

Dépôts recommandés pour les données de recherche sur la COVID-19

Préparé par le groupe de travail sur la COVID-19 du réseau Portage pour le compte de l'Association des bibliothèques de recherche du Canada (ABRC)

Roger Reka (Université de Windsor)
Minglu Wang (Université York)
Erin Clary (Portage)
Beth Knazook (Portage)
Qian Zhang (Université de Waterloo)
Kristi Thompson (Université Western)
Lucia Costanzo (Université de Guelph)
Kathy Szigeti (Université de Waterloo)
Rebecca Dickson (COPPUL)
Mark Leggott (DRC-RDC)
Felicity Tayler (Université d'Ottawa)
Chantal Ripp (Université d'Ottawa)
Jane Fry (Université Carleton)
Nick Rochlin (Université de la Colombie-Britannique)
Melanie Parlette-Stewart (Portage)

SEPTEMBRE 2020

Réseau Portage
Association des bibliothèques de recherche du Canada
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Table des matières

Choisir entre un dépôt disciplinaire ou général.....	2
Option 1 : Dépôt de données disciplinaire	2
<i>Dépôts disciplinaires recommandés</i>	3
Option 2 : Dépôt de données de recherche canadien ou international général	5
<i>Dépôts généraux recommandés</i>	5
Option 3 : Choisir son propre dépôt	6
Survol des caractéristiques importantes des dépôts	7
<i>Quelques éléments à considérer</i>	7
Dépôts.....	9
Références	9

Dépôts recommandés pour les données de recherche sur la COVID-19

La communication des données de recherche impose une nouvelle recherche et façonne les politiques et la prise de décisions. En temps de pandémie, l'accès rapide aux données les plus récentes éclaire les réponses des responsables de la santé publique et aide à sauver des vies. Préserver ces données en vue de leur utilisation future contribue aussi à clarifier les effets économiques, politiques et culturels à long terme de la santé publique et des décisions stratégiques, et à mieux se préparer en vue d'autres éventuelles pandémies. Le présent document vous aidera à choisir un dépôt qui vous donnera un accès immédiat et à long terme à vos données sur la COVID-19.

Choisir entre un dépôt disciplinaire ou général

Option 1 : Dépôt de données disciplinaire

Les données de recherche devraient être déposées dans un dépôt qui est reconnu par les autres chercheurs de votre communauté de pratique. Si un tel dépôt n'existe pas ou si aucun ne convient, vous pouvez déposer vos données de recherche dans un dépôt général (voir l'option 2).

Les avantages d'un dépôt disciplinaire sont les suivants :

- vos données de recherche sont plus visibles pour les collègues qui entreprennent le même genre de recherche ;
- le dépôt peut être conçu de manière à mieux accepter votre type de données ;
- de nombreux dépôts de données de recherche sont des collections de curation des données de recherche sur la COVID 19 pour améliorer la divulgation de ce type de données.

Au moment de choisir un dépôt de données de recherche disciplinaire, attardez-vous sur :

- les suggestions ou les exigences formulées dans les directives des bailleurs de fonds, p. ex. [IRSC](#)¹.
- les listes des dépôts de données disciplinaires recommandés qui sont publiées dans diverses revues, dont [Scientific Data](#) et [PLOS ONE](#).²
- la popularité au sein de la communauté (quels sont les dépôts largement utilisés dans votre domaine ? Ce sont ceux qui sont susceptibles d'être consultés par vos pairs).

¹ "CIHR Open Access Policy: Annex," Canadian Institutes of Health Research, dernière modification le 15 mars 2015, <https://cihr-irsc.gc.ca/e/46068.html#8>.

² Scientific Data, "Recommended Data Repositories," Nature Research, consulté le 4 août 2020,, <https://www.nature.com/sdata/policies/repositories>; PLOS ONE, "Recommended Repositories." PLOS, consulté le 4 août 2020, <https://journals.plos.org/plosone/s/recommended-repositories>.

Dépôts disciplinaires recommandés

Voici la liste retenue des dépôts qui conviennent pour la communication des données sur la COVID-19. Elle est loin d'être complète. Vous trouverez aussi d'excellentes recommandations dans les [recommandations et directives de la RDA en ce qui a trait à la COVID-19](#).³

- Santé publique et épidémiologie
 - [Dépôt de données sur la COVID-19](#) d'OpenICPSR
 - Le risque de divulgation rattaché aux ensembles de données [déposés](#) dans OpenICPSR ne fait pas l'objet d'un examen, mais, en répondant à une série de questions, vous pourrez déterminer si les données peuvent être rendues publiques. Les données qui contiennent des renseignements sensibles ou qui permettent d'établir l'identité d'une personne seront publiées en accès restreint (serveurs localisés aux États-Unis).
 - Il est possible de se procurer une trousse de curation professionnelle comprenant une analyse de divulgation. Communiquez à cette fin avec le service des acquisitions de l'ICPSR at deposit@icpsr.umich.edu pour en savoir plus.
- Sciences sociales
 - [Dépôt de données sur la COVID-19](#) d'OpenICPSR
 - [Dépôt de données qualitatives](#)
 - Tous les dépôts peuvent faire l'objet de consultation et de curation. Les données sensibles peuvent être assujetties au [contrôle de l'accès](#) (serveurs aux États-Unis).⁴ Les chercheurs associés à des institutions non membres auront à payer des [frais de service](#).⁵
- Logiciels
 - [Software Heritage Archive](#)
 - [Zenodo](#) (avec intégration à GitHub)
- Données cliniques
 - [Vivli](#) pour les données des essais cliniques
 - Accès restreint. [D-wise renoncera aux frais d'anonymisation des études sur la COVID-19](#) qui seront faites au moyen de la plateforme Vivli, et Vivli renoncera aux frais relatifs à la communication et à l'archivage de toutes les données des essais cliniques liés à la COVID-19 de même qu'à leur accès.⁶
 - [OpenNeuro](#) pour les IRM, MEG, EEG, iEEG et ECoG

³ RDA COVID-19 Working Group, "RDA COVID-19 Recommendations and Guidelines on Data Sharing," 30 juin 2020, <https://doi.org/10.15497/rda00052>.

⁴ "Human Participants – General Guidance," Qualitative Data Repository, consulté le 5 août 2020,, <https://qdr.syr.edu/guidance/human-participants>.

⁵ "Deposit Fees," Deposit, Qualitative Data Repository, consulté le 5 août 2020, <https://qdr.syr.edu/deposit#deposit-fees>.

⁶ "d-wise offers anonymization services to studies available on Vivli COVID-19 portal," Vivli, 13 avril 2020, <https://vivli.org/vivli-covid-19-portal/>.

- Dépôt en libre accès. Les données doivent être dépersonnalisées avant d'être partagées.
 - [EBRAINS](#) pour les données brutes ou traitées sur le cerveau humain, les modèles, les analyses, les méthodes, les codes, les logiciels
 - Dépôt en libre accès. Les données doivent être dépersonnalisées avant d'être partagées.
 - [base de données ouverte de radiologie sur la COVID-19 de RSNA International](#) pour l'imagerie thoracique (TDM et rayons X)
 - Dépôt en libre accès. Les images doivent être dépersonnalisées. RSNA a publié un [protocole de dépersonnalisation](#) et le logiciel Anonymizer à cette fin.
 - The [portail des bases de données et de l'analyse en immunologie](#) (ImmPort) pour la cytométrie et l'immunologie
- Acides nucléiques et « omiques »
 - Archives de la [collaboration internationale sur la base de données de séquences nucléotidiques \(INSDC\)](#) pour les données de génomique (données brutes sur la séquence du virus, données sur l'expression des gènes)
 - La [GenBank](#) de NCBI pour les données en génomique (génomomes assemblés et annotés)
 - La [Worldwide Protein Data Bank \(wwPDB\)](#) pour les données de structure
 - Voir la section 4.4. (pages 28 à 36) des [directives de la RDA](#) sur les recommandations concernant le génome du virus et le génome hôte, les données de structure, la protéomique, la métabolomique et la lipidomique.⁷

⁷ RDA COVID-19 Working Group, "RDA COVID-19 Recommendations and Guidelines on Data Sharing."

Option 2 : Dépôt de données de recherche canadien ou international général

Il existe un bon nombre de dépôts de données de recherche généraux, notamment les dépôts de bibliothèques universitaires canadiennes. Ils fournissent un lieu où déposer des données à partager sans égard au sujet de la recherche et ils sont sans doute plus aptes à recevoir d'une variété de données.

Votre choix pourrait s'arrêter sur ce genre de dépôts si :

- aucun dépôt disciplinaire ne convient ;
- vous souhaitez déposer avec vos données des programmes d'analyse et des renseignements complémentaires ;
- vous aimeriez partager vos données avec un public élargi.

Les dépôts canadiens vous permettront de garder vos données sur des serveurs canadiens, profiter de services de curation dans vos institutions (le cas échéant) et rendre vos données visibles à de larges communautés de données.

L'interopérabilité et la visibilité entre les dépôts canadiens et internationaux ne cessent de s'améliorer. Le Dépôt fédéré des données de recherche (DFDR) canadien recueille des métadonnées de nombreux autres dépôts canadiens, dont tous les dépôts du Scholars Portal Dataverse, et les partage avec [OpenAIRE](#) pour que les données canadiennes puissent être consultées à l'échelle internationale.⁸

Dépôts généraux recommandés

- [Scholars Portal Dataverse](#) (canadien)*
 - Dataverse est en accès restreint, mais les serveurs ne sont pas chiffrés. Le dépôt devrait servir à partager uniquement des données dépersonnalisées et non confidentielles.
- [Dépôt fédéré de données de recherche \(DFDR\)](#) (canadien)*
 - DFDR est un dépôt en libre accès. Les données doivent être dépersonnalisées avant d'être partagées.
- [Zenodo](#) (International)
 - Dépôt recommandé pour la [communauté de recherche sur les maladies à coronavirus](#)⁹
Dépôt en accès restreint, mais les serveurs ne sont pas chiffrés. Le dépôt devrait servir à partager uniquement des données dépersonnalisées et non confidentielles.

*En 2019, Portage a publié un guide intitulé [Options de dépôt au Canada : un guide de Portage](#) qui décrit et compare Dataverse et DFDR en détail. Si vous avez du mal à choisir entre ces deux options, nous vous recommandons de lire la section intitulée « Comment décider quelle option de dépôt utiliser ? »¹⁰.

⁸ "Explore," OpenAIRE, consulté le 5 août 2020, <https://explore.openaire.eu/search/find>.

⁹ "Coronavirus Disease Research Community," Zenodo, consulté le 5 août 2020, <https://zenodo.org/communities/covid-19/?page=1&size=20>.

¹⁰ Réseau Portage, Calcul Canada, Association des bibliothèques de recherche du Canada et Scholars Portal, *Options de dépôt au Canada : un guide de Portage*, 1^{er} juin 2019, <https://doi.org/10.5281/zenodo.3966356>.

Option 3 : Choisir son propre dépôt

Nous avons mis en évidence quelques dépôts dans nos recommandations précédentes, mais il existe beaucoup plus d'options disponibles pour différentes disciplines et différents objectifs. Vous aurez peut-être la chance de trouver un dépôt pertinent en utilisant les outils de découverte suivants.

Sachez que la plupart de ces outils répertorient une grande variété de dépôts, dont certains n'offrent pas de conservation à long terme ou n'acceptent pas les dépôts de données (il s'agit uniquement de plateformes de publication ou de découverte). Il peut être utile de garder à l'esprit ce [Survol des caractéristiques importantes des dépôts](#) lors votre exploration de ces ressources :

- Ce [Tableau de comparaison des dépôts généralistes](#) vous donne un aperçu des grands dépôts généralistes internationaux.¹¹
- L'outil de [repérage de dépôt](#) fouille dans le [répertoire re3data de dépôts de recherche](#) pour aider les chercheurs à trouver des options convenables de dépôt.¹²
- Vous pouvez [explorer re3data par sujet](#), puis utiliser des filtres pour raffiner vos résultats.¹³ Les filtres de téléversement (data upload) et de restrictions de téléversement (data upload restrictions) vous aident à éliminer les dépôts qui n'acceptent pas de nouveaux dépôts ou ceux qui imposent des limites quant aux déposants (par exemple, des dépôts réservés aux établissements membres).
- [FAIRsharing](#) peut servir à trouver diverses normes et bases de données dans votre domaine de recherche.¹⁴ Sachez que la catégorie des bases de données comprend à la fois les dépôts qui peuvent accueillir de nouveaux dépôts de données, et d'autres ressources qui cataloguent et hébergent des données, mais qui peuvent ne pas accepter de dépôts directs de données de la part des chercheurs.
 - L'équipe de FAIRsharing a créé une [collection de normes et de bases de données](#) qui sont convenables à l'utilisation dans le cadre de la pandémie COVID-19.¹⁵
 - Le groupe de travail sur la COVID-19 de la Research Data Alliance a également créé une [collection de bases de données, de politiques et de normes](#) pour les chercheurs et les décideurs politiques.¹⁶

¹¹ Shelley Stall, Maryann E. Martone, Ishwar Chandramouliswaran, Mercè Crosas, Lisa Federer, Julian Gautier, Mark Hahnel, *et al.* "Generalist Repository Comparison Chart", 15 juillet 2020, <http://doi.org/10.5281/zenodo.3946720>.

¹² DataCite, and American Geophysical Union, "Repository Finder", DataCite, consulté le 4 août 2020, <https://repositoryfinder.datacite.org/>; Berlin School of Library and Information Science, DataCite, Helmholtz Open Science Office, KIT Library, and Purdue University Libraries, "Registry of Research Data Repositories," re3data, consulté le 4 août 2020, <https://doi.org/10.17616/R3D>.

¹³ "Browse by Subject," re3data, consulté le 4 août 2020, <https://www.re3data.org/browse/by-subject/>.

¹⁴ "Standards, Databases, Policies," FAIRsharing, consulté le 4 août 2020, <https://www.FAIRsharing.org>.

¹⁵ "COVID-19 Resources," FAIRsharing, consulté le 5 août 2020, <https://fairsharing.org/collection/COVID19Resources>.

¹⁶ RDA COVID-19 Working Group, "RDA Covid-19 WG Resources," FAIRsharing, consulté le 5 août 2020, <https://fairsharing.org/collection/RDACovid19WG>.

Survol des caractéristiques importantes des dépôts

En plus de rendre vos données accessibles, les bons dépôts garantiront la fiabilité de celles-ci. Pour ce faire, les dépôts doivent disposer d'une infrastructure fiable, d'une bonne gouvernance, de politiques qui soutiennent leur communauté (tant les déposants que les utilisateurs) et ils doivent faire preuve d'une gestion responsable des données. Les recommandations suivantes sont adaptées des [principes TRUST pour les dépôts numériques](#) généralement reconnus.¹⁷

- Le dépôt a la possibilité d'attribuer des **identifiants uniques** aux ensembles de données (par exemple, des [identifiants d'objet numérique](#) ou des numéros d'acquisition).¹⁸
 - *Pourquoi est-ce important ?* Cela vous permettra et permettra aux autres de citer vos ensembles de données en reliant vos données aux résultats de vos recherches et en veillant à ce que vos données puissent être découvertes.
- Le dépôt a des **conditions d'utilisation** claires et **s'engage à conserver les fonds numériques**.
 - *Pourquoi est-ce important ?* Certains dépôts ne permettent pas de conserver les données. Si vous voulez vous assurer que vos données sont conservées à long terme, vous devez choisir un dépôt qui a une politique claire en matière de conservation.
- Le dépôt assure **l'intendance du fonds de données**, y compris la validation technique, la documentation et le contrôle de la qualité.
 - *Pourquoi est-ce important ?* Les déposants et les utilisateurs doivent avoir confiance dans la qualité durable des données contenues dans un dépôt.
- Le dépôt dispose d'un mécanisme d'**assistance aux utilisateurs**
 - *Pourquoi est-ce important ?* Les dépôts doivent être transparents et facilement compréhensibles par l'utilisateur final, en suivant l'évolution des attentes et en y répondant.
- Le dépôt est **reconnu au sein de sa communauté de recherche** et soutient les métadonnées et les normes de conservation désignées de la communauté.
 - *Pourquoi est-ce important ?* Où allez-vous pour rechercher des données dans votre domaine ? Les chercheurs doivent pouvoir trouver vos données dans les endroits où ils s'attendent à les trouver.

Quelques éléments à considérer

- Quels types de résultats de recherche devez-vous archiver ?
 - Les [lignes directrices de la RDA](#) recommandent aux chercheurs d'archiver les données situationnelles, les modèles analytiques, les résultats scientifiques et les rapports utilisés dans la prise de décision.¹⁹ Certains dépôts n'acceptent pas tous ces types de documents. Vous pouvez mettre en évidence les ressources connexes dans votre fichier README ou dans d'autres documents ; de nombreux dépôts ont un champ « ressources connexes » qui vous permettra de faire des liens directs à des ressources externes.
- Où vos données seront-elles stockées ? Vos données doivent-elles être hébergées au Canada ?
 - Si vous avez des préoccupations en matière de confidentialité ou de sécurité, vous devrez considérer le lieu de stockage du dépôt en plus de ses autres caractéristiques.

¹⁷ Dawei Lin, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giarretta, Marisa De Giusti, et al, "The TRUST Principles for Digital Repositories," *Scientific Data* 7, 144 (2020), <https://doi.org/10.1038/s41597-020-0486-7>.

¹⁸ "The DOI System, ISO 26324," DOI. Dernière modification 4 février 2020, <https://www.doi.org>.

¹⁹ RDA COVID-19 Working Group, "RDA COVID-19 Recommendations and Guidelines on Data Sharing."

- Est-ce que toutes vos données peuvent être partagées ? Avez-vous besoin d'un dépôt qui peut en restreindre l'accès ?
 - Si vous avez des questions sur le degré d'ouverture de vos données, consultez le processus décisionnel [Puis-je partager mes données ?](#) et le [Guide de dépersonnalisation](#) avant d'agir.²⁰

²⁰ Groupe de travail COVID-19 de Portage, « Puis-je partager mes données? » 23 septembre 2020, <https://doi.org/10.5281/zenodo.4046659>; Groupe de travail COVID-19 de Portage, « Directive sur la dépersonnalisation des données », 23 septembre 2020, <https://doi.org/10.5281/zenodo.4047175>.

Dépôts

1. EBRAINS <https://ebrains.eu/services/data-knowledge/share-data>
2. Dépôt fédéré de données de recherche (DFDR) <https://www.frdr-dfdr.ca/repo/>
3. Immunology Database and Analysis Portal (ImmPort) <https://www.immport.org/resources/dataTemplates>
4. International Nucleotide Sequence Database Collaboration (INSDC) archives <http://www.insdc.org/>
5. NCBI GenBank <https://www.ncbi.nlm.nih.gov/genbank/>
6. OpenICPSR COVID-19 Data Repository <https://www.openicpsr.org/openicpsr/covid19>
7. OpenNeuro <https://openneuro.org/>
8. Qualitative Data Repository <https://qdr.syr.edu/deposit/process>
9. RSNA International COVID-19 Open Radiology Database <https://www.rsna.org/covid-19/COVID-19-RICORD>
10. Scholars Portal Dataverse <https://dataverse.scholarsportal.info/fr/>
11. Software Heritage Archive <https://www.softwareheritage.org/save-and-reference-research-software/>
12. Vivli <https://vivli.org/>
13. Worldwide Protein Data Bank (wwPDB) <http://www.wwpdb.org/>
14. Zenodo <https://zenodo.org/>

Références

1. Berlin School of Library and Information Science, DataCite, Helmholtz Open Science Office, KIT Library et Purdue University Libraries. *Registry of Research Data Repositories*. re3data.org. Consulté le 4 août 2020, <https://doi.org/10.17616/R3D>.
2. Canadian Institutes of Health Research. *CIHR Open Access Policy: Annex*, Canadian Institutes of Health Research. Dernière modification le 15 mars 2015. <https://cihr-irsc.gc.ca/e/46068.html#8>.
3. DataCite, et American Geophysical Union. *Repository Finder*. Consulté le 4 août 2020, <https://repositoryfinder.datacite.org/>.
4. "The DOI System, ISO 26324," DOI. Dernière modification 4 février 2020, <https://www.doi.org>.
5. GitHub. *GitHub Guides : Making Your Code Citable*. Édité en octobre 2016. Consulté le 5 août 2020. <https://guides.github.com/activities/citable-code/>.
6. Groupe de travail COVID-19 de Portage. « Directive sur la dépersonnalisation des données ». 23 septembre 2020. <https://doi.org/10.5281/zenodo.4047175>.
7. Groupe de travail COVID-19 de Portage. « Puis-je partager mes données? ». 23 septembre 2020. <https://doi.org/10.5281/zenodo.4046659>.

8. FAIRsharing. *Standards, Databases, Policies*. Consulté le 4 août 2020 <https://www.FAIRsharing.org>.
9. The FAIRsharing Team. *COVID-19 Resources*. FAIRsharing. Consulté le 5 août 2020 <https://fairsharing.org/collection/COVID19Resources>.
10. Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, et al. *The TRUST Principles for Digital Repositories, Scientific Data*, 7: 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>.
11. PLOS ONE. *Recommended Repositories*. PLOS. Consulté le 4 août 2020. <https://journals.plos.org/plosone/s/recommended-repositories>.
12. RDA COVID-19 Working Group. *RDA Covid-19 WG Resources*, FAIRsharing. Consulté le 5 août 2020 <https://fairsharing.org/collection/RDACovid19WG>.
13. RDA COVID-19 Working Group. *RDA COVID-19 Recommendations and Guidelines on Data Sharing*. 30 juin 2020 <https://doi.org/10.15497/rda00052>.
14. Réseau Portage, Calcul Canada, Association des bibliothèques de recherche du Canada et Scholars Portal. *Options de dépôt au Canada : un guide de Portage*. 1^{er} juin 2019. <https://doi.org/10.5281/zenodo.3966349>.
15. Scientific Data. *Recommended Data Repositories*. Nature Research. Consulté le 4 août 2020 <https://www.nature.com/sdata/policies/repositories>.
16. Stall, Shelley, Maryann E. Martone, Ishwar Chandramouliswaran, Mercè Crosas, Lisa Federer, Julian Gautier, Mark Hahnel, et al. *Generalist Repository Comparison Chart*. Zenodo. 15 juillet 2020 <http://doi.org/10.5281/zenodo.3946720>.