

# Options de dépôt au Canada : un guide de Portage

juin 2019



[www.carl-abrc.ca](http://www.carl-abrc.ca)

**portage**

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



## Table des matières

Pourquoi publier mes données? .....	2
Où puis-je publier mes données de recherche?.....	3
Options de dépôt nationales .....	4
Dataverse Nord.....	5
Dépôt fédéré des données de recherche (DFDR).....	6
Comment décider quelle option de dépôt utiliser? .....	7
Ressources .....	13

## Pourquoi publier mes données?

La publication des données de recherche en facilite la réutilisation intra ou interdisciplinaire. Certaines données publiées sont ouvertes et prêtes à être partagées ou réutilisées sans restriction (p. ex. sous licence [Creative Commons](#) ou autre licence de données libres); dans certains cas, il peut être approprié d'imposer des restrictions sur le partage et la réutilisation.

Les organismes de financement et les revues exigent de plus en plus la publication des données dans un dépôt fiable dans le cadre d'un plan de gestion des données bien établi (consulter [l'ébauche de la Politique des trois organismes sur la gestion des données de recherche](#)).<sup>1</sup>

La publication des données est également une bonne façon d'en assurer l'accessibilité pérenne, même après la conclusion de l'étude pour laquelle elles ont été recueillies. La capacité de trouver et de réutiliser des données est d'une importance grandissante pour la vérification des travaux de recherche publiés et pour appuyer de nouvelles études. Pour de plus amples renseignements, consultez les modules de formation sur la [gestion des données de recherche 101](#).

---

<sup>1</sup> [L'Assistant PGD](#) du Réseau Portage est un outil en ligne et bilingue qui aide les chercheurs dans leur GDR.

## Où puis-je publier mes données de recherche?

Le Canada peut compter sur un riche environnement de dépôts de données, qu'ils soient d'envergure nationale, régionale, locale ou institutionnelle. Les dépôts de données sont variés et nombreux, mais peuvent être catégorisés de manière globale comme étant propres à un domaine, gouvernementaux ou généralistes.

Comment décider quel type de dépôt utiliser?

- Y a-t-il un dépôt propre à un domaine pour votre discipline?
  - Si oui, il pourrait s'agir du meilleur endroit pour y déposer vos données puisqu'ils sont souvent adaptés aux besoins spécialisés de votre discipline.<sup>2</sup>
- Menez-vous vos travaux de recherche pour le compte du gouvernement, que ce soit une agence ou un ministère du gouvernement fédéral ou provincial, ou une administration municipale?
  - Si oui, vous pourriez envisager un dépôt propre à un domaine ou un dépôt gouvernemental, le cas échéant.
- Si un dépôt propre à un domaine n'est pas la meilleure option, vous pourriez envisager un dépôt généraliste acceptant généralement tout type de données de recherche.
- Vous pouvez choisir différents dépôts généralistes; certains offrent un modèle tarifaire par service alors que d'autres ne facturent aucuns frais directement aux chercheurs. Le présent guide couvre deux options de dépôts offertes à la communauté de la recherche au Canada sans frais grâce à un partenariat avec les bibliothèques universitaires.

Avant de continuer, la première étape est de communiquer avec votre [représentant en gestion de données de recherche local](#) pour discuter de vos besoins en matière de publication de données.

---

<sup>2</sup> Consultez <https://www.re3data.org/> pour obtenir un registre des dépôts de données.

## Options de dépôt nationales

Portage appuie deux options nationales de dépôt des données : le Dépôt fédéré des données de recherche (DFDR) et Dataverse Nord. Dans l'esprit du modèle **d'architecture fédéré de Portage**, nous avons établi un partenariat avec Scholars Portal du CBUO, Calcul Canada et divers établissements d'enseignement offrant Dataverse pour offrir des options d'hébergement ne nécessitant pas l'installation, la configuration, les résolutions de bogues ou la mise à niveau pour les établissements ou les chercheurs.

Les efforts de développement sont en cours en termes de financement, de fonctionnalités et de services. Le présent document sera mis à jour pour refléter les développements au fur et à mesure de l'évolution du paysage.

Ces options de dépôt vous permettent de :

- Publier vos jeux de données, peu importe le type ou le format de fichiers, aux fins de découverte et de réutilisation;
- Décrire les jeux de données à l'aide de champs de métadonnées appropriés à votre discipline;
- Obtenir immédiatement un Digital Object Identifier (DOI) pour la publication et la citation de vos données;
- Établir les conditions de licence relativement à l'utilisation de vos jeux de données;
- Bénéficier d'un espace de stockage sécuritaire et basé au Canada;
- Mesurer l'impact de vos travaux de recherche en suivant l'utilisation de vos jeux de données;
- Respecter les exigences des organismes subventionnaires et des revues concernant la publication des données; et
- Rendre vos données découvrables en libre accès et réutilisables pour le reste de la planète.

Que vos données soient hébergées sur une instance Dataverse, dans le DFDR ou dans une liste croissante de dépôts de données propres à un domaine ou gouvernementaux, elles sont toutes facilement découvrables **dans le Service national de découverte de Portage** (<http://frdr.ca>). Ce service moissonne et indexe les métadonnées pour rendre les données de partout au Canada découvrables à partir d'un portail web unique. Cette approche fédérée pancanadienne encourage un riche paysage de dépôt servant une grande diversité de besoins, amplifie la consultation des dépôts hôtes et aide à éliminer les silos de données et disciplinaires.

## Dataverse Nord

Dataverse Nord est une communauté de pratique de fournisseurs et d'utilisateurs de Dataverse au Canada. Dataverse est une plateforme de dépôt libre pour la publication des données de recherche qui est conçue et maintenue par l'Université Harvard et des contributeurs de partout dans le monde.

Scholars Portal, un service du Conseil des bibliothèques universitaires de l'Ontario (CBUO), offre Dataverse à un [nombre croissant d'universités au Canada](#). Scholars Portal est partenaire de Portage pour offrir Dataverse comme service aux établissements d'enseignement canadiens contributeurs.

Vous trouverez d'autres fournisseurs de Dataverse [ici](#).

Principales caractéristiques :

- **Personnalisable** : Les établissements et les groupes de recherche peuvent créer des collections unifiées par une marque commune pour conserver leurs données de recherche à un même endroit.
- **Politiques locales** : Les politiques de dépôt peuvent être précisées pour l'établissement et varient des dépôts avec médiation aux modèles d'autopublication.
- **Contrôle de l'accès aux données** : Les chercheurs peuvent choisir le niveau d'ouverture de leurs données, y compris pour chaque fichier individuel. Vous pouvez changer ces paramètres en tout temps.
- **Une bonne option dans la plupart des cas** : Adapté pour les jeux de données de petite ou de moyenne taille (jusqu'à 2 Go) fait de Dataverse une bonne solution pour la «longue traîne» des données.

Pour de plus amples renseignements, veuillez consulter [Dataverse : l'essentiel](#) et [Dataverse : informations de base](#).

## Dépôt fédéré des données de recherche (DFDR)

Le DFDR compte sur une infrastructure de recherche numérique pancanadienne, grâce à un partenariat entre Portage de l'ABRC et Calcul Canada, pour offrir un dépôt robuste et rentable à grande échelle.

Principales caractéristiques :

- **Soutien pour les jeux de données volumineux** : La technologie de transferts des données du DFDR est adaptée aux jeux de données volumineux.
- **Maintien de l'arborescence des fichiers** : Le DFDR vous permet de maintenir les relations hiérarchiques des fichiers et dossiers durant le transfert.
- **Connexion à l'infrastructure nationale de recherche numérique** : Globus File Transfer est utilisée par les systèmes de Calcul Canada, offrant ainsi aux chercheurs qui utilisent déjà ces systèmes dans le cadre de leurs processus de recherche un moyen simple de transférer et de publier des données dans le DFDR.
- **Grande allocation par chercheur** : L'équipe travaille actuellement à offrir une « allocation de stockage générale » à tous les chercheurs au Canada (potentiellement jusqu'à 3 To sans application ou justification autre que la pertinence de la publication, l'évaluation par des curateurs et le respect des conditions d'utilisation du DFDR) ainsi que la possibilité de demander un espace de stockage supplémentaire.

Pour de plus amples renseignements, consultez le [guide rapide sur le DFDR](#) de Portage.

## Comment décider quelle option de dépôt utiliser?

Il n'y a pas de mauvaise option, mais les chercheurs pourraient trouver un dépôt plus approprié qu'un autre pour leurs besoins ou le type d'utilisation souhaitée (consulter le tableau ci-dessous). Vous n'avez pas à choisir l'une ou l'autre des options : vous pouvez opter pour Dataverse pour vos jeux de données de petite ou moyenne taille, tout en tirant profit des capacités du DFDR pour vos jeux de données massives.

Questions sur l'utilisation	Dataverse	DFDR
Mon établissement offre-t-il un dépôt?	Communiquer avec votre représentant en GDR pour vérifier les options locales.  Consultez le <a href="#">site web de Portage</a> pour vérifier si votre établissement fait partie de Dataverse Nord.	Offert à tous les chercheurs principaux au Canada et à leurs personnes désignées. Option idéale si vous n'avez pas de dépôt institutionnel.
Devrez-vous mettre à jour votre jeu de données périodiquement?	Dataverse comporte une fonction de versionnage robuste qui permet aux usagers de suivre et de télécharger les vieilles versions de leurs données. La mise à jour des ensembles de données peut se faire manuellement et facilement.	La mise à jour de jeux de données dans le DFDR se fait en communiquant avec un curateur. Le DFDR ne conserve pas les différentes versions de vos données.



<p>Avez-vous des fichiers de données volumineux (plus de 2 Go), un grand nombre de fichiers ou une arborescence complexe?</p>	<p>Actuellement, Dataverse supporte les fichiers jusqu'à 2 Go. Les efforts de développement sont en cours pour améliorer l'extensibilité et la capacité de traitement des fichiers volumineux.</p>	<p>Le DFDR prend en charge les transferts de fichiers volumineux (que ce soit la taille des fichiers ou le nombre de fichiers) facilement et maintient les arborescences complexes. Les services de stockage du DFDR sont extensibles et rentables pour la recherche hautement axée sur les données.</p>
<p>Devez-vous contrôler l'accès à vos jeux de données et à vos fichiers?*</p>	<p>Dataverse offre un contrôle granulaire pour l'accès aux données; les permissions peuvent être définies pour les collections, les ensembles de données ou encore les fichiers.</p>	<p>Dans le DFDR, vous pouvez restreindre temporairement l'accès aux jeux de données au moyen d'un embargo, mais tous les fichiers doivent éventuellement être publiés.</p>
<p>Quels types de solutions de stockage sont intégrées?</p>	<p>Les instances de Dataverse au Canada sont hébergées sur les serveurs des établissements participants. Vous pouvez téléverser dans Dataverse à partir de certaines solutions de stockage d'infonuagique aux États-Unis, y compris Open Science Framework et Dropbox.</p>	<p>Le DFDR est connecté au réseau de stockage national de Calcul Canada, qui comprend des options de stockage à distance. Vous pouvez transférer des données dans le DFDR à partir des systèmes de Calcul Canada, de serveurs locaux ou institutionnels ou encore de votre ordinateur de bureau ou portable personnel.</p>

<p>Exigez-vous une collection spéciale pour vos ensembles de données?</p>	<p>Dataverse permet la création de contenants personnalisés et marqués (des dataverses) qui stockent vos ensembles de données dans des collections qui peuvent être partagées et consultées. Les permissions relatives au dépôt et au téléchargement peuvent être définies pour les collections, les ensembles de données ou encore les fichiers.</p>	<p>Les groupes de recherche peuvent demander des collections spéciales pour leurs ensembles de données. Les pages de collection comprennent votre marque, de l'information sur le projet, des liens vers des portails web existants et une liste recherchable d'ensembles de données publiés. Vous contrôlez qui peut déposer dans la collection.</p>
<p>La prévisualisation et la visualisation des données sont-elles importantes?</p>	<p>Dataverse offre l'exploration d'ensembles de données tabulaires, offrant aux usagers une liste complète de variables et de statistiques générales à même un navigateur web.</p>	<p>Dans le DFDR, vous pouvez parcourir les dossiers et l'arborescence avant de décider si vous souhaitez télécharger un jeu de données, mais la prévisualisation est limitée aux fichiers téléchargeables par protocole HTTPS et visualisables dans un navigateur web.</p>

<p>Avez-vous à collaborer avec d'autres membres de votre équipe de recherche?</p>	<p>Dataverse comporte des fonctionnalités appuyant la recherche collaborative. Vous pouvez suivre les versions des fichiers, décrire des ensembles de données et fournir des notes de version, contrôler l'accès aux fichiers pour les membres de l'équipe, assigner différents rôles relativement aux accès tels que « administrateur », « contributeur » ou « curateur » pour un dataverse (collection) ou un ensemble de données.</p>	<p>Le DFDR est un dépôt pour la publication des données de recherche. Il ne supporte pas les fonctionnalités actives de gestion des données de recherche.</p>
<p>Le système permet-il les embargos de jeux de données? Est-ce automatisé?</p>	<p>Actuellement, Dataverse ne supporte pas l'embargo automatique des ensembles de données. Vous pouvez maintenir un ensemble de données non publié pour en restreindre l'accès et générer une URL privée pour la partager au besoin avec un éditeur ou un organisme subventionnaire.</p>	<p>Le DFDR supporte les périodes d'embargo, typiquement jusqu'à deux ans. Vous recevrez un avis avant que votre jeu de données soit publié et vous pourrez demander une extension.</p>

<p>Avez-vous besoin d'API pour intégrer ou automatiser certains aspects de votre processus de publication?</p>	<p>Dataverse comporte une série d'API robustes pour la recherche, le repérage et les téléversements. L'API native vous permet d'automatiser de nombreux aspects de votre processus de publication.</p>	<p>Le DFDR maintient une série complète d'API pour le dépôt et la découverte de jeux de données. Pour de plus amples renseignements, consultez la <a href="#">documentation</a>.</p>
<p>Aurez-vous besoin de champs de métadonnées spécialisés ou propres à une discipline pour décrire vos données?</p>	<p>Dataverse appuie les champs de citation et de description généraux, ainsi que les schémas de métadonnées spécialisés en sciences sociales, en géosciences, en santé et sciences de la vie, ainsi qu'en astronomie.</p>	<p>Par défaut, le DFDR utilise des normes générales de description et de découverte des données. Vous pouvez faire la demande de champs de métadonnées personnalisés et des formulaires web propres à votre discipline ou votre utilisation.</p>
<p>Quels flux de travail sont disponibles pour la préservation à long terme de mes données?</p>	<p>Les bibliothèques universitaires offrant accès à Dataverse peuvent avoir intégré des solutions ou des services de préservation numérique et d'archivage des données. Communiquez avec votre représentant local de GDR pour en savoir davantage sur l'archivage des données.</p>	<p>Le DFDR offre un processus de préservation au moyen de la création d'un paquet d'information archivé (AIP) pour les jeux de données de toute taille. Vous pouvez demander que cet AIP vous soit transféré ou vers un fournisseur de service de préservation de votre choix pour l'archivage à long terme.</p>

\* À l'heure actuelle, ni l'une, ni l'autre des options de dépôt n'est adaptée aux données de recherche sensibles.

## Ressources

- Ébauche de la Politique des trois organismes sur la GDR : [http://www.science.gc.ca/eic/site/063.nsf/fra/h\\_97610.html](http://www.science.gc.ca/eic/site/063.nsf/fra/h_97610.html)
- Dépôt fédéré des données de recherche : <https://www.frdr.ca/>
- Dataverse : l'essentiel : [https://portagenetwork.ca/wp-content/uploads/2019/05/Datavere-Essentiel\\_FR.pdf](https://portagenetwork.ca/wp-content/uploads/2019/05/Datavere-Essentiel_FR.pdf)
- Dataverse : Informations de base : [https://portagenetwork.ca/wp-content/uploads/2019/05/DVN\\_Information-de-base\\_FR.pdf](https://portagenetwork.ca/wp-content/uploads/2019/05/DVN_Information-de-base_FR.pdf)
- DFDR : l'essentiel : [https://portagenetwork.ca/wp-content/uploads/2019/05/FRDR\\_Essentiel\\_FR.pdf](https://portagenetwork.ca/wp-content/uploads/2019/05/FRDR_Essentiel_FR.pdf)
- Assistant PGD de Portage : <https://assistant.portagenetwork.ca/fr>
- Réseau Portage : <https://reseauportage.ca/>
- Personnes-ressources dans votre institution : <https://portagenetwork.ca/fr/comment-gerer-vos-donnees/personne-ressource-dans-votre-institution/>
- Registre des dépôts de données de recherche (en anglais) : <https://www.re3data.org/>
- Module de formation sur la Gestion des données de recherche 101 : <https://portagenetwork.ca/fr/ressources-formation/ressources-de-formation-par-portage/gestion-des-donnees-de-recherche-gdr-101/>
- Dataverse de Scholars Portal : <https://dataverse.scholarsportal.info/>