

Dépôt de données de recherche dans le Dataverse de l'Université de Sherbrooke

Lignes directrices

Document rédigé par le Service des bibliothèques et archives et le Service d'appui à la recherche, à la création et à l'innovation

Hiver 2023

TABLE DES MATIERES

Présentation de Dataverse	3
Personnes autorisées à déposer	5
Rôles.....	5
Quoi déposer ?.....	4
Documents autorisés	4
Documents non autorisés	4
Gestion des accès.....	5
Embargo	5
Accès restreint	5
Données obsolètes et retrait de données	6
Durée de conservation.....	6
Responsabilités des personnes déposantes	6
Conformité éthique et légale	6
Licences pour réutilisation	7
Intégrité du contenu	7
Description et documentation.....	7
Conditions d'utilisation de la plateforme	8
Rôles et responsabilités du personnel administrateur et curateur	8
Gestion du Dataverse de l'Université de Sherbrooke.....	8
Médiation et curation des dépôts	8
Création et organisation des collections.....	9
Noms et catégorisation des collections.....	9
Conventions de nommage	9
Collections.....	9

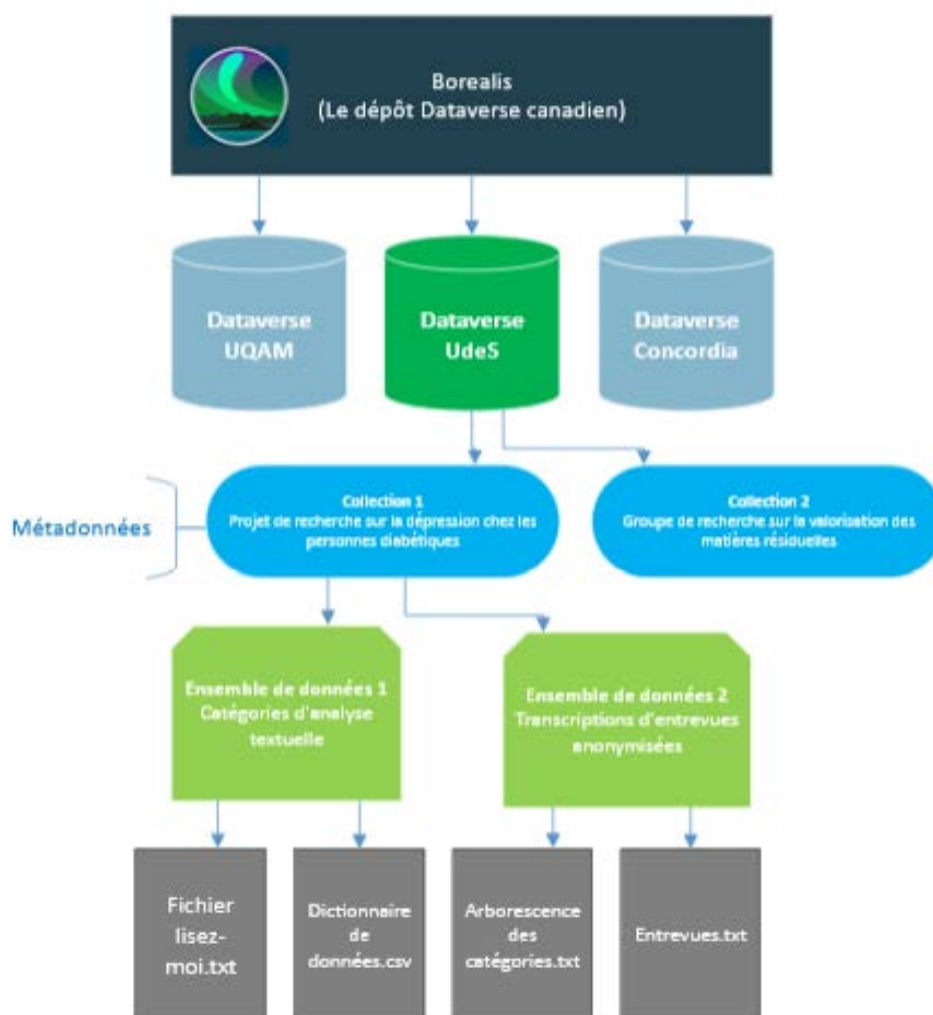
Fichiers et ensembles de données.....	10
Annexe 1 - Définitions.....	11
Annexe 2 - Formats recommandés.....	133
Annexe 3 - Licences.....	14

PRESENTATION DE DATAVERSE

Le Dataverse de l'Université de Sherbrooke (UdeS) est une plateforme permettant aux membres de la communauté universitaire de déposer, de partager et de conserver leurs données de recherche¹. Elle permet également le repérage des données partagées par les autres chercheuses et chercheurs de toutes les institutions. Il s'agit d'un logiciel libre et sans frais d'utilisation pour les personnes déposantes. Ce dépôt est une instance locale de la plateforme Borealis : le dépôt Dataverse canadien qui chapeaute les dépôts institutionnels de la plupart des universités du Québec et du Canada et dont les serveurs sont localisés à l'Université de Toronto.

Lors d'un dépôt de données, une collection est créée, qui contiendra un certain nombre d'ensembles de données. Chaque ensemble est un contenant qui comprend un fichier de données particulier ainsi que la documentation, le code et les métadonnées en lien avec le fichier de données.

À noter que le dépôt des données peut se faire sur une autre plateforme que Dataverse, qui est un dépôt institutionnel offert à la communauté de recherche de l'Université de Sherbrooke.



¹ [Dataverse de l'Université de Sherbrooke](#)

QUOI DEPOSER ?

DOCUMENTS AUTORISÉS

Le [Dataverse de l'Université de Sherbrooke](#) comporte certaines restrictions imposées par les contraintes éthiques et légales en vigueur (ex : licence associée à un ensemble de données) ainsi que par la limite technologique de la plateforme qui ne permet pas le dépôt de document de plus de 3 Go par fichier. Les documents suivants peuvent être déposés dans le Dataverse:

- Données qui soutiennent les résultats d'un projet de recherche;
- Données de recherche et documentation technique liées à un article scientifique publié par le déposant ou la déposante (par exemple, données exigées par un éditeur lors de la publication d'un article);
- Toute documentation permettant de comprendre et de réutiliser les données de recherche (ex. : Fichier « Lisez-moi », dictionnaire de données, instruments de collecte, grilles d'analyse, renseignements sur les données manquantes, traitements effectués, état des données, données liées, choix méthodologiques, métadonnées descriptives, formulaires de consentement vierges...);
- Tout code informatique permettant de traiter les données (ex. : script permettant d'extraire les concepts d'un verbatim, d'en dresser la liste avec le nombre d'occurrences ; script permettant de visualiser ces données...);
- Métadonnées d'un fichier ou d'un ensemble de données sous embargo ou non déposées, ou déposées sur une autre plateforme, avec l'adresse web du dépôt.

DOCUMENTS NON AUTORISÉS

Les documents suivants ne peuvent pas être déposés dans le Dataverse de l'Université de Sherbrooke :

- Données de recherche non conformes aux exigences éthiques, légales et réglementaires, ainsi qu'aux obligations en matière de gestion des données confidentielles et autres obligations contractuelles (ex : renseignements personnels (qu'ils soient dépersonnalisés ou non), renseignements confidentiels nécessaires pour la protection d'une invention, éléments protégés par de la propriété intellectuelle, entre autres);
- Articles publiés et travaux académiques diplômants (mémoires, thèses, essais). Ces documents doivent être déposés dans la plateforme [Savoirs UdeS](#), prévue à cette fin;
- Outils ou contenus dérivés du projet de recherche (infographies, ressources éducatives libres, billets de blogue...);
- Documents administratifs liés au projet (PGD, certificat d'éthique, formulaires de consentement remplis, demande de subvention, évaluation des coûts...);
- Données actives, c'est-à-dire des données qui ne soutiennent pas les résultats d'un projet de recherche.

PERSONNES AUTORISEES A DEPOSER

L'accès interne au Dataverse institutionnel est régi par l'authentification institutionnelle. De ce fait, tous les membres de la communauté universitaire, soit ceux possédant un CIP actif, pourront s'y connecter. Toutefois, seuls les déposants et déposantes autorisés pourront y déposer des données.

Les personnes déposantes autorisées sont :

- les professeures ou professeurs;
- les postdoctorantes ou postdoctorants;
- tout autre membre de la communauté universitaire délégué par une professeure ou un professeur;
- les personnes étudiantes, si elles respectent l'ensemble des conditions suivantes :
 1. la personne déposante certifie que les données de recherche qu'elle souhaite déposer sont sous sa responsabilité et celle de l'Université ;
 2. les données ont été produites dans le cadre d'un mémoire, d'une thèse ou d'un autre projet diplômant, une fois qu'il est terminé.

Note : les données des étudiants seront déposées dans une collection prévue à cet effet.

Si une personne membre de la communauté UdeS ne relevant pas de l'une de ces catégories souhaite effectuer un dépôt, elle peut en faire la demande à dataverse_udes@usherbrooke.ca. Les demandes seront examinées au cas par cas.

Les collaboratrices et collaborateurs externes à l'UdeS ne seront pas autorisés à déposer.

ROLES

Par défaut, toutes les personnes déposantes auront le rôle « Collaboration » pour leur propre collection. Ce rôle permet de modifier les métadonnées, téléverser des fichiers, modifier les fichiers, modifier les conditions et soumettre les ensembles de données aux fins de curation. Cependant, sous certaines conditions et suivant une entente avec le Service des bibliothèques et archives (SBA), il pourrait être envisagé d'attribuer un rôle plus permissif à certaines personnes qui le demandent, leur donnant ainsi plus de latitude et d'autonomie dans leur collection. Cette possibilité n'est offerte qu'aux membres du corps professoral ou aux personnes mandatées par celles et ceux-ci.

GESTION DES ACCES

EMBARGO

Dans certains cas, la chercheuse ou le chercheur peut restreindre l'accès à un ensemble ou à un fichier de données dans le temps en planifiant, à partir du moment du dépôt, un délai avant que les données soient accessibles au public. C'est le cas notamment lorsque l'article est en voie d'être publié, ou encore lorsque les données utilisées font l'objet d'un futur brevet.

ACCES RESTREINT

Le fait de publier un ensemble de données dans Dataverse rend cet ensemble ouvert au public par défaut et accessible pour consultation.

Bien que cette ouverture soit souhaitable et fortement encouragée, la chercheuse ou le chercheur peut, dans certains cas, restreindre l'accès à ses données. Elle ou il peut alors indiquer des conditions d'accès spécifiques ou permettre l'affichage d'un bouton qui générera un courriel de demande d'accès qui lui sera envoyé. Dans ce genre de cas, le chercheur ou la chercheuse devra tout de même publier les métadonnées signalant l'existence de l'ensemble de données.

DONNEES OBSOLETES ET RETRAIT DE DONNEES

Si des corrections ou des modifications à des données publiées sont requises, il est possible de créer une nouvelle version de l'ensemble de données qui remplacera la précédente.

Il est aussi possible de supprimer un fichier à l'intérieur d'un ensemble de données publié. Selon les versions existantes pour l'ensemble de données, il est possible que ce fichier ne soit supprimé que dans la version actuelle et non pas dans toutes les versions antérieures.

Par ailleurs, il n'est pas possible de supprimer un ensemble complet de données publiées, mais il est toutefois possible de le rendre inaccessible au public. Un enregistrement de ces données subsistera dans Dataverse. Si un ensemble de données est rendu inaccessible par erreur, le seul moyen de corriger la situation serait de recréer l'ensemble de données.

DUREE DE CONSERVATION

Lors du dépôt d'un ensemble de données, la chercheuse ou le chercheur doit indiquer dans le champ « Remarques » la durée de conservation souhaitée en nombre d'années, en tenant compte notamment de l'intérêt qu'auront ses données pour les recherches futures ainsi que de leur pérennité (dans la forme et le contenu). Cette information sera utilisée à titre indicatif par le personnel administrant le Dataverse de l'Université de Sherbrooke.

RESPONSABILITES DES PERSONNES DEPOSANTES

La chercheuse ou le chercheur souhaitant déposer ses données dans le Dataverse de l'Université de Sherbrooke doit s'assurer de répondre aux conditions mentionnées dans la Politique sur la gestion des données de recherche². La chercheuse ou le chercheur doit aussi s'assurer de respecter la future Politique sur la protection de la propriété intellectuelle et assumer les responsabilités précisées à la section 4.3 « Responsabilités des chercheurs et chercheuses »³. En outre, elle ou il doit vérifier les éléments énumérés ci-après avant de soumettre son ensemble de données pour publication. En cas d'erreur ou d'omission, l'équipe du Dataverse de l'Université de Sherbrooke lui demandera des corrections ou, dans certains cas, les apportera directement après vérification avec la personne déposante.

CONFORMITE ETHIQUE ET LEGALE

La chercheuse ou le chercheur déposant ses données dans le Dataverse de l'Université de Sherbrooke doit se conformer aux obligations éthiques, légales et commerciales liées à ses données, à la Politique sur la gestion des données de recherche en matière d'éthique ainsi qu'à la future Politique sur la protection de la propriété intellectuelle.

² [Politique sur la gestion des données de recherche](#)

³ La Politique sur la protection de la propriété intellectuelle est actuellement en révision.

LICENCES POUR REUTILISATION

Lors du dépôt, la personne déposante qui souhaite partager ses données est fortement encouragée à attribuer aux données publiées une licence ouverte permettant la réutilisation, telle que [Creative Commons](#)⁴ ou [Open Data Commons](#)⁵, sauf dans la mesure où cela n'est pas conforme aux exigences éthiques, légales, règlementaires ou contractuelles applicables à ces données. De même, la personne déposante s'engage à prendre connaissance des licences et de bien comprendre leur implication avant de les utiliser.

INTEGRITE DU CONTENU

Pour assurer l'intégrité du contenu, la personne déposante doit valider que :

- les fichiers s'ouvrent correctement et ne sont pas vides ;
- le format des documents est approprié (voir l'[Annexe 2 - Formats recommandés](#)) ;
- les données sont complètes, uniformes et exactes.

DESCRIPTION ET DOCUMENTATION

Afin que les données déposées soient repérables et accessibles, la personne déposante doit préparer ses données conformément aux bonnes pratiques suivantes :

- remplir les champs obligatoires de métadonnées suivants :
 - titre
 - autrice ou auteur et identifiant unique (exemple : [ORCID](#)), le cas échéant
 - personne-ressource
 - adresse courriel institutionnelle
 - description
 - sujet (liste déroulante)
 - mots-clés (de l'auteur)
- remplir les champs recommandés de métadonnées suivants :
 - personne-ressource
 - nom, prénom
 - affiliation
 - **Note** : inclure au moins deux personnes-ressources par ensemble de données.
 - publication connexe (si les données sont liées à une publication)
 - Remarques (durée de conservation et suppression des données le cas échéant)
 - personne déposante (si différente de l'autrice ou de l'auteur)
 - date de dépôt
- ajouter des métadonnées spécifiques du domaine de recherche pour améliorer la visibilité des données déposées⁶ ;
- nommer les fichiers en respectant les règles d'écritures (voir la sous-section [Conventions de nommage](#)) ;

⁴ [Creative Commons](#)

⁵ [Open Data Commons](#)

⁶ Voir Dataverse Nord (2021). [Guide des pratiques exemplaires sur les métadonnées.](#)

- ajouter la documentation permettant de comprendre et d'utiliser les données. Tous les ensembles de données doivent inclure un fichier « Lisez-moi » » (voir <https://zenodo.org/record/4058962#.Y8lwC3bMK3B>). La documentation doit également inclure :
 - les liens entre les fichiers,
 - la liste des répertoires et noms de fichiers, y compris une description du contenu,
 - les formats des fichiers,
 - la version du fichier,
 - une explication des codes, acronymes et abréviations utilisées,
 - les procédures de collecte, d'analyse et de traitement des données,
 - les logiciels utilisés pour produire ou traiter les données,
 - de l'information sur l'exhaustivité de l'ensemble de données,
 - les limites à la réutilisation (selon les licences choisies lors du dépôt).

CONDITIONS D'UTILISATION DE LA PLATEFORME

Les personnes qui déposent dans le Dataverse de l'Université de Sherbrooke, ainsi que celles qui accèdent et utilisent les données déposées, s'engagent à respecter les [conditions d'utilisation](#) de Borealis. Les utilisatrices et les utilisateurs devront également respecter les licences déterminées par la déposante ou le déposant.

ROLES ET RESPONSABILITES DU PERSONNEL ADMINISTRATEUR ET CURATEUR

GESTION DU DATAVERSE DE L'UNIVERSITE DE SHERBROOKE

L'administration du Dataverse institutionnel, comprenant notamment l'autorisation des personnes déposantes et la création des collections, relève du SBA.

Les responsabilités assumées par les personnes administratrices du Dataverse de l'Université de Sherbrooke sont les suivantes :

- autoriser les accès
- créer et assigner les collections aux personnes déposantes (processus restrictif, processus intermédiaire/restrictif)
- vérifier les métadonnées (voir la section [Médiation et curation des dépôts](#))
- gérer les rôles
- offrir un soutien dans l'utilisation de la plateforme
- veiller au bon fonctionnement et à l'évolution de la plateforme.

MEDIATION ET CURATION DES DEPOTS

Lors du dépôt des données par la chercheuse ou le chercheur, le personnel administrateur du Dataverse de l'Université de Sherbrooke est immédiatement avisé. S'ensuit une étape de vérification des données et des métadonnées : la curation. Cette étape, préalable à la publication des données, est effectuée par une équipe de bibliothécaires et de techniciennes et techniciens en documentation du SBA.

La curation effectuée dans le Dataverse de l'Université de Sherbrooke, inspirée des recommandations des experts de l'Alliance de recherche numérique du Canada⁷, consistera à faire les vérifications suivantes :

- le choix de la collection,
- la capacité d'ouvrir les fichiers (de formats standards),
- les métadonnées et la documentation,
 - structure et nommage des fichiers (voir la section Conventions de nommage),
 - présence des fichiers mentionnés dans la documentation,
 - présence d'un fichier « Lisez-moi »,
 - champs de métadonnées,
- le format des fichiers (ouvert, si possible),
- la présence d'une licence de réutilisation.

L'équipe responsable de l'administration du Dataverse de l'Université de Sherbrooke pourrait soutenir la personne déposante en lui apportant des suggestions et/ou corrections à effectuer avant la publication des fichiers de données. Au besoin, elle se réserve également le droit d'apporter des corrections :

- à la forme des métadonnées;
- aux noms des collections, ensembles de données et répertoires, par souci d'uniformité, conformément aux conventions de nommage établies.

CREATION ET ORGANISATION DES COLLECTIONS

NOMS ET CATEGORISATION DES COLLECTIONS

Afin de maximiser l'organisation structurelle et ainsi optimiser le repérage, il est recommandé de se doter de lignes directrices pour la création et la nomenclature des collections, des ensembles de données et des fichiers.

Les collections ne seront pas créées au nom de la chercheuse ou du chercheur dans la plupart des cas. Il est suggéré de créer les collections par nom de projet, de laboratoire, ou de groupe de recherche. Le nom de la chercheuse ou du chercheur sera choisi pour le nommage d'une collection en dernier recours.

Attention : les noms de collection ne peuvent pas être changés après leur création.

Une catégorie liée au nom de la collection est attribuée à la collection au moment de sa création.

CONVENTIONS DE NOMMAGE

Collections

Pour la nomenclature des collections, le groupe de travail sur la gestion des données de recherche recommande :

- l'omission des articles (le, la, etc.) avant le nom d'un laboratoire, d'un groupe ou d'un projet de recherche ou autre entité.

⁷ <https://zenodo.org/record/5579827#.YkintjWQhPY>

Exemples :

- Groupe de recherche sur les changements climatiques
- Laboratoire d'intelligence artificielle en sciences sociales
- pour les noms de chercheuses ou chercheurs, utiliser la forme : Nom de famille, Prénom.
- ne jamais indiquer « Université de Sherbrooke », celui-ci apparaît par défaut, entre parenthèses, à la suite du nom d'une collection.

Fichiers et ensembles de données

Par souci d'uniformité et pour faciliter le repérage des contenus, il est recommandé de nommer les fichiers et les ensembles de données en appliquant les règles proposées dans le document

« [Recommandations générales : Nommage des fichiers numériques](#) » produit par le SBA.

ANNEXE 1 - DEFINITIONS

Collection

Sous-ensemble situé sous le Dataverse institutionnel. Ces sous-ensembles héritent automatiquement du processus choisi pour ce dernier. La particularité d'une collection est qu'il est possible d'y configurer un schéma de métadonnées ainsi que des filtres de recherche qui sont différents de ceux utilisés par le Dataverse institutionnel.

Curation/médiation des données

De façon générale, « la curation des données est une gestion active des données de recherche à chaque étape de production, soit la création, la sauvegarde, l'utilisation, l'archivage, le partage et la réutilisation des données⁸. » À l'UdeS, la curation des données durant la vie active de celles-ci est de la responsabilité de l'équipe de recherche. Au moment du dépôt dans le Dataverse institutionnel, les personnes administratrices s'impliqueront dans le choix de la collection ainsi que dans la vérification de l'intégrité des fichiers, des métadonnées, de la documentation, du format des fichiers et de la licence.

Délégation

Action pour une professeure ou un professeur de nommer une autre personne pour agir à titre de personne déposante. La délégation se fait via le formulaire du dépôt de données et en le soumettant aux personnes qui administrent le Dataverse de l'Université de Sherbrooke.

Documentation

Ensemble des documents requis pour le dépôt des données de recherche. Inclut généralement un fichier Lisez-moi, un dictionnaire de données, les valeurs possibles pour chaque variable, les codes de traitement des données, etc.

Document administratif

Formulaire, certificat, document de planification ou autre document lié à la gestion du projet de recherche.

Ensemble de données

Ensemble des fichiers (fichiers de données, fichier « Lisez-moi », code, métadonnées, etc.) se rapportant aux données produites dans le cadre d'un projet de recherche. Typiquement, un ensemble de données contient plusieurs fichiers de données et plusieurs fichiers de documentation.

Fichier de données

Fichier unique contenant des données de recherche.

Métadonnées

Les métadonnées sont de l'information structurée qui décrit, explique, localise ou facilite autrement l'obtention, l'utilisation ou la gestion d'une ressource d'information. Les métadonnées sont souvent appelées données des données ou information sur l'information.

Personne déposante

Personne ayant sollicité et obtenu la création d'une collection en vue d'y déposer ses données de recherche (voir la sous-section Personnes autorisées à déposer). La personne déposante peut

⁸ Clary et al., 2020, p. 5.

ajouter des ensembles de données, modifier la licence et les conditions et éditer les autorisations sur les fichiers et répertoires.

Renseignement personnel

Tout renseignement qui concerne une personne physique et qui permet de l'identifier. Le nom d'une personne, pris isolément, n'est pas un renseignement personnel. Cependant, lorsque ce nom est associé ou jumelé à un autre renseignement visant cette même personne, il devient alors un renseignement personnel.

Voici des exemples de renseignement personnel :

- Le nom d'une personne et sa date de naissance;
- Matricule étudiant et/ou code permanent;
- Numéro d'assurance sociale;
- Numéro de carte de crédit;
- Numéro d'assurance maladie;
- Renseignement de nature médicale ou financière;
- Le nom d'une personne et son numéro de téléphone personnel;
- Le nom d'une personne et son adresse de domicile;
- Le nom d'une personne et ses résultats scolaires.

Schéma de métadonnées

Modèle composé d'un ensemble déterminé de champs de métadonnées, conçu dans un but spécifique. Un schéma de métadonnées spécifie généralement les noms des champs et la sémantique des renseignements contenus dans ces champs. Il peut également contenir des règles de syntaxe, c'est-à-dire des règles indiquant comment les champs et leurs contenus doivent être encodés. Plusieurs schémas de métadonnées ont été conçus pour une variété de disciplines et d'utilisateurs. Un des plus connus est le Dublin Core.

<https://statistique.quebec.ca/fr/fichier/etat-des-lieux-sur-les-metadonnees-relatives-aux-contenus-culturels-glossaire.pdf>

Téléversement et publication des données

Le téléversement est effectué par la personne déposante qui télécharge un ou plusieurs fichiers, qui constituent un ensemble de données. Après le téléversement des fichiers, ceux-ci ne seront visibles que par les personnes qui en ont l'autorisation (authentification requise). Il faudra que les fichiers soient publiés par la personne administratrice, selon le processus retenu, afin que ceux-ci soient disponibles pour tous en libre accès dans Dataverse.

ANNEXE 2 - FORMATS RECOMMANDÉS

Type de fichier	Formats recommandés
Audio	WAV ou AIFF FLAC MP3
Image	TIFF PNG JPG
Texte	PDF/A TXT XML CSV
Balisateur	XML HTML et fichiers associés : .css, .xslt, .js, .es
Transcription	PDF/A
Vidéo	MPEG-4
Analyse statistique	R SPSS STATA

Source : [UOttawa \(2021\). Instructions pour le dépôt dans UOttawa Dataverse.](#)

ANNEXES 3 - LICENCES

Ressources disponibles sur les types de licences

- [licences Creative Commons](#)
- [licences Open Data Commons](#)
- [licence ouverte / open licence d'Etalab](#)
- [licences pour des logiciels : GNU](#)

Pour en savoir plus sur les licences d'exploitation des données, veuillez consulter le [UK Digital Curation Centre](#).