

# MDClone Synthetic data access and User acknowledgement form

MUHC – McGill University Health Centre

Digital Health Hackathon – Réseau Santé Numérique

## Background

The McGill University Health Centre (MUHC), in partnership with MDClone and the Digital Health Network (Réseau Santé Numérique - RSN), provides synthetic clinical datasets designed to emulate the structure and statistical characteristics of real MUHC patient data. These datasets serve as secure, ethical, and realistic resources for **training**, without involving any real patient information.

## Purpose

To provide hackathon participants with hands-on experience using **synthetic clinical data** that simulate real-world health information while fully protecting privacy and confidentiality. The use of these datasets aims to foster learning in data-driven health innovation, analytics, and responsible numerical tools development.

## Availability and Access

- The datasets will be made available to registered hackathon participants via a **secure OneDrive environment** managed under the Université de Montréal institutional account, or the duration of the event.
- Access credentials are personal and non-transferable. **Sharing of files is strictly prohibited.**
- The data may only be accessed and processed within the event's approved environment. Participants may download a **local copy** during the hackathon for analysis, provided that all copies are permanently deleted after the event.

## Content

The synthetic datasets are generated by **MDClone** based on MUHC clinical data models. Depending on the research theme, each dataset may include simulated variables such as:

- Demographics (age, sex, postal code, language)
- Clinical variables (diagnoses, laboratory values, medications)
- Hospital visits (admissions, readmissions, emergency visits)
- Outcomes (length of stay, discharge status)

All records are entirely synthetic and contain **no identifiable or real patient data**.

## Usage

- The dataset is provided **exclusively for educational, training, and innovation purposes** during the Digital Health Hackathon organized by the RSN and other partners.
- The data and outputs **cannot** be used for formal research, publication, or dissemination outside the hackathon.
- Participants may share **their analysis scripts or code** publicly (e.g., GitHub), but not the dataset itself.
- Participants must uphold principles of **ethical digital health innovation** aligned with the RSN Charter: openness, equity, trust, sustainability, and citizen engagement.

### Ownership and Intellectual Property

- The synthetic datasets remain the property of MUHC.
- Users do not acquire any intellectual property rights over the datasets or derived results.

### Liability

- The MUHC and MDClone make no warranties regarding the accuracy or completeness of the synthetic data.
- The use of these datasets is at the participant's own risk. The MUHC, MDClone, RSN, and their affiliates shall not be liable for any damages resulting from their use or misuse.

### Data retention and destruction

- Participants must **cease all use of the dataset** immediately upon completion of the hackathon.
- All local copies or temporary files must be deleted, and no portion of the dataset may be retained, shared, or reproduced in any form, after the event.
- MUHC and RSN reserve the right to verify compliance and revoke access in case of misuse.

### Acknowledgement

By signing this acknowledgment, the participant agrees to:

1. Use the synthetic data only within the approved scope of the hackathon.
2. Respect confidentiality and data governance standards applicable to MUHC and the RSN.
3. Adhere to the principles of **responsible innovation, FAIR data, and science integrity** promoted by the RSN.

I have read, understood, and agree to comply with the terms and conditions set out in this MDClone Synthetic Data User Acknowledgement.

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

# Formulaire d'engagement pour l'utilisation des données synthétiques MDClone

CUSM – Centre universitaire de santé McGill

Hackathon en santé numérique – Réseau Santé Numérique

## Contexte

Le Centre universitaire de santé McGill (CUSM), en collaboration avec **MDClone** et le **Réseau Santé Numérique (RSN)**, met à disposition des jeux de données cliniques **synthétiques** reproduisant la structure et les caractéristiques statistiques de données réelles du CUSM.

Ces jeux de données constituent des ressources sécurisées et éthiques pour **la formation**, sans contenir de données réelles ni identifiables.

## Objectif

Permettre aux participant·es du hackathon d'expérimenter l'utilisation de **données cliniques simulées** afin de favoriser l'apprentissage, l'analyse et le développement d'outils numériques en santé, dans un cadre sécurisé et respectueux de la confidentialité.

## Disponibilité et accès

- Les jeux de données seront accessibles uniquement aux participant·es inscrit·es via un **environnement OneDrive sécurisé** géré par l'Université de Montréal pendant toute la durée de l'événement.
- Les identifiants d'accès sont personnels et ne doivent pas être partagés.
- Les participant·es peuvent télécharger une copie locale pour leurs analyses pendant le hackathon, à condition de supprimer toutes les copies à la fin de l'événement.

## Contenu

Les données synthétiques sont générées par **MDClone** à partir de modèles cliniques du CUSM.

Elles peuvent inclure, selon les thématiques :

- Variables démographiques : âge, sexe, code postal, langue
- Variables cliniques : diagnostics, résultats de laboratoire, médicaments
- Données hospitalières : admissions, réadmissions, urgences
- Résultats : durée du séjour, statut de sortie

Toutes les données sont **entièrement fictives** et ne comportent aucune information identifiable.

## Utilisation

- Les données sont destinées **exclusivement à des fins pédagogiques** dans le cadre du Hackathon en santé numérique.
- Elles **ne peuvent être utilisées à des fins de recherche, publication ou diffusion académique**.
- Les participant·es peuvent partager leur code et leurs résultats analytiques, mais **jamais les données elles-mêmes**.
- Les participant·es s'engagent à respecter les **principes du RSN** : transparence, équité, science ouverte, confiance numérique et durabilité.

### Propriété intellectuelle

- Les données demeurent la propriété du CUSM.
- Les utilisateur·rices ne détiennent aucun droit de propriété intellectuelle sur les données ni sur les résultats qui en découlent.

### Responsabilité

Le CUSM et MDClone ne garantissent ni l'exactitude ni la pertinence des données synthétiques.

L'utilisation de ces jeux de données se fait **aux risques de l'utilisateur·rice**, sans responsabilité du CUSM, de MDClone ou du RSN en cas de perte, dommage ou usage inapproprié.

### Conservation et destruction

- À la fin du hackathon, les participant·es doivent **cesser toute utilisation et supprimer toutes les copies locales**.
- Aucune portion du jeu de données ne doit être conservée, transférée ou reproduite.
- Le CUSM et le RSN se réservent le droit de vérifier le respect de ces conditions et de **retirer l'accès** en cas de non-conformité.

### Engagement

En signant ce formulaire, l'utilisateur·rice s'engage à :

1. Utiliser les données uniquement dans le cadre autorisé du hackathon.
2. Respecter la confidentialité et les politiques de gouvernance des données du CUSM et du RSN.
3. Adhérer aux principes d'innovation responsable et d'intégrité scientifique promus par le RSN.

J'ai lu, compris et j'accepte de me conformer aux conditions énoncées dans le présent formulaire d'engagement relatif aux données synthétiques de MDClone.

Nom : \_\_\_\_\_

Signature : \_\_\_\_\_

Date : \_\_\_\_\_