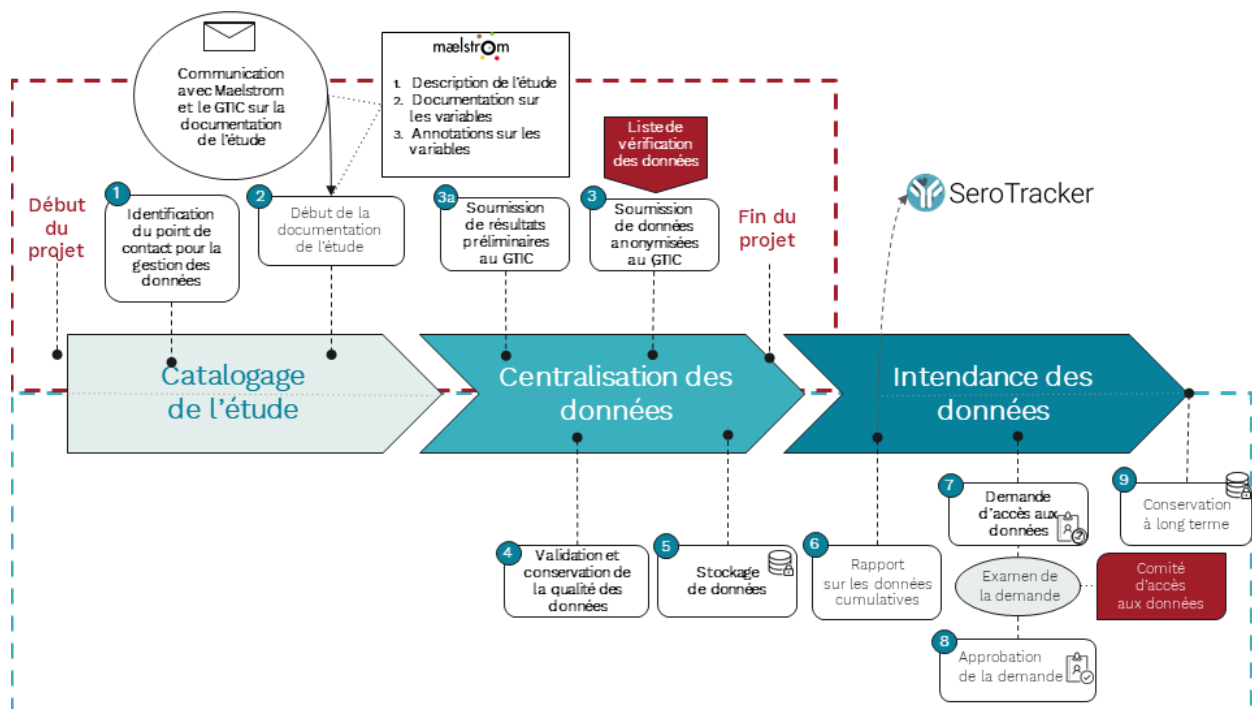


Figure 1. Ce diagramme illustre le cycle de vie général des données produites par les études financées par le GTIC, ce qui comprend le partage de données après la fin du projet. Les activités réalisées par les équipes de recherche sont illustrées dans la moitié supérieure du diagramme (délimitée par le pointillé rouge) et les activités réalisées par le Secrétariat du GTIC sont illustrées dans la moitié inférieure (délimitée par le pointillé turquoise).

2. Catalogage des études

Afin d'optimiser l'utilisation des données produites par les études financées par le GTIC et de promouvoir l'étude scientifique de l'immunité contre la COVID-19 au Canada, le GTIC prévoit de collaborer avec le groupe *Maelstrom Research*, dirigé par la professeure Isabel Fortier, de l'Institut de recherche du Centre universitaire de santé McGill, à la mise en place d'un catalogue de métadonnées portant sur l'ensemble de ces projets de recherche. La trousse de catalogage de *Maelstrom* est largement utilisée par des initiatives d'envergure internationale pour consigner les données d'études de cohortes observationnelles et diffuser leurs métadonnées au sein de la communauté scientifique.

Le processus de catalogage appliqué par l'équipe de *Maelstrom Research* pour consigner chaque étude se divise en trois grandes étapes : la description de l'étude, la consignation des variables et l'annotation des variables. Chacune des étapes est décrite ci-dessous. Les études financées par le GTIC devraient collaborer avec *Maelstrom* pour amorcer le processus de catalogage avant le début de la collecte de données – dans le cas des études déjà en cours, cela pourrait être impossible. Nous recommandons de nommer une personne-ressource qui sera en liaison avec *Maelstrom*.

2.1 Description de l'étude

La première étape consiste à décrire le plan de l'étude, la ou les populations ciblées et les activités de collecte de données. L'équipe de *Maelstrom Research* commencera par décrire l'étude en consultant différentes sources, notamment les articles publiés et le site Web de l'étude, pour ensuite confirmer les renseignements auprès des chercheurs principaux.

2.2 Consignation des variables

Aux premiers stades du projet, il faut sélectionner ou créer des dictionnaires de variables normalisés. L'équipe de *Maelstrom Research* obtiendra les questionnaires et les dictionnaires de données auprès de l'équipe de projet. Chaque variable doit au moins avoir un nom et une étiquette et, le cas échéant, des codes et des étiquettes pour chaque catégorie. Toutes les variables de l'étude (et non seulement éléments de données de base) doivent être cataloguées à mesure qu'elles sont définies.

2.3 Annotation des variables

La dernière étape consiste à catégoriser chaque variable de l'étude dans au moins un des domaines et sous-domaines du système de classification de *Maelstrom Research*. L'équipe de *Maelstrom* associera chaque variable à un ou plusieurs sous-domaines des catégories de données, en s'aidant des questionnaires et des renseignements obtenus aux étapes précédentes du processus de catalogage. L'exactitude de l'association des variables sera vérifiée auprès des chercheurs principaux pour éviter toute erreur.

Les caractéristiques générales de l'étude, les populations ciblées, les activités de collecte de données et les dictionnaires de données seront rendus accessibles sur le site Web de *Maelstrom* et liés aux données de l'étude, qui seront consignées dans la base de données du GTIC une fois qu'elles seront disponibles.

Pour la liste complète des champs de métadonnées de *Maelstrom*, voir l'annexe 1.

Remarque : Les équipes de recherche ayant déjà commencé le processus de description de l'étude en dehors du contexte de leur collaboration avec le GTIC devraient communiquer avec le Secrétariat du GTIC afin de prendre des mesures pour intégrer le mieux possible leurs renseignements actuels dans le processus.

3. Éléments de données

Compte tenu de la nature des données recueillies dans le cadre des études financées par le GTIC, la centralisation peut être considérablement optimisée en utilisant des éléments de données communs (c'est-à-dire des éléments de données de base). Le GTIC a défini ces éléments de données de base qui devraient être saisis par l'ensemble des études et fournis au Secrétariat du GTIC, en plus des résultats d'analyses sérologiques et d'analyses de laboratoire des sujets individuels. Ces éléments de données appartiennent aux domaines généraux suivants, définis par le GTIC : démographie, COVID-19, symptômes, exposition, santé/comportements liés à la santé et sérologie.

Les éléments de données de base ne constituent pas l'ensemble complet des éléments de données que chaque étude saisira. Néanmoins, il s'agit du minimum requis pour permettre au GTIC de fournir régulièrement à l'Agence de la santé publique du Canada les données scientifiques les plus récentes sur les analyses sérologiques et l'évolution des connaissances sur l'immunité contre le SRAS-CoV-2 au Canada.

Pour la liste des éléments de données de base, consulter le [site Web du GTIC](#).

Remarque : Les équipes qui ont commencé leur collecte de données avant la diffusion de ces éléments de données de base devraient communiquer avec le Secrétariat du GTIC afin de prendre des mesures pour harmoniser le mieux possible leurs données avec celles définies par le GTIC.

Domaine	Visite	Numéro	Élément	Question	Type	Carac. techniques
DÉMOGRAPHIE	Début	1	Date de l'entrevue	Date de l'entrevue	Date	JJ/MM/AA
DÉMOGRAPHIE	Début	2	Âge	Quel âge avez-vous?	Numérique	ANS

DÉMOGRAPHIE	Début	3	Sexe	Quel sexe vous a été assigné à la naissance?	Catégorie	MASCULIN/FÉMININ/PRÉFÈRE SE DÉCRIRE/PRÉFÈRE NE PAS RÉPONDRE
DÉMOGRAPHIE	Début	4	Sexe	Quel est votre sexe aujourd'hui?	Catégorie	MASCULIN/FÉMININ/PRÉFÈRE SE DÉCRIRE/PRÉFÈRE NE PAS RÉPONDRE

4. Approbations juridiques et éthiques

Le partage de données réduit le dédoublement du travail, favorise l'efficacité et la transparence et maximise le rendement des fonds octroyés par le GTIC. Toutefois, malgré le besoin impératif d'une diffusion rapide des données relatives à la COVID-19, celle-ci ne doit pas avoir lieu au détriment de la vie privée des individus.

4.1 Consentements et approbations

Afin de s'assurer du respect des lois locales, provinciales et fédérales, le GTIC s'attend à ce que chaque équipe obtienne tous les permis requis pour effectuer ses recherches conformément aux lois provinciales et fédérales, ainsi qu'aux exigences des autorités locales.

Le GTIC offrira de l'encadrement et des « gabarits » aux équipes pour l'obtention de consentements et les soutiendra relativement à l'inclusion de données recueillies antérieurement.

4.2 Anonymisation des données

Toutes les données fournies au GTIC doivent être anonymisées ou, au minimum, dépersonnalisées (en éliminant les renseignements d'identification comme le nom, la date de naissance, etc.).

En ce qui concerne la gestion des données autres que les éléments de données de base qui pourraient poser un risque accru de réidentification, des précautions supplémentaires seront appliquées aux mécanismes de partage de données du GTIC. Les plans de partage de données seront élaborés en fonction de chaque projet, en collaboration avec le Secrétariat du GTIC.

Une confirmation de l'anonymisation des données et des autorisations sera demandée au moment de soumettre les données aux fins de stockage dans la base de données du GTIC.

Remarque : Les équipes ayant besoin de soutien relativement à l'un ou l'autre de ces aspects peuvent demander conseil au Secrétariat du GTIC.

5. Partage de données avec le GTIC aux fins de centralisation

Pour assurer la transmission sécurisée des données des études financées, le GTIC a créé un portail Web sécurisé qui permettra aux équipes de recherche d'accéder des espaces publics.

Le GTIC déposera dans les espaces publics des documents d'orientation et des modèles que les équipes de recherche pourront utiliser de façon générale.

Les espaces privés seront utilisés par chaque équipe pour le téléversement de données et de documents à l'intention du GTIC. Ils serviront également à l'exportation de données, ainsi qu'au téléversement d'ententes et de formulaires signés. Les données doivent pouvoir être exportées dans un format ouvert standard répandu, par exemple le format « csv ».

Le portail de téléversement de données (<https://pydio.mchi.mcgill.ca>) devrait être accessible de n'importe où. Il s'agit d'une plateforme privée, hébergée dans un environnement sécurisé géré par le GTIC à l'Université McGill. Toutes les interactions concernant des données provenant d'études ou destinées à des études passeront par ce portail.

Annexe 1

Champs de métadonnées du catalogue de *Maelstrom Research*.

ÉTUDE	
Champ	Définition
Nom	Nom officiel de l'étude.
Acronyme	Acronyme qui désigne l'étude.
Site Web	Adresse URL du site Web de l'étude.
Chercheurs	Nom, établissement d'attache et coordonnées des chercheurs principaux.
Personne-ressource	Nom, établissement d'attache et coordonnées de la personne à joindre pour obtenir de plus amples renseignements sur l'étude.
Objectifs	Objectifs principaux de l'étude.
Échéancier de l'étude	Date du recrutement des premiers participants et date de fin de l'étude, si celle-ci est terminée.
Plan de l'étude	Renseignements sur la conception de l'étude. <ol style="list-style-type: none">1. Étude de cohorte2. Étude cas-témoin3. Étude de cas4. Étude transversale5. Essai clinique6. Autre
Renseignements généraux sur le suivi	Profil et fréquence du suivi des participants (p. ex., les participants sont suivis aux 5 ans).
Renseignements supplémentaires sur le plan de l'étude	Renseignements additionnels sur le plan de l'étude (p. ex., suréchantillonnage intentionnel d'une sous-population).
Cible de recrutement	Type d'unités de participants ciblés par l'étude. <ol style="list-style-type: none">1. Individus2. Familles3. Autre
Nombre de participants	Nombre de participants à recruter. Si l'étude est terminée, nombre total de participants recrutés.
Nombre de participants avec échantillons biologiques	Si l'étude comprend le prélèvement d'échantillons biologiques, nombre de participants qui devraient fournir des échantillons. Si l'étude est terminée, nombre total de participants qui ont fourni des échantillons.

Renseignements supplémentaires sur le nombre de participants	Renseignements supplémentaires sur le nombre de participants visé (p. ex., pour la population 2 nous préleverons des échantillons additionnels).
Accès	Accès permis ou prévu, ou non, aux données de l'étude, aux échantillons biologiques ou à d'autres éléments de l'étude, par des chercheurs externes ou des tiers.
Marqueur(s) de publication	Référence(s) bibliographique(s) à utiliser pour citer l'étude et, le cas échéant, identifiant Pubmed ID de la publication.
Logo	Logo de l'étude.
Documents	Documents pertinents se rapportant à l'étude (p. ex., questionnaires, procédures opérationnelles normalisées ou listes de codage).

POPULATION	
Champ	Définition
Nom	Nom de la population étudiée.
Description	Courte description de la population.
Sources de recrutement	Caractéristiques des sources de recrutement. 1. Population générale (inscription volontaire, sélection d'un échantillon, composition aléatoire) 2. Population définie (patients de cliniques, membres d'une association, etc.) 3. Participants d'autres études 4. Autre source
Renseignements supplémentaires sur les sources de recrutement	Renseignements additionnels sur les procédures de recrutement (p. ex., les participants ont été recrutés à partir de listes électorales et de listes de patients en médecine générale).
Critères de sélection	Critères de sélection appliqués aux participants parmi les suivants, dans la mesure il est pertinent de les indiquer. 5. Genre (femmes ou hommes) 6. Âge (âge minimal et âge maximal) 7. Lieu de résidence (pays, territoire ou ville) 8. Femmes enceintes (premier, deuxième ou troisième trimestre) 9. Nouveau-nés 10. Jumeaux 11. Origine ethnique 12. État de santé 13. Autre
Renseignements supplémentaires sur les critères de sélection	Renseignements additionnels sur les critères appliqués à la population (p. ex., tous les sujets présentant au début de l'étude un déficit cognitif sans démence étaient admissibles à la phase longitudinale qui avait lieu après une année).

Nombre de participants	Nombre de participants à recruter pour former la population. Si l'étude est terminée, nombre total de participants recrutés.
Nombre de participants avec échantillons biologiques	Si l'étude comprend le prélèvement d'échantillons biologiques, nombre de participants qui devraient fournir des échantillons pour la population. Si l'étude est terminée, nombre total de participants qui ont fourni des échantillons.
Renseignements supplémentaires sur le nombre de participants	Renseignements supplémentaires sur le nombre de participants. Habituellement, nombre de participants à chaque édition de l'étude (p. ex., nombre de participants pour chaque activité de collecte de données : Édition 1 : 7 175 participants Édition 2 : 3 145 participants Édition 3 : 1 733 participants).

ACTIVITÉ DE COLLECTE DE DONNÉES	
Champ	Définition
Nom	Nom de l'activité de collecte de données.
Description	Courte description de l'activité de collecte de données.
Date de l'activité de collecte de données	Dates de début et de fin de la collecte de données.
Sources de données	Sources utilisées pour la collecte de renseignements. 14. Questionnaires 15. Mesures physiques 16. Mesures cognitives 17. Échantillons biologiques (sang, sang de cordon, cellules buccales, tissus, salive, urine, cheveux, ongles et autres) 18. Bases de données administratives (bases de données en santé, statistiques de l'état civil, bases de données socioéconomiques, bases de données environnementales) 19. Autres (p. ex., dossiers médicaux)

ENSEMBLE DE DONNÉES	
Champ	Définition
Nom	Nom de l'ensemble de données.
Acronyme	Acronyme de l'ensemble de données.
Description	Courte description de l'ensemble de données qui précise son contenu.
Type d'entité	Ce que décrivent les données (habituellement, le participant).