

# Modèle de plan de gestion des données : Centres de données de recherche et analyses externes (RCCDR)

## Résumé

Le modèle du Réseau canadien des Centres de données de recherche (RCCDR) résume la gestion de données effectuée par Statistique Canada et le RCCDR au nom des chercheurs. Bien que la gestion de données dans un centre de données de recherche (CDR) a des avantages, elle a aussi des inconvénients substantiels: les données d'un CDR ne pourront jamais être versées dans un dépôt conformément aux meilleures pratiques recommandées pour la gestion des données de recherche. Étant donné cette lacune, les chercheurs doivent connaître d'autres options pour s'engager aux meilleures pratiques. En plus de s'assurer que le dossier de projet du CDR est bien documenté et conforme aux résultats de la recherche, les chercheurs doivent conserver leurs données justificatives dans un dépôt reconnu dans leur discipline ou dans le Dépôt fédéré de données de recherche (DFDR) contiendront les métadonnées, la syntaxe (code qui produit une sortie statistique) et tous les documents de soutien pour le projet de recherche.

Ce modèle s'adresse aux chercheurs qui effectuent des travaux dans des CDR en utilisant les données de Statistique Canada et les données de recherche qu'ils ont apportées dans les « données supplémentaires » des CDR ou qu'ils analysent parallèlement à leurs travaux dans les CDR (comme les méthodes mixtes) ou aux statistiques à usage public qui complètent les travaux des CDR (désignées comme *données externes* dans le présent document). Les chercheurs doivent savoir que toutes les données introduites dans les CDR seront stockées avec le reste des documents relatifs à leur projet, sous réserve des protocoles de gestion de l'information de Statistique Canada. Il s'agit d'un processus gratuit et relativement simple ; les chercheurs peuvent obtenir plus d'informations en s'adressant à leur analyste de CDR.

Si votre travail est effectué dans le CDR en utilisant uniquement les données fournies par le programme du CDR, utilisez le modèle du CDR uniquement plutôt que ce modèle.

## Détails administratifs

**Auteur (s) du modèle :** Grant Gibson, Réseau canadien des Centres de données de recherche (RCCDR)

**Publié :** 9 avril 2021

**DOI :** [10.5281/zenodo.4667841](https://doi.org/10.5281/zenodo.4667841)

**Contact :** Le Réseau Portage - [portage@engagedri.ca](mailto:portage@engagedri.ca), [reseauportage.ca](http://reseauportage.ca)

*Modèle de plan de gestion des données :*  
**Centres de données de recherche et analyses externes (RCCDR)**

**Licence :** Attribution - Pas d'Utilisation Commerciale 4.0 International (CC BY-NC 4.0)



**Version :**

Version	Date	Changements
1.0	2021-04-09	Préparé pour la publication inaugurale.

## Collecte de données

Toutes les recherches effectuées dans des Centres de données de recherche (désignées comme CDR dans le présent document) sont secondaires de nature. Il n'y a pas de collecte de données dans cette partie du projet. Ces données appartiennent à Statistique Canada qui est aussi responsable de leur maintien; le stockage de ces données et l'accès à celles-ci sont offerts par le Réseau canadien des Centres de données de recherche.

Les données brutes dans le CDR sont stockées dans de multiples formats, notamment : .SAS (SAS), .dta (STATA), et .shp (fichiers de forme) selon le cas. La disponibilité du logiciel StatTransfer™ dans les CDR et la gestion continue par Statistique Canada garantiront que les données seront accessibles indéfiniment si les formats de fichiers actuellement utilisés deviennent obsolètes. Les chercheurs peuvent ajouter des données dans les CDR (celles-ci seront appelées « données supplémentaires »). Ainsi, ces données sont stockées avec tous les autres produits de recherche du CDR liés à ce contrat et elles sont archivées.

### Quels jeux de données des CDR seront utilisés dans la recherche?

Le numéro d'enregistrement est disponible sur le site web de Statistique Canada; on peut y accéder directement ou via notre site web [crdcn.org/fr/données](http://crdcn.org/fr/données). Par exemple : Enquête auprès des peuples autochtones 2017, Numéro d'enregistrement : 3250 [https://www23.statcan.gc.ca/imdb/p2SV\\_f.pl?Function=getSurvey&SDDS=3250](https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3250).

**Exemple** : Les sources de données pour ce projet sont <<INSCRIRE LE NOM DES ÉTUDES/DOSSIERS ADMINISTRATIFS APPROUVÉS>>. Les versions actuelles sont : <<Numéro d'enregistrement>>.

### Veillez décrire le processus de collecte des données supplémentaires ou externes qui feront partie de votre projet.

Les données externes ou supplémentaires sont les données utilisées pour votre projet de recherche qui ne vous sont pas fournies par Statistique Canada dans le cadre du programme des centres de données de recherche.

**Dans quels formats de fichier les données supplémentaires seront-elles collectées et traitées? Ces formats permettront-ils le partage des données et leur accès à long terme? Comment allez-vous structurer et nommer ces fichiers et gérer leurs versions de manière à ce qu'ils soient facilement compréhensibles par les autres?**

## Documentation et métadonnées

*La documentation fournie par Statistique Canada dans le CDR sera mise à la disposition de tout futur utilisateur potentiel de ces données. Cette documentation est rendue disponible gratuitement aux chercheurs dont les projets ont été approuvés, et contient des informations sur le processus de sélection de l'échantillon, une copie du questionnaire et un guide de codification. Les chercheurs devraient également réfléchir à la manière dont les métadonnées de leurs données externes peuvent être fournies à d'autres chercheurs. Selon les meilleures pratiques, il faut coordonner la gestion des données internes et externes; la meilleure façon de le faire dépend de la nature des données externes.*

**Que ferez-vous pour vous assurer que vos contributions de données de recherche (syntaxe, résultats, etc.) dans votre dossier de projet du CDR et votre analyse externe (le cas échéant) sont correctement documentées, organisées et accessibles?**

Des ressources sont disponibles sur le site web du RCCDR pour vous aider. Vous trouverez ici une recommandation du RCCDR sur la manière de documenter vos contributions à la recherche. Pour des idées sur la façon de conserver des recherches reproductibles, vous pouvez consulter le site suivant :

<https://labordynamicsinstitute.github.io/replication-tutorial-2019/#/> (lien en anglais).

**Comment ferez-vous en sorte que la syntaxe archivée dans votre dossier de projet (et celle créée pour votre analyse externe, le cas échéant) est créée de manière cohérente tout au long de votre projet?**

La syntaxe est définie comme tout code utilisé par un chercheur pour transformer les données brutes en résultats de recherche. Généralement, elle est sous forme de fichiers .do (Stata), de fichiers .sas (SAS) et de code .r (R).

**Veillez indiquer ici les informations concernant la disponibilité des métadonnées pour votre projet (les données des CDR et vos données externes). Certaines métadonnées pour les jeux de données des CDR sont disponibles en contactant l'analyste des CDR.**

**Comment allez-vous vous assurer que les données externes/supplémentaires sont facilement compréhensibles et correctement documentées (y compris les métadonnées)?**

Pour une bonne ressource de départ sur les métadonnées, consultez le lien suivant :  
<https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/> (lien en anglais).

## **Stockage et sauvegarde**

*Le stockage des données est géré par le RCCDR en partenariat avec Statistique Canada sur des serveurs situés dans le réseau. La politique actuelle du RCCDR est de stocker les données des projets (syntaxe, versions et tout autre élément stocké dans le dossier du projet) pendant dix ans. Ces données sont sauvegardées sur place et accessibles par un réseau hautement sécurisé à partir de n'importe quel autre site du RDC. Les données brutes liées au projet de recherche sont stockées à perpétuité par Statistique Canada.*

*Pour les données de recherche externes, le stockage et la sauvegarde sont de la responsabilité du chercheur exclusivement. Veuillez tenir compte des questions suivantes en ce qui concerne les données externes. Ces questions doivent également être prises en compte pour les données supplémentaires si vous envisagez de faire un stockage et une sauvegarde parallèles de ces données.*

**Quels sont les besoins de stockage prévus pour votre projet, en termes d'espace de stockage (en mégaoctets, gigaoctets, téraoctets, etc.) et de durée de stockage?**

Étant donné la structure des accords en vertu desquels les données supplémentaires sont introduites dans le CDR, nous recommandons fortement un stockage et une sauvegarde parallèles pour simplifier le partage de ces données de recherche. Notez que le terme « données » comprend les données brutes et toutes les données générées au cours de la recherche.

**De quelle manière et à quel endroit vos données seront-elles stockées et sauvegardées pendant votre projet de recherche?**

**Comment l'équipe de recherche et les autres collaborateurs vont-ils accéder aux données, les modifier et y contribuer tout au long du projet?**

## Préservation

*Le travail effectué dans le CDR pour ce projet est conservé sur la base du numéro de contrat fourni par le programme du CDR qui peut être utilisé par n'importe quel membre de l'équipe du projet pour récupérer le code et les documents justificatifs pendant une période de 10 ans, selon les indications dans la partie « Stockage et sauvegarde ». Les données brutes qui sont la propriété de Statistique Canada, c'est-à-dire les données des CDR, sont stockées de façon permanente par Statistique Canada, mais ne peuvent jamais être données aux chercheurs. Les chercheurs peuvent également conserver toutes les données de recherche des CDR générées par les utilisateurs qui répondent aux critères de diffusion par le biais d'une demande de vérification via un dépôt tel que le DFDR (encore ici, les données brutes des CDR ne peuvent pas être partagées). Les meilleures pratiques pour les travaux reproductibles exigent la préservation à long terme des données de recherche (donc, dans le cas de la recherche en CDR, cela signifie les métadonnées, la syntaxe, la méthodologie). En plus de cette préservation pour les travaux des CDR, les données externes (et la syntaxe, les métadonnées et la méthodologie qui s'y rapportent) doivent également être préservées.*

**Déposerez-vous vos données de syntaxes et autres données de recherche dans un dépôt pour préserver vos fichiers? Veuillez décrire ici votre intention de préserver toutes les données de recherche, en précisant comment vous traiterez les enjeux de confidentialité liés à vos données supplémentaires/externes.**

## Partage et réutilisation

*Étant donné que les fichiers de microdonnées de Statistique Canada sont recueillis avec une garantie de leur confidentialité et qu'ils sont détenus et contrôlés par Statistique Canada, ils ne peuvent être partagés par aucun membre de l'équipe de recherche.*

*L'accès aux données dans les CDR est régi par la politique d'accès et de frais de service du RCCDR en anglais ou en français. Cette politique offre un accès gratuit aux chercheurs universitaires qui sont membres du réseau et donne accès aux autres sur la base du recouvrement des coûts.*

*Le RCCDR et Statistique Canada font la promotion de leurs fonds de données par le biais des médias sociaux et de leurs sites web respectifs. De plus, les données du RCCDR doivent être citées dans toutes les publications avec le numéro d'enregistrement afin que les lecteurs puissent trouver les données. Par ailleurs, toutes les publications utilisant les données du RCCDR doivent inclure l'identifiant du contrat du RCCDR afin que les utilisateurs potentiels puissent trouver des informations sur le contrat original. Ces informations sont disponibles sur le site web du RCCDR ([crdcn.org/fr/publications](http://crdcn.org/fr/publications)). Pour vos données supplémentaires/externes, veuillez répondre aux questions suivantes afin de satisfaire aux principes FAIR (lien en anglais).*

**En dehors du partage et de la réutilisation des données qui se fait automatiquement dans le dossier de votre projet, quelles données allez-vous partager, où les partagerez-vous et sous quelle forme les partagerez-vous (par exemple, données brutes, traitées, analysées, définitives)?**

Considérez également le format de fichier que vous utiliserez. Ce format de fichier sera-t-il utilisable à l'avenir ? Est-il propriétaire ?

**Quel type de licence d'utilisateur final sera attribué à ces données partagées?**

**Quelles mesures allez-vous prendre pour faire savoir à la communauté des chercheurs que ces données existent?**

## **Responsabilités et ressources**

*Le RCCDR et Statistique Canada conserveront les données de recherche même si le chercheur quitte leur organisation.*

*Le RCCDR bénéficie du soutien des IRSC, du CRSH et de la FCI ainsi que des fonds des universités partenaires. Les utilisateurs des CDR n'ont aucuns frais à payer pour la gestion des données effectuée sous les auspices du RCCDR et de Statistique Canada, comme décrit dans le présent PGD.*

*Le RCCDR n'utilise pas de contrôle de cohérence pour s'assurer que le code fourni avec les demandes de communication des résultats de recherche à partir de l'installation sécurisée crée réellement le résultat demandé. La responsabilité de s'assurer que le code et les documents décrivant leur utilisation fonctionnent comme prévu et sont clairs pour les autres utilisateurs qui pourraient y accéder incombe aux chercheurs du CDR. Le RCCDR dispose d'un mécanisme permettant de garantir que le code est sauvegardé avec tous les résultats de recherche utilisés pour étayer les conclusions de tout travail publié.*

*Les chercheurs doivent réfléchir à la manière de gérer leurs données de recherche externes et doivent se demander qui, au sein de l'équipe de projet, sera responsable de la gestion des données de recherche et quelles ressources pourraient être nécessaires pour ce faire. Dans la mesure du possible, les données de recherche provenant du CDR doivent être gérées de manière coordonnée avec la gestion des données de recherche externes.*

*Modèle de plan de gestion des données :*  
**Centres de données de recherche et analyses externes (RCCDR)**

*En plus de la gestion des données employée par Statistique Canada, il est possible pour les chercheurs de faire sortir du CDR des résultats de recherche qui ne contiennent pas de données confidentielles, notamment des tableaux, de la syntaxe et d'autres informations qui pourraient être conservés dans un dépôt au choix du chercheur, comme décrit dans la section *Préservation*. Si vous envisagez de procéder à un stockage ou à une préservation supplémentaire de vos données de recherche (soit les données de recherche générées par les utilisateurs du CDR, soit les données externes/supplémentaires), veuillez indiquer à qui incombe la responsabilité de la préservation et de la gestion de ces archives.*

**Pour les données externes/supplémentaires, précisez qui sera responsable de la gestion des données de ce projet pendant et après le projet et les principales tâches de gestion des données dont cette personne sera responsable.**

**Pour les données externes/supplémentaires, comment seront traitées les responsabilités en matière de gestion des activités liées aux données si des changements importants interviennent au sein du personnel chargé de superviser les données du projet, notamment un changement de chercheur principal?**

**Pour les données externes/supplémentaires, de quelles ressources aurez-vous besoin pour mettre en œuvre votre plan de gestion des données pour toutes vos données de recherche (c'est-à-dire les données des CDR et les données externes/supplémentaires)? Quel est le coût approximatif global de la gestion des données?**

L'outil fourni suivant fourni par OpenAIRE peut aider les chercheurs à estimer le coût de la gestion des données de recherche

<https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs> (*lien en anglais*).

## **Conformité éthique et juridique**

*Tous les utilisateurs du CDR doivent être des « employés réputés » de Statistique Canada. Pour devenir un employé réputé, le Conseil du Trésor impose un processus de vérification de la sécurité comprenant une vérification des antécédents criminels, une vérification de la solvabilité et la prise d'empreintes digitales. L'approbation de l'accès aux données nécessite un processus d'examen par les pairs d'une proposition de recherche et une revue institutionnelle à Statistique Canada. Dans les cas où les travaux d'un chercheur ont été évalués dans le cadre du processus d'examen de la titularisation, ils sont considérés comme préapprouvés par les pairs et seul l'examen institutionnel est nécessaire.*



*Modèle de plan de gestion des données :*  
**Centres de données de recherche et analyses externes (RCCDR)**

*Une fois qu'un chercheur est autorisé à accéder au CDR, il doit prêter un serment de confidentialité - en promettant de ne jamais divulguer de données confidentielles. Des sanctions pénales peuvent être appliquées en vertu de la loi sur les statistiques en cas de violation de ce serment.*

*La propriété intellectuelle des travaux effectués au sein des CDR devient la propriété de Statistique Canada, y compris le code utilisé pour manipuler les données. La collecte et la diffusion de microdonnées confidentielles, ainsi que l'accès à celles-ci, sont effectués en vertu de la Loi sur la statistique et sont conformes à toutes les exigences légales. Les microdonnées confidentielles pour ce projet ne peuvent être partagées, affichées ou copiées. L'accès aux données est disponible exclusivement par l'intermédiaire de Statistique Canada et du programme de CDR. Plus d'informations sur la façon d'accéder aux données sont disponibles ici en [anglais](#) ou en [français](#).*

*En général, l'autorisation d'éthique de la recherche n'est pas requise pour les recherches menées dans les CDR. Une déclaration du RCCDR sur le sujet est disponible ici en [anglais](#) ou en [français](#).*

*Veuillez répondre aux questions suivantes sur la conformité éthique en ce qui concerne vos données externes/supplémentaires. Si votre projet a fait l'objet d'un examen du Comité d'éthique de la recherche de votre établissement, vous pouvez résumer la soumission au lieu de répondre à ces questions.*

**Si votre projet de recherche comporte des données sensibles, comment ferez-vous en sorte qu'elles soient gérées de manière sécurisée et qu'elles soient uniquement accessibles aux membres approuvés du projet?**

Examinez où, comment et à qui les données sensibles dont la valeur à long terme est reconnue doivent être rendues disponibles, et combien de temps elles doivent être archivées. Les décisions doivent être conformes aux exigences du comité d'éthique de la recherche. Les méthodes utilisées pour partager les données dépendront du type, de la taille, de la complexité et du degré de sensibilité des données. Décrivez les problèmes prévus liés au partage des données, ainsi que leurs causes et les mesures possibles pour les atténuer. Les problèmes peuvent comprendre, entre autres, la confidentialité, l'absence d'ententes de consentement ou les inquiétudes concernant les droits de propriété intellectuelle.

Repris du : Digital Curation Centre (2013). [Liste de contrôle pour un plan de gestion de données](#). v.4.0. Des restrictions peuvent être imposées en limitant l'accès physique aux dispositifs de stockage, en sauvegardant les données sur des ordinateurs n'ayant pas accès à l'internet ou en protégeant les fichiers par un mot de passe et en cryptant ces derniers. Les données sensibles ne doivent jamais être partagées par courrier électronique ou par des services de stockage infonuagiques tels que Dropbox.

## **Quelles stratégies allez-vous mettre en œuvre pour traiter les utilisations secondaires des données sensibles, le cas échéant?**

L'obtention du consentement approprié des participants à la recherche est une étape importante pour garantir aux comités d'éthique de la recherche que les données peuvent être partagées avec des chercheurs en dehors de votre projet. La déclaration de consentement peut identifier certaines conditions clarifiant les utilisations des données par d'autres chercheurs. Par exemple, elle peut stipuler que les données ne seront partagées qu'à des fins de recherche sans but lucratif ou que les données ne seront pas liées à des données personnelles identificatoires provenant d'autres sources. En savoir plus sur la sécurité des données, consultez le site suivant : le [UK Data Service](#) (*lien en anglais*).

## **Comment allez-vous gérer les questions juridiques, éthiques et de propriété intellectuelle?**

Le respect de la législation relative à la protection de la vie privée et des lois susceptibles d'imposer des restrictions sur le contenu des données doit être discuté avec le responsable de la protection de la vie privée ou le bureau des services de recherche de votre institution. Les comités d'éthique de la recherche sont au cœur du processus de recherche.

Donnez ici une description de la propriété, des licences et des droits de propriété intellectuelle des données. Les conditions de réutilisation doivent être clairement énoncées, conformément aux exigences juridiques et éthiques applicables le cas échéant (par exemple, consentement du sujet, autorisations, restrictions, etc.).

**Si vous estimez qu'il y a d'autres exigences juridiques ou éthiques pour votre projet, veuillez les décrire ici.**