

# 1. Modèle de cadre de curation des collections institutionnelles de Borealis

Rév. 2025-02-19

Rédigé par : Michael Steeleworthy et Erin Clary

Groupe de travail sur les politiques, Groupe d'experts Dataverse Nord

Réviseurs : Y.G. Rancourt et Robyn Stobbs. Groupe d'experts sur la préservation, Alliance de recherche numérique du Canada

Le présent cadre de curation de données des collections institutionnelle vise à aider les administratrices et administrateurs de Borealis et leurs établissements commanditaires (p. ex., bibliothèques universitaires) à établir des services de curation de données ou à améliorer ceux déjà en place. Il présente différents éléments de la curation (p. ex., formats de fichiers, métadonnées, qualité des données) et vous demande de juger de leur incidence potentielle sur vos pratiques de curation. De nombreux aspects de ce cadre rejoindront les politiques, pratiques, ressources et parties prenantes de votre établissement ou en dépendront. Il faudra les passer en revue lorsque vous établirez les objectifs de vos services de curation. Vous devrez peut-être aussi déterminer si d'autres parties prenantes ou changements opérationnels seront requis pour fournir le niveau de service souhaité.

Pour des précisions sur le processus de dépôt et d'évaluation, veuillez consulter les Directives sur le dépôt et l'évaluation des collections institutionnelles de Borealis (Borealis Institutional Collection Deposit & Appraisal Guidelines).

Les éléments matériels fondés sur ce modèle peuvent étayer vos réponses aux exigences du document 2023-2025 CTS figurant dans [le modèle de demande](#) : R08 Dépôt et évaluation, R10 Assurance qualité, et R13 Réutilisation

## 1.1. Rappel important

Toutes les politiques et procédures institutionnelles liées à votre collection institutionnelle dans Borealis devraient être complémentaires afin de faciliter la gestion de cette dernière. Leur harmonisation est une étape importante à ne pas négliger lors de leur élaboration. Veuillez consulter ce [diagramme pour une représentation visuelle des possibles interactions harmonieuses entre vos politiques et vos procédures de collections institutionnelles dans Borealis](#) afin de vous assurer de couvrir tous éléments requis ou nécessaires.

## 1. Modèle de cadre de curation des collections institutionnelles de Borealis

### 1.1. Rappel important

### 1.2. Introduction : La curation en tant que pratique

#### Scénarios de services de curation

## Niveaux de curation

- 1.3. Critères du cadre de curation de votre bibliothèque
  - 1.3.1. Pertinence de l'ensemble de données pour la collection
  - 1.3.2. Métadonnées et documents
  - 1.3.3. Intégrité des données et chaîne de contrôle
  - 1.3.4. Qualité des données
  - 1.3.5. Format des fichiers
  - 1.3.6. Organisation des fichiers et conventions pour leur donner un nom
  - 1.3.7. Sensibilité des données, droits d'auteur et conditions d'utilisation (gestion des risques et des droits)
  - 1.3.8. Données sous embargo
  - 1.3.9. Savoir traditionnel et données recueillies auprès de partenaires autochtones
  - 1.3.10. Logiciels et code
  - 1.3.11. Licences

## 1.2. Introduction : La curation en tant que pratique

La curation de données est la gestion active des données de recherche lors de leur création, leur maintenance, leur utilisation, leur archivage, leur partage et leur réutilisation<sup>1</sup>. Il s'agit d'un ensemble de processus automatisés et médiatisés par des spécialistes qui peuvent servir à ajouter de la valeur, à mieux respecter les principes selon lesquels les données doivent être faciles à trouver, accessibles, interopérables et réutilisables (principes FAIR)<sup>2</sup>, et à préparer les données pour leur préservation à long terme.

Les politiques et objectifs stratégiques qui régissent votre collection institutionnelle dans Borealis transparaissent souvent dans vos services de curation de données. Le fonctionnement de ces derniers — p. ex., quantité de services fournis et portée des activités de curation réalisées — dépendra de la mission et des ressources de votre organisation ainsi que des fonctionnalités de la plateforme d'hébergement du dépôt. Les bibliothèques doivent établir leurs motifs de curation des données, évaluer leur capacité à déployer des services de curation et déterminer les éléments de la curation les plus appropriés pour leurs paramètres opérationnels.

Le présent cadre comprend des éléments qu'une bibliothèque peut envisager d'ajouter à ses pratiques de curation<sup>3</sup>. Ceux-ci sont appuyés par les différents scénarios de services (curation non médiatisée, semi-médiatisée et médiatisée) et les trois niveaux pour les

---

<sup>1</sup> Groupe d'experts sur la curation du Réseau Portage (2020), *Introduction à curation de données*, Zenodo. <https://doi.org/10.5281/zenodo.4015595>

<sup>2</sup> M. D. Wilkinson, M. Dumontier, IJ. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers et B. Mons (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific Data*, vol. 3, n° 1, 160018. <https://doi.org/10.1038/sdata.2016.18>

<sup>3</sup> Pour en savoir plus sur les activités de curation de données, veuillez consulter : L. R. Johnston, J. Carlson, C. Hudson-Vitale, H. Imker, W. Kozlowski, R. Olendorf et C. Stewart (2016), *Definitions of Data Curation Activities used by the Data Curation Network*. <http://conservancy.umn.edu/handle/11299/188638>

diverses tâches de curation, comme résumé dans le Guide pour la curation dans Dataverse<sup>4</sup>. Le modèle cadre de curation de collections institutionnelles recourt à ces niveaux présentés ci-dessous. Les scénarios de service sont aussi tirés de ce guide, mais leur description a été légèrement modifiée afin d'inclure les activités de curation de données exigeant peu de communication, voire aucune, avec la personne ayant créé les données (sauvetage de données, données ouvertes). Ces scénarios de service et ces niveaux de curation peuvent aider votre bibliothèque à élaborer son propre modèle de services, car ils définissent et situent les étapes des pratiques de curation :

### Scénarios de services de curation

Curation non médiatisée	Aucune intervention du service de gestion des données de recherche (GDR) n'est requise. La chercheuse ou le chercheur crée ses propres collections et ensembles de données dans Borealis, téléverse ses données, ajoute les métadonnées et publie les collections/ensembles de données. Le service de GDR s'appuie sur les activités de curation automatisées du dépôt Borealis et peut proposer des séances de formation ou des éléments matériels à l'appui des dépôts.
Curation semi-médiatisée	Les chercheuses et chercheurs pourront créer des ensembles de données grâce à des permissions automatiques liées à leur affiliation ou grâce à des permissions ajoutées manuellement par le personnel administratif. Après avoir téléversé les fichiers et ajouté les métadonnées, les chercheuses et chercheurs présentent les ensembles de données au personnel administratif du service de GDR aux fins d'examen. Ce dernier recevra une notification lorsque les ensembles de données sont soumis pour examen et pourra soit les renvoyer aux auteurs leur demandant d'y apporter des modifications, soit les publier.
Curation médiatisée	Le service de GDR crée la collection (ou l'ensemble de données) dans Borealis. Ensuite, la ou le bibliothécaire assure la curation des données, puis les publie, sous réserve de l'approbation de la chercheuse ou du chercheur ou, dans certains cas (p. ex., sauvetage de données, curation de données ouvertes), du service de GDR.

### Niveaux de curation

Niveau 1	Minimum requis pour publier du contenu dans Borealis et faciliter la recherche d'ensembles de données (p. ex., soumission des ensembles de données dans la bonne collection et exactitude des métadonnées requises).
Niveau 2	Activités facilitant la découverte d'ensembles de données et améliorant leur utilisabilité au fil du temps (p. ex., champs de métadonnées recommandés remplis et ensembles de données comprenant suffisamment de documents pour

<sup>4</sup> A. Cooper, M. Steeleworthy, È. Paquette-Bigras, E. Clary, E. MacPherson, L. Gillis, L. Wilson et J. Brodeur (2021), *Guide pour la curation dans Dataverse*. <https://doi.org/10.5281/zenodo.5579827>. In English : <https://zenodo.org/record/5579820>.

	permettre à une utilisatrice ou à un utilisateur ayant un profil similaire de le comprendre ainsi que d'ouvrir et d'utiliser les fichiers).
Niveau 3	Activités de curation intensives visant à préparer les ensembles de données en vue de leur préservation et à augmenter les possibilités d'utiliser les données et le code pour reproduire ou répéter une étude connexe (p. ex., amélioration des documents connexes, révision du contenu des fichiers et du code et conversion de fichiers de données en formats appropriés pour la préservation à long terme).

### 1.3. Critères du cadre de curation de votre bibliothèque

Les éléments de curation présentés ci-dessous peuvent contribuer à l'élaboration ou à l'amélioration des pratiques de curation de votre bibliothèque. Notons que, pour bon nombre de ces points, vous devrez tenir compte des conditions d'utilisation de votre dépôt, de votre politique sur la collection, des ententes avec les déposantes et déposants, et de votre politique de préservation.

Les activités de curation doivent être transparentes et effectuées en collaboration avec la déposante ou le déposant. Si des activités sont automatisées (p. ex., conversion de formats de fichiers), il est recommandé de l'indiquer explicitement dans les documents destinés au public.

#### 1.3.1. Pertinence de l'ensemble de données pour la collection

*« Comment déterminerez-vous si un ensemble de données cadre avec votre collection ? »*

Pour déterminer la pertinence d'un ensemble de données pour votre dépôt, vous devez tenir compte de vos politiques et procédures, y compris les politiques sur les données de recherches, les collections générales, les archives ou les critères relatifs à la portée des dépôts complémentaires de votre organisation, ainsi que des services et ressources externes. Idéalement, la portée et l'étendue des collections du dépôt de données seront déterminées avant le lancement des services de dépôt afin d'uniformiser les pratiques et de faciliter la gestion des collections de données.

Comme de nombreuses politiques sur les collections, l'établissement des priorités permet à la bibliothèque de se concentrer sur des sujets particuliers ou d'harmoniser ses collections avec un mandat de service élargi. Vos critères peuvent accorder la priorité à des ensembles de données associés à une publication, à des ensembles de données originaux ayant une valeur intrinsèque, ou encore à des données associées à un domaine de recherche particulier. La pertinence peut aussi être déterminée en fonction d'autres facteurs, comme le lien entre les ensembles de données et l'établissement, l'état du financement ou la valeur historique.

Il est difficile de prévoir la valeur future d'un ensemble de données, car elle est liée non seulement à la recherche associée aux données et à la reproductibilité, mais également

aux systèmes techniques et organisationnels utilisés pour consulter, stocker et préserver les données<sup>5</sup>. Par conséquent, dans le cadre du processus décisionnel, il peut être utile de tenir compte du contexte de recherche actuel : à quel point ces données sont-elles utiles pour communiquer les connaissances et les perspectives actuelles ? Existe-t-il des indicateurs externes (p. ex., citations, subventions, prix, statistiques sur la consultation) ou des normes disciplinaires (p. ex., périodes de rétention) permettant de déterminer la valeur ainsi que son incidence sur la pertinence des données pour votre collection ? Portez une attention particulière aux données qui contribuent à représenter la recherche sur les populations sous-documentées ou marginalisées, ou réalisées par celles-ci, ou qui font état de changements dans un champ de pratique.

Ces domaines prioritaires et lignes directrices sur le dépôt doivent être mentionnés dans les conditions d'utilisation et les politiques de votre bibliothèque relatives aux collections, à la collecte de données et à la préservation, ou dans d'autres documents stratégiques, et publiés à un emplacement auquel les déposantes potentielles et déposants potentiels ont accès.

Pour en savoir plus, veuillez consulter notre Modèle de politique sur les collections institutionnelles de Borealis.

### 1.3.2. Métadonnées et documents

« *Quel est le niveau de curation approprié pour les métadonnées et les documents ?* »

L'enrichissement des métadonnées est peut-être l'activité de curation la plus évidente et la mieux comprise. Le dépôt Borealis procure de solides fonctions d'enrichissement des métadonnées, dont la plupart sont décrites dans le *Guide des pratiques exemplaires sur les métadonnées de Dataverse Nord*<sup>6</sup>, le *Guide pour la curation dans Dataverse* mentionné précédemment, et le guide d'utilisation de Borealis<sup>7</sup>. Votre bibliothèque doit établir ses propres normes et pratiques exemplaires minimales d'enrichissement des métadonnées et des documents. Ces décisions reposent en partie sur des considérations opérationnelles, comme les ressources dont vous disposez, votre modèle de service de curation (p. ex., entièrement ou partiellement médiatisé), et le degré ou niveau d'enrichissement des métadonnées et des documents recherché. Par exemple, une bibliothèque peut limiter ses activités de curation à l'amélioration des métadonnées de citations, tandis qu'une autre pourrait exiger l'amélioration des éléments de métadonnées qui facilitent la découverte et la reproductibilité par l'intermédiaire de mots-clés, de descriptions d'ensembles de données, de documents ouverts, de définitions des variables, entre autres.

Pour normaliser les pratiques, on recommande aux bibliothèques de se doter de leurs propres pratiques exemplaires encadrant leur travail lié aux métadonnées. Le Guide des

<sup>5</sup> B. Lavoie (2012), « Sustainable research data », dans G. Pryor (éd.), *Managing Research Data*, Facet Publishing, p. 67-82. Facet. doi:10.29085/9781856048910.005

<sup>6</sup> T. Bascik, P. Boisvert, A. Cooper, M. Gagnon, M. Goodwin, J. Huck, A. Leahey, K. Stathis et M. Steeleworthy (2021), *Guide des pratiques exemplaires sur les métadonnées de Dataverse Nord V. 3.0*. <https://doi.org/10.5281/zenodo.5576430>. In English : <https://doi.org/10.5281/zenodo.5576411>.

<sup>7</sup> Scholars Portal, (2022), *Guide d'utilisation de Borealis*. <https://learn.scholarsportal.info/fr/guides/borealis/>.

pratiques exemplaires sur les métadonnées de Dataverse Nord peut servir de point de départ, car il contient des renseignements sur les champs obligatoires du dépôt Borealis ainsi que des lignes directrices sur ses conventions relatives aux métadonnées.

Pour déterminer comment intégrer l'enrichissement des métadonnées et des documents à leurs pratiques de curation, les bibliothèques peuvent appliquer l'un des trois niveaux suivants :

1. Les métadonnées et les documents ne font pas l'objet d'un processus de curation. Il incombe aux déposantes et déposants de fournir des descriptions complètes et exactes des fichiers, de leur contenu et de leur utilisation.
2. À la discrétion du service de dépôt de données, les métadonnées et les documents peuvent faire l'objet d'un processus de curation pour vérifier leur exhaustivité et leur exactitude. L'enrichissement peut être évalué en fonction du niveau de curation exigé pour le dépôt (p. ex., ensembles de données publiables, amélioration de la découverte et de l'utilisabilité ou de la capacité de préservation).
3. Les métadonnées et les documents seront enrichis en fonction du niveau de curation exigé pour le dépôt. Les données ne seront publiées que si tous les champs de métadonnées requis sont remplis et les documents nécessaires sont fournis.

Des critères peuvent également être définis pour appliquer différents niveaux à différentes sous-collections ou ensembles de données.

### 1.3.3. Intégrité des données et chaîne de contrôle

*« Quelle est l'importance de la chaîne de contrôle pour votre bibliothèque ? »*

La provenance des métadonnées liées aux processus d'intégration, y compris le nom et le courriel de la personne ayant soumis les données et la date de soumission, est automatiquement consignée dans le dépôt Borealis, mais votre bibliothèque devra déterminer l'importance de la chaîne de contrôle des ensembles de données avant le dépôt et vérifier l'intégrité des fichiers. La consignation de la chaîne de contrôle d'un ensemble de données permet de s'assurer que les modifications sont autorisées et justifiées, et contribue à la transparence et à la redevabilité.

Lors du processus de dépôt de données pour un service de curation médiatisé, par exemple, la consignation pourrait notamment permettre d'établir les méthodes privilégiées ou exigées pour le transfert des données dans le dépôt afin de réduire les risques liés à l'intégrité, à l'exhaustivité et à la provenance des ensembles de données. Par exemple, elle pourrait exiger qu'une déposante ou un déposant verse le contenu directement dans le système Borealis, ou encore que le contenu soit soumis au personnel du dépôt par transfert médiatisé dans des dossiers infonuagiques, dans des systèmes de transfert de fichier sécurisés, ou même par courriel.

L'imposition d'une méthode de transfert donnée peut parfois favoriser l'application de pratiques exemplaires, mais cette mesure pourrait ne pas convenir à toutes les tailles d'ensembles de données, ce qui pourrait contrarier les utilisatrices et utilisateurs. Par exemple, une convention de transfert de fichiers comme BagIt permettrait d'automatiser la

collecte de métadonnées essentielles au point de transfert, mais peut créer un obstacle perçu comme inutile par les déposantes et déposants qui ignorent l'importance des métadonnées et de la fixité pour le transfert et la curation de fichiers.

Une fois les données reçues, les activités de curation menées par le personnel et les déposantes et déposants peuvent également déboucher sur des changements intentionnels visant à améliorer la découverte, l'accès et l'utilisation des ensembles de données. Un point d'attention particulier peut être l'examen d'éléments matériels afin de s'assurer que les informations contextuelles pertinentes, telles que les exigences logicielles, les sources de données tierces, les licences et les accords, ainsi que les restrictions d'accès, sont fournies. Votre bibliothèque peut tenir un registre de curation pour consigner la nature et la raison des modifications apportées avant la publication. Votre bibliothèque peut choisir de consigner les activités de curation liées à des fichiers individuels dans Borealis par le biais de fichiers de provenance ou de métadonnées, ou de conserver des journaux dans un système externe.

Pour déterminer comment intégrer la chaîne de contrôle et l'intégrité des données aux pratiques de curation, les bibliothèques peuvent appliquer l'un des trois niveaux suivants :

1. L'intégrité des données et la chaîne de contrôle ne seront pas évaluées dans le cadre du processus de curation. Le service de dépôt assurera la curation des fichiers sans remettre en question leur fixité ni leur provenance.
2. À la discrétion du service de dépôt de données, l'intégrité et la chaîne de contrôle des ensembles de données pourraient être vérifiées. Si cette évaluation entraîne des changements dans les ensembles de données, des journaux de curation sont créés et conservés.
3. Tous les ensembles de données doivent être versés dans le dépôt conformément au modèle de transfert établi, et leur intégrité et leur chaîne de contrôle seront vérifiées. Toutes les évaluations seront consignées dans les journaux de curation, que des modifications y aient été apportées ou non.

#### 1.3.4. Qualité des données

*« Dans quelle mesure votre bibliothèque évaluera-t-elle la qualité des données ? »*

La mesure dans laquelle une bibliothèque évaluera la qualité des données reçues dépend de son expertise dans le domaine de recherche et de ses ressources. Souvent, mais pas toujours, les dépôts institutionnels recevront des données qui ont déjà été soumises à un nettoyage, à une analyse des procédures et à un contrôle de la qualité rigoureux. Les dépôts institutionnels de données peuvent collaborer avec des groupes de recherche pour vérifier la qualité des ensembles de données reçus avant de les verser dans le dépôt, s'en remettant aux spécialistes pour gagner du temps dans le processus de curation. Une telle décision opérationnelle tiendrait compte des limites du dépôt relatives aux sujets, et laisserait aussi plus de temps pour les pratiques de curation liées à l'accès, à la découverte et à la préservation.

Pour déterminer comment évaluer la qualité des données dans le cadre du processus de curation, les bibliothèques peuvent appliquer l'un des trois niveaux suivants :

1. La qualité des données n'est pas évaluée dans le cadre du processus de curation. Les déposantes et déposants sont tenus de veiller à ce que les données et documents déposés puissent être compris par leurs pairs et soient prêts pour l'utilisation et la réutilisation.
2. À la discrétion du service de dépôt, la qualité des données peut être évaluée dans le cadre du processus de curation. Une évaluation de la qualité des données pourra être réalisée, selon les ressources, la valeur des ensembles de données et l'expertise de l'équipe de curation dans le domaine. Elle peut donner lieu à des recommandations ou à des demandes de nettoyage des données en vue du processus d'intégration.
3. Toutes les données seront évaluées dans le cadre du processus de curation. Elle peut donner lieu à des recommandations ou à des demandes de nettoyage des données en vue du processus d'intégration.

### 1.3.5. Format des fichiers

« *Quelles seront les conventions de votre bibliothèque quant aux formats de fichiers acceptés ?* »

La bibliothèque doit déterminer dans quelle mesure elle autorisera, transformera ou rejettera les fichiers de format fermé ou exclusif. Les formats pouvant être intégrés auront une incidence opérationnelle et stratégique, surtout si votre dépôt préserve le contenu à long terme à des fins de réutilisation et de reproductibilité. Le fait d'accepter seulement les formats libres facilitera la préservation et accroîtra les chances que le contenu de la collection du dépôt puisse être utilisé à long terme, mais pourrait prolonger le processus de curation, alourdir le fardeau des curatrices et curateurs et des déposantes et déposants, et entraîner une réduction de la demande de services de dépôt. Inversement, un service de dépôt qui permet aux déposantes et déposants de partager des formats fermés ou exclusifs peut initialement réduire le fardeau des déposantes et déposants ou des curatrices et curateurs, mais au risque d'assurer l'intendance de données qui deviendront inutilisables avec le temps. Il peut être nécessaire d'offrir une certaine flexibilité, de connaître les formats largement employés dans les différentes disciplines, et de savoir si les logiciels à utiliser pour extraire, transformer ou visualiser les données dans un format donné sont en développement actif : vous pourrez ainsi déterminer si ce type de fichier fermé continuera d'être pris en charge à court terme.

Si vous prévoyez transformer des fichiers, vous pourriez avoir à collaborer avec les déposantes ou les déposants au point d'intégration pour éviter de perdre de l'information au cours du processus de conversion. Cela peut aussi être nécessaire si vous n'avez pas de logiciel pour visualiser et exporter le contenu dans un autre format. Il est recommandé de publier une liste des formats privilégiés pour indiquer qu'ils sont pris en charge par votre bibliothèque et que les autres formats ne bénéficieront pas du même niveau de service. Si vous n'avez pas encore publié de liste des formats privilégiés, le tableau de la Bibliothèque du Congrès sur la [viabilité des formats numériques](#)<sup>8</sup> est une bonne ressource pour constater

---

<sup>8</sup> Bibliothèque du Congrès (2004), *Sustainability of Digital Formats : Planning for Library of Congress Collections*. <https://www.loc.gov/preservation/digital/formats/>

les effets du temps et des changements technologiques sur l'accessibilité des formats de fichiers.

Pour bien évaluer l'incidence des formats de fichiers sur leurs pratiques de curation, les bibliothèques ayant recours au dépôt Borealis doivent savoir comment ce dernier gère les données tabulaires (p. ex., feuilles de calcul, fichiers de données SPSS) et les fichiers Microsoft Excel. Au moment du téléversement, le dépôt Borealis convertit les données tabulaires — XLS, SAV, CSV, etc. — en fichiers de données tabulaires textes non exclusifs (TAB) et crée des métadonnées de citations et des métadonnées DDI pour les variables. Ce processus d'ingestion tabulaire favorise autant la préservation que la découverte et la présentation dans l'outil d'exploration de données, une application web intégrée pour l'exploration et la curation de variables dans le Dataverse. Le fichier TAB intégrera et affichera seulement la première feuille ou le premier onglet du fichier de calcul. Par conséquent, les fichiers Excel comprenant plusieurs feuilles ou une mise en forme perdront des données et du contexte au cours du processus de conversion en format TAB. Le dépôt Borealis conserve tous les originaux des données tabulaires à des fins de consultation et de téléchargement, mais les curatrices et curateurs doivent connaître ce processus de conversion en format TAB et l'apparence du fichier TAB créé dans l'ensemble de données.

Pour déterminer l'incidence potentielle des formats de fichiers sur le processus de curation, les bibliothèques peuvent appliquer l'un des trois niveaux suivants :

1. Le service de dépôt n'assure pas la curation des formats de fichiers, et fournit ou non des lignes directrices sur les formats de fichiers privilégiés aux déposantes et déposants ; s'il en fournit, les déposantes et déposants doivent soumettre les données dans les formats privilégiés, mais les formats non conformes déposés seront également intégrés. Aucune conversion en formats libres ne sera effectuée.
2. À la discrétion du service de dépôt, les formats de fichiers exclusifs, fermés ou courants peuvent être acceptés. De plus, une conversion peut être effectuée au cours de ce processus ; dans ce cas, les formats originaux seront conservés à des fins de consultation.
3. Le service de dépôt accepte seulement les formats de fichiers ouverts ou les fichiers qui peuvent être convertis en formats libres. Au cours du processus de conversion, le format de tous les fichiers sera vérifié. De plus, une conversion peut être effectuée au cours de ce processus ; dans ce cas, les formats originaux seront conservés à des fins de consultation. Le service pourrait demander aux déposantes et déposants de remplacer les fichiers non ouverts par des fichiers ouverts.

On recommande aux bibliothèques de se référer aux [introductions à la curation de données du Data Curation Network](#) pour normaliser la gestion des formats de fichier durant le processus de curation.

### 1.3.6. Organisation des fichiers et conventions pour leur donner un nom

*« Quelles seront les conventions pour hiérarchiser et nommer les fichiers et dossiers de votre bibliothèque ? »*

Comme pour les formats de fichiers, la bibliothèque doit déterminer les conventions

imposées ou privilégiées pour organiser et nommer les fichiers, et comparer les avantages de la normalisation avec les risques que pose l'ajout d'un fardeau pour les déposantes et déposants et les curatrices et curateurs. Idéalement, l'approche du stockage de données numériques doit être structurée, transparente et prévisible, de sorte qu'il soit facile d'évaluer l'exhaustivité des données, de les récupérer en cas de perte ou d'erreur, et de les consulter, autant pour les êtres humains que pour les machines.

Une approche uniforme des conventions pour structurer et nommer les dossiers suffit pour atteindre un certain degré de contrôle et de transparence, mais il existe aussi des spécifications, notamment la spécification [Oxford Common File Layout \(OCFL\)](#), qui encadrent l'accès au stockage de fichiers sans logiciel.

L'organisation des fichiers et les conventions de nommage peuvent dépendre des conventions du dépôt, des préférences des déposantes et déposants, du code associé renvoyant à un fichier statique, ou d'une publication qui contient déjà un nom de fichier. Pour déterminer l'incidence que peuvent avoir les conventions pour organiser et nommer les fichiers sur le processus de curation, les bibliothèques peuvent appliquer l'un des trois niveaux suivants :

1. Le service de dépôt n'appliquera aucune convention pour organiser et nommer les fichiers lors du processus de curation.
2. À sa discrétion, le service de dépôt pourra appliquer des conventions pour organiser et nommer les fichiers lors du processus de curation ; des modifications pourront être recommandées ou exigées avant la fin de l'intégration.
3. Des conventions seront appliquées pour organiser et nommer les fichiers d'ensembles de données lors du processus de curation. Des modifications pourront être recommandées ou exigées avant la fin de l'intégration.

Les bibliothèques devraient envisager de publier des guides destinés au public pour informer les déposantes et déposants des exigences et de leur application (p. ex., guides de l'Université d'Ottawa : [File naming and organization of data](#) / [Nommage de fichiers et gestion de versions](#)).

### 1.3.7. Sensibilité des données, droits d'auteur et conditions d'utilisation (gestion des risques et des droits)

*« Comment votre bibliothèque gèrera-t-elle les données sensibles, les droits d'auteur et les éléments connexes ? »*

Les activités de curation de votre bibliothèque visant les données sensibles doivent être balisées par la capacité du dépôt Borealis à gérer celles-ci de manière sécuritaire. Votre service de dépôt et ses curatrices et curateurs doivent veiller à ce que les déposantes et déposants ne versent pas de données contenant des renseignements identificatoires sur des sujets humains ou des renseignements réputés sensibles ou confidentiels. Les services de dépôt doivent aussi veiller à ce que les données protégées par des droits d'auteur, le contenu appartenant à des tiers et les données pouvant contrevenir aux conditions d'utilisation de Borealis ne soient pas versées dans le dépôt, ou se doter de procédures pour atténuer et résoudre de tels problèmes.

Avant la création de son service de dépôt, la bibliothèque doit déterminer à qui il incombe de s'assurer que le contenu publié n'est pas sensible et ne contrevient pas aux conditions d'utilisation locales ou du fournisseur (p. ex., Borealis). Les bibliothèques peuvent élaborer des conditions d'utilisation conférant la responsabilité aux déposantes et déposants au moment du téléversement, ce qui peut accélérer le processus de curation, mais exigera une intervention en aval dans le cas des ensembles de données sensibles ou contrevenant aux conditions d'utilisation. Elles peuvent aussi prévoir un processus de curation plus élaboré comprenant des étapes supplémentaires pour veiller à ce que les déposantes et déposants et les ensembles de données respectent ces conditions, par exemple :

1. Les déposantes et déposants doivent accepter les ententes les visant ou les conditions d'utilisation de votre dépôt avant de commencer une nouvelle demande.
2. À la discrétion du service de dépôt, une curatrice ou un curateur peut examiner les documents ou les fichiers pour s'assurer que le contenu ne contrevient pas aux conditions d'utilisation du dépôt. Cet examen peut être effectué lorsqu'une curatrice ou un curateur détermine que la confirmation de la présence (ou du retrait) de données sensibles contrevenant aux conditions d'utilisation est requise (p. ex., d'après le titre ou la description de l'étude, le format de fichier ou la discipline).
3. À la discrétion de la bibliothèque, une curatrice ou un curateur peut demander à obtenir et examiner une entente de recherche ou un formulaire de consentement de participante ou participant vierge pour confirmer que les données peuvent être publiées et partagées.
4. Le contenu publié qui contrevient aux conditions d'utilisation du dépôt sera soumis à un processus établi pour gérer les demandes de retrait, les restrictions ou l'archivage.

Les bibliothèques devraient prendre en compte le document [Joint FORCE11 & COPE Research Data Publishing Ethics Working Group Recommendations](#)<sup>9</sup>, qui décrit les pratiques éthiques de partage et de publication des données de recherche.

### 1.3.8. Données sous embargo

« Dans quelle mesure votre bibliothèque acceptera-t-elle et gèrera-t-elle les données sous embargo ? »

L'élaboration de lignes directrices sur la gestion des embargos pour votre dépôt permettra de normaliser vos pratiques de curation et facilitera la gestion des attentes des déposantes et déposants relatives au service. Le dépôt Borealis comprend des fonctions automatisées d'embargo qui réduisent le fardeau des curatrices et curateurs quant à la gestion des dates de publication, mais votre service de dépôt doit d'abord déterminer si les embargos feront partie de ses pratiques. Une bibliothèque peut décider que son dépôt de données sera entièrement ouvert, comme sa collection générale, ou encore autoriser des embargos d'une durée limitée pour répondre aux besoins des déposantes et déposants ou des éditrices ou éditeurs.

---

<sup>9</sup> I. Puebla, D. Lowenberg et Force11 Research Data Publishing Ethics WG (2021), Joint FORCE11 & COPE Research Data Publishing Ethics Working Group Recommendations, Zenodo. <https://doi.org/10.5281/zenodo.5391293>

Toutes les versions du dépôt Borealis depuis la version 5.8 permettent aux curatrices et curateurs d'appliquer aux fichiers des embargos d'une durée limitée. Il faut savoir qu'après la publication des fichiers sous embargo, il est impossible d'écourter ou de lever ce dernier : toute consultation publique est suspendue jusqu'à la fin de l'embargo.

Pour déterminer si les embargos feront partie du processus de curation, les bibliothèques peuvent appliquer l'un des deux niveaux suivants :

1. Le service de dépôt n'accepte pas les embargos.
2. À la discrétion des curatrices et curateurs et en consultation avec les déposantes et déposants, une période d'embargo peut être accordée. Celle-ci sera déterminée par le service de dépôt en fonction des besoins des déposantes et déposants, conformément aux pratiques établies.

### 1.3.9. Savoir traditionnel et données recueillies auprès de partenaires autochtones

« Dans quelle mesure vos pratiques de curation respecteront-elles la souveraineté des données autochtones ? »

De plus en plus d'instruments de politique affirment les droits moraux des peuples autochtones en matière de souveraineté des données — y compris la création, la propriété et l'intendance des données et des savoirs — ainsi que l'importance, pour les chercheuses colonisatrices et les chercheurs colonisateurs, ainsi que leurs organisations, de reconnaître et respecter ces droits. Par exemple, la [Déclaration des Nations Unies sur les droits des Peuples Autochtones](#)<sup>10</sup>, les [appels à l'action](#) de la Commission de vérité et réconciliation<sup>11</sup>, la Politique des trois organismes sur la gestion des données de recherche<sup>12</sup>, les [principes de PCAP® \(propriété, contrôle, accès et possession\) des Premières Nations](#) du Centre de gouvernance de l'information des Premières Nations<sup>13</sup> et les [principes CARE](#) (avantage collectif, pouvoir de contrôle, responsabilité, éthique) de la Global Indigenous Data Alliance<sup>14</sup> soulignent l'importance de respecter le contrôle que la communauté autochtone exerce sur ses collections ainsi que sur l'analyse, le stockage, l'archivage et la consultation de ses données.

Quelles que soient les pratiques adoptées, votre service de dépôt doit pouvoir gérer les données de recherche autochtones dans le cadre de son processus de curation. Conformément à la Politique des trois organismes sur la gestion des données de recherche,

---

<sup>10</sup> Assemblée générale des Nations Unies (2 octobre 2007), *Déclaration des Nations Unies sur les Droits des Peuples Autochtones*, A/RES/61/295. <https://undocs.org/A/RES/61/295>

<sup>11</sup> Commission de vérité et réconciliation du Canada (2012, 2015), *Commission de vérité et réconciliation du Canada : appels à l'action*. <https://publications.gc.ca/site/fra/9.801240/publication.html>

<sup>12</sup> Gouvernement du Canada (15 mars 2021), *Politique des trois organismes sur la gestion des données de recherche*, [https://www.science.gc.ca/eic/site/063.nsf/fra/h\\_97610.html](https://www.science.gc.ca/eic/site/063.nsf/fra/h_97610.html) (consulté le 2 mars 2022).

<sup>13</sup> Centre de gouvernance de l'information des Premières Nations (s.d.), *Les principes de PCAP® des Premières Nations*. <https://fnigc.ca/fr/les-principes-de-pcap-des-premieres-nations/>.

<sup>14</sup> Global Indigenous Data Alliance (2019), *CARE Principles for Indigenous Data Governance*. <https://www.gida-global.org/care>

les établissements d'enseignement postsecondaire admissibles à des subventions — et donc leur bibliothèque — doivent reconnaître que « les données créées dans le contexte de la recherche menée « par et avec » les collectivités, les regroupements collectifs et les organisations des Premières Nations, des Métis et des Inuit seront gérées selon des principes élaborés et approuvés par ces intervenants, et en partenariat avec ces derniers »<sup>15</sup>. Au minimum, votre service doit s'informer des ententes officielles et informelles entre les chercheuses et chercheurs et les partenaires autochtones qui régissent le partage, la propriété, la consultation et l'utilisation des données ; vous pourriez aussi choisir d'examiner tout document pertinent avant l'intégration des données. Certaines de ces ententes peuvent avoir une incidence sur votre processus de curation (p. ex., processus d'enrichissement des métadonnées, documents requis et emplacement de stockage des données) et sur les dispositions relatives à la consultation et à l'utilisation en aval.

On recommande aux curatrices et curateurs de données et aux gestionnaires de services de dépôt du Canada de suivre le cours du Centre de gouvernance de l'information des Premières Nations sur les principes de PCAP des Premières Nations. Le personnel de bibliothèque et les curatrices et curateurs doivent aussi consulter leurs bureaux de recherche et des affaires autochtones et collaborer avec leurs partenaires de recherche autochtones dans un esprit de réconciliation.

### 1.3.10. Logiciels et code

*« Dans quelle mesure votre bibliothèque assurera-t-elle la curation des logiciels et du code associés aux données ? »*

La mesure dans laquelle votre bibliothèque assurera la curation des logiciels et du code associés aux ensembles de données dépend de l'expertise de votre équipe de curation en la matière, de sa connaissance des langages de programmation ou de la syntaxe propre aux applications, de même que de l'importance de la reproductibilité ou de la réutilisation pour le dépôt. La reproductibilité consiste à analyser à nouveau des données pour confirmer les recherches précédentes<sup>16</sup>. Les normes de reproductibilité peuvent varier d'un domaine à l'autre, mais il faut généralement disposer des données, du code entier et des documents sur la collecte, l'instrumentation et le traitement.

Un dépôt qui assure la curation de données de recherche associées à une publication peut permettre la vérification des analyses citées dans l'article. Encore mieux, un ensemble de données contenant des données brutes et leur équivalent après le traitement, le code annoté qui a servi à nettoyer et à traiter les données brutes, ainsi que des renseignements sur la collecte et l'instrumentation des données se rapprochent des normes de reproductibilité.

Le code nécessaire pour exécuter, nettoyer et traiter les données à des fins d'analyse doit être joint à l'ensemble de données déposé afin de faciliter la vérification des données et des méthodes. Les logiciels de curation présentent leurs propres contraintes liées au matériel et au système d'exploitation. Il faut savoir que les logiciels et le code peuvent désigner

<sup>15</sup> [https://www.science.gc.ca/eic/site/063.nsf/fra/h\\_97610.html](https://www.science.gc.ca/eic/site/063.nsf/fra/h_97610.html)

<sup>16</sup> I. V. Pasquetto, B. M. Randles et C. L. Borgman (2017), *On the Reuse of Scientific Data*, *Data Science Journal*, vol. 16, n° 8. <http://doi.org/10.5334/dsj-2017-008>

différents objets, selon les domaines, et que certains domaines et certaines études ne nécessitent pas de logiciel ni de code.

Pour déterminer le processus de curation nécessaire pour les logiciels et le code, la bibliothèque peut appliquer l'un des quatre niveaux suivants :

1. Le service de dépôt accepte les logiciels et le code, mais n'assure pas leur curation. On s'attend à ce que les déposantes et déposants soumettent un code et des logiciels libres ou fournissent des documents décrivant en détail les méthodes de collecte et de traitement des données.
2. Le service de dépôt accepte le code et les logiciels libres. À sa discrétion, il assumera la curation des logiciels et du code de certains ensembles de données afin de vérifier leur exactitude et leur exhaustivité. Dans certains domaines, tous les logiciels et les codes subiront le processus de curation, selon les forces de la collection et les exigences du service de dépôt.
3. Tout logiciel ou code déposé subira le processus de curation afin de vérifier son exactitude et son exhaustivité.
4. Toutes les données soumises au service de dépôt doivent être accompagnées des logiciels et du code utilisés. Ces objets subiront le processus de curation afin de vérifier leur exactitude et leur exhaustivité.

### 1.3.11. Licences

« *Quelles licences seront exigées ou recommandées pour votre dépôt ?* »

La bibliothèque doit tenir compte des licences qu'elle veut et peut prendre en charge. Prévoyez-vous recommander ou exiger des licences particulières ? Est-ce que les déposantes et déposants disposeront d'une liste d'options, ou vous les autoriserez à utiliser leurs propres licences ou conditions d'utilisations ?

Les curatrices et curateurs doivent connaître les licences Creative Commons préchargées dans Borealis ainsi que d'autres instruments d'attribution de licence (p. ex. MIT, licence publique générale GNU). Si le nombre d'options est restreint, il sera plus facile pour les curatrices et curateurs de recommander une licence, pour les déposantes et déposants de faire un choix, et pour les utilisatrices et utilisateurs de trouver les conditions d'utilisation qui s'appliquent à tout ensemble de données téléchargé à partir de votre dépôt ; il vous sera également plus facile de gérer les actifs numériques à long terme. Cependant, la restriction des licences pourrait vous obliger à rejeter certains ensembles de données. Les données fournies aux déposantes et déposants par un tiers ou dérivées de données déjà publiées peuvent être assujetties à des conditions d'utilisation particulières et nécessiter une licence conforme à celles-ci. De même, les données recueillies en collaboration avec des partenaires autochtones ou du secteur privé peuvent parfois n'être partagées que si certaines conditions ou restrictions sont appliquées aux ensembles de données. Les logiciels présentent d'autres défis, puisque les licences Creative Commons recommandées par de nombreux dépôts ne conviennent généralement pas pour le code et les logiciels<sup>17</sup>.

---

<sup>17</sup> Creative Commons ne recommande pas d'utiliser ses licences pour le code et les logiciels. Pour en savoir plus, voir : Creative Commons (s.d.), *Frequently Asked Questions: Can I apply a Creative*

Pour aider une déposante ou un déposant à choisir la licence qui convient, les curatrices et curateurs devront peut-être examiner les données et les documents, les conditions d'utilisation des données ou du code fournis par un tiers ou dérivés de ses données, et les ententes officielles et informelles de recherche et de partage des données. Comme pour les données sensibles, vous devrez déterminer qui devra veiller à ce que le contenu publié respecte vos conditions d'utilisation ou celles du fournisseur (p. ex., Borealis). Les bibliothèques peuvent établir des conditions d'utilisation conférant la responsabilité aux déposantes et aux déposants au moment du téléversement.

Pour déterminer les types de licences autorisés et les responsabilités du service de curation relatives à l'examen du contenu des déposantes et déposants et à la sélection de la licence appropriée, les points suivants pourraient vous être utiles :

1. Les déposantes et déposants doivent accepter les ententes les visant ou les conditions d'utilisation de votre dépôt avant de commencer une nouvelle demande.
2. À la discrétion du service de dépôt, une curatrice ou un curateur peut collaborer avec les déposantes et déposants pour sélectionner une licence qui convient à leurs ensembles de données parmi les licences publiques ou préchargées dans le dépôt Borealis.
3. À la discrétion du service de dépôt, une curatrice ou un curateur peut examiner les documents ou les fichiers pour s'assurer que le contenu n'enfreint pas les droits de propriété intellectuelle ou les accords d'utilisation des données. Les curatrices et curateurs peuvent aussi collaborer avec les déposantes et déposants pour bien citer les sources tierces ou créer des licences personnalisées conformes aux conditions d'utilisation des différentes sources de données.
4. Le contenu publié qui contrevient aux conditions d'utilisation du dépôt sera soumis à un processus établi pour gérer les demandes de retrait, les restrictions ou l'archivage.