



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada

Guide d'évaluation pour la préservation des données de recherche

Rédaction par le groupe de travail sur l'évaluation en vue de la préservation :
Jonathan Dorey (Institut national de la recherche scientifique, École nationale
d'administration publique et Université TÉLUQ)
Grant Hurley (Scholars Portal, Université de Toronto)
Beth Knazook (Alliance de recherche numérique du Canada)

Février 2022



Table des matières

Objectif du guide	3
1. Définitions	4
2. Parties prenantes de l'évaluation	4
3. Concepts et pratique d'évaluation et de sélection	5
4. Processus d'évaluation	7
5. Critères d'évaluation	9
6. Résultats de l'évaluation	14
Sources et citations	15
Remerciements	17



Objectif du guide

L'évaluation et la sélection sont des activités essentielles pour l'intendance responsable des données de recherche. Les données n'ont pas forcément une valeur de recherche à long terme et le volume croissant de données produites et publiées pour répondre à des besoins à court et moyen terme crée un fardeau à la fois pour les dépôts qui stockent et maintiennent l'accès aux ressources ainsi que pour les chercheurs à la recherche de données de qualité. Les pratiques d'évaluation des dépôts, souvent effectuées dans le processus de curation au moment de la consignation des données afin d'optimiser celles-ci pour le partage et la réutilisation, doivent mieux prendre en compte la pérennité des pratiques [FAIR](#) en matière de données¹.

En plus de l'incertitude quant à la valeur à long terme de certains ensembles de données, une grande partie des données produites aujourd'hui pourraient malheureusement ne pas rester accessibles et réutilisables selon la manière dont elles ont été consignées à l'origine. Les changements de tendances en matière de réutilisation pourraient nécessiter que les formats de fichiers, les métadonnées du dépôt et la documentation des ensembles de données soient actualisés ou révisés pour répondre aux besoins de la communauté désignée d'un dépôt. Les bonnes pratiques d'évaluation sont donc une composante inestimable du processus de curation, car elles peuvent guider les futures activités de préservation et les décisions de conservation. Idéalement, elles éclairent l'approche des chercheurs quant au partage de l'information, mais surtout, elles déterminent les décisions d'acquisition des dépôts et l'affectation continue des ressources pour maintenir l'accès aux données aussi longtemps que nécessaire. Des processus formels d'évaluation et de sélection font en sorte que les dépôts soient transparents et réalistes dans leur prise de décision concernant les données à accepter, la meilleure façon de les partager et la durée de leur préservation.

Ce guide est conçu pour être utilisé parallèlement aux politiques des établissements en matière d'acquisition, d'agrandissement des collections, de préservation et de retrait, ainsi qu'à d'autres documents stratégiques institutionnels généraux, afin d'aider les curateurs à travailler avec les chercheurs et les spécialistes de la préservation pour évaluer les données de recherche en vue de leur préservation à long terme. Pour en savoir plus sur l'élaboration de politiques, consultez la documentation [Policy Tools and Guidance](#) du Digital Curation Centre ou le guide [Policy-making for Research Data in Repositories : A Guide](#) de DataShare².

¹ Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3, no. 1 (2016): 1-9, consulté le 28 janvier 2022, <https://doi.org/10.1038/sdata.2016.18>.

² "Policy Tools and Guidance," *Digital Curation Centre*, consulté le 28 janvier 2022, <https://www.dcc.ac.uk/guidance/policy/policy-tools-and-guidance>; Ann Green, Stuart Macdonald, and Robin Rice, *Policy-making for Research Data in Repositories: A Guide*, v. 1.2 (Edinburgh: DISC-UK DataShare et EDINA et University Data Library, Université d'Édimbourg, mai 2009), <https://www.coar-repositories.org/files/guide.pdf>.



1. Définitions

L'*évaluation* pour la préservation est le processus servant à déterminer si un ensemble de données a une valeur archivistique à long terme suffisante pour mériter le travail de surveillance, de gestion, de stockage et de maintien d'accès à ces données ainsi qu'aux systèmes et flux de travail connexes de manière pérenne.

On parle de *réévaluation* lorsqu'une nouvelle évaluation est effectuée à une date ultérieure pour déterminer si un ensemble de données conserve sa valeur archivistique.

La *sélection* est le processus servant à déterminer les fichiers d'un ensemble de données qui doivent être conservés après que la valeur archivistique à long terme a été établie.

La *préservation* décrit la "série d'activités gérées pour garantir un accès continu aux matériels numériques aussi longtemps que nécessaire"³.

2. Parties prenantes de l'évaluation

Les créateurs de données peuvent contribuer au processus d'évaluation en réfléchissant à ce qui sera nécessaire pour que d'autres puissent comprendre et reproduire leurs recherches. Ils doivent faire en sorte que leurs fichiers soient bien organisés et décrits ou contextualisés afin de soutenir le travail des curateurs de données qui prennent souvent des décisions d'évaluation au nom des dépôts et des établissements responsables. Les créateurs de données peuvent également s'adresser aux dépôts pour toute question quant aux durées de conservation et aux intentions de préservation en fonction des exigences des bailleurs de fonds et de leur connaissance approfondie du domaine de recherche qui doivent être documentées dans toute décision d'évaluation du dépôt.

Les curateurs de données travaillent avec les créateurs de données pour prendre des décisions d'évaluation et de sélection fondées sur la qualité des données disponibles et le contexte de la recherche que les données soutiennent. Ils peuvent collaborer avec les créateurs de données pour améliorer le potentiel de réutilisation durable d'un ensemble de données pouvant inclure notamment l'ajout d'une description plus détaillée et la contextualisation des données pour la communauté de recherche au sens large. Lorsque les curateurs examinent les ensembles de données au moment de leur consignation dans un dépôt, ils doivent également tenir compte des priorités, des politiques en matière de collection et des normes de ce dépôt ainsi que des ressources qui pourraient être nécessaires pour garantir un accès pérenne.

³ « Glossary - Digital Preservation », Digital Preservation Coalition, consulté le 28 janvier 2022, <https://www.dpconline.org/handbook/glossary#D>.



Les préservateurs⁴ travaillent avec les dépôts pour stocker les données et maintenir l'accès à celles-ci aussi longtemps que nécessaire. Leurs activités comprennent la création de métadonnées de préservation, la migration des fichiers lorsque les formats originaux deviennent problématiques ou obsolètes, la duplication des données entre les lieux de stockage et la vérification de l'intégrité de ces données au fil du temps.

Les chercheurs et les intervenants de la communauté sont les personnes qui accèdent aux données et les réutilisent. Elles peuvent informer les curateurs et les spécialistes de la préservation à propos de l'évolution des besoins et des contextes dans le temps.

3. Concepts et pratiques d'évaluation et de sélection

L'évaluation fait depuis longtemps l'objet de discussions et de pratiques dans le domaine des archives. En procédant à l'évaluation, les archivistes reconnaissent que leur rôle n'est pas neutre et que leurs décisions auront une incidence sur les futurs chercheurs et la préservation de la mémoire enregistrée. De plus, les archivistes reconnaissent qu'il est impossible et irresponsable de conserver tout. Il faut des ressources pour garantir l'accès à des informations fiables, bien contextualisées et bien organisées dans le temps ; les coûts de stockage et de préservation ainsi que l'impact environnemental du stockage numérique doivent tous être durables.

Les archivistes peuvent analyser la collecte des matériels selon : la pertinence d'une zone géographique ou d'une communauté particulière, le désir d'une représentation équilibrée et équitable des événements, des idées et des peuples ou les intérêts de recherche de leur établissement, notamment. De manière critique, les archivistes doivent considérer leur propre capacité à fournir un accès équitable et durable à toutes les parties prenantes des archives. Les documents d'archives ne peuvent pas être réservés à une élite ayant accès aux fonds et à la technologie nécessaires pour « débloquer » les documents. L'évaluation et la réévaluation sont devenues des outils de plus en plus précieux pour évaluer et prioriser les formats de fichiers numériques menacés afin d'assurer un accès permanent. Pour en savoir plus sur l'évaluation et la sélection des matériels numériques, consultez [Approaching Appraisal](#) de Nathan Tallman et Lauren Work⁵.

⁴ Certains dépôts n'ont pas de spécialiste de la préservation ; ainsi, une partie ou la totalité de ces responsabilités peuvent être confiées à d'autres membres du personnel dans certaines circonstances. L'absence de personnel de soutien à la préservation oblige les curateurs à tenir compte de la capacité de préservation des dépôts pour leurs décisions sur l'évaluation et la réévaluation.

⁵ Nathan Tallman and Lauren Work, "308.1 Approaching Appraisal: Framing Criteria for Selecting Digital Content for Preservation," OSF, 20 Juin 20 2019, <https://doi.org/10.17605/OSF.IO/8Y6DC>.



Les conseils pour l'évaluation et la sélection des données de recherche sont également assez bien établis au sein de la communauté de gestion des données de recherche ; dans ce contexte, ces conseils visent à garantir que les données soient partagées selon les principes FAIR (facile à trouver, accessible, interopérable et réutilisable) et de manière équitable. On conseille aux déposants de prendre des décisions d'évaluation personnelle au début du cycle de vie des données de recherche, lors de la création ou de la révision d'un plan de gestion des données (PGD) au début d'un projet de recherche, ou lorsqu'ils commencent à nettoyer leurs données pour les partager avec d'autres. Les déposants peuvent travailler avec des curateurs à n'importe quel moment de ce processus, mais souvent l'évaluation en vue de la curation se fait d'abord lorsque les données sont consignées dans un dépôt. Le niveau et la qualité de la curation assurée par ce dépôt dépendront des ressources, des politiques et des priorités de chaque dépôt, mais la curation peut inclure les activités suivantes, notamment : la vérification de la conformité aux modalités d'utilisation du dépôt, l'assurance de la qualité des fichiers de données et de la documentation et des métadonnées qui fournissent des informations contextuelles, l'identification des considérations éthiques, la documentation des droits appropriés pour l'accès, la réutilisation et la préservation ainsi que la confirmation d'utilisation de formats de fichiers libres lorsque cela est possible. Parmi les cadres pour ces activités, il y a notamment la [liste de contrôle pour la CURATION](#) de l'Alliance de recherche numérique du Canada, les étapes [CURATED](#) du Data Curation Network et le [Lifecycle Model](#) du Digital Curation Centre⁶.

Malgré un chevauchement considérable des critères d'évaluation utilisés par les archivistes et les curateurs de données, une différence essentielle entre les deux domaines est que l'évaluation archivistique reflète et renforce un engagement envers l'intendance à long terme de tout matériel dans les collections. L'évaluation archivistique oriente et soutient les décisions d'acquisition générales tout en fournissant un mécanisme pour évaluer la valeur continue d'une ressource face aux nouveaux défis liés à l'obsolescence technique, à l'évolution des priorités de l'établissement et aux besoins changeants des communautés d'intervenants. Les pratiques d'évaluation archivistique exigent un investissement continu en temps et en efforts, ainsi qu'une défense permanente des intérêts des parties prenantes du secteur des archives, longtemps après l'acquisition initiale d'une ressource.

En ce qui concerne les données de recherche, les pratiques d'évaluation donnent souvent la priorité à la réutilisation immédiate des données et à l'impact du partage des données, parfois au détriment de la préservation à long terme⁷. Rien n'indique non plus que la réévaluation à l'appui

⁶ Alexandra Cooper, Michael Steeleworthy, Ève Paquette-Bigras, Erin Clary, Erin MacPherson, Louise Gillis, Lee Wilson and Jason Brodeur, « Guide pour la curation des données », Zenodo, 19 octobre 2021, <https://doi.org/10.5281/zenodo.5579820>; "The DCN Curation Workflow: The CURATED Steps," Data Curation Network, consulté le 26 janvier 2022, <https://datacurationnetwork.org/outputs/workflows/>; "The DCC Curation Lifecycle Model," Digital Curation Centre, consulté le 26 janvier 2022, <https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>.

⁷ C'est courant lorsqu'il n'y a pas de recommandations de format de préservation appropriées au moment de l'ajout dans un dépôt et que la meilleure option peut entraver la réutilisation en raison de la perte du



de la gestion de la préservation soit actuellement une pratique formalisée dans de nombreux dépôts de données⁸. Des processus clairs d'évaluation et de réévaluation sont requis pour que les dépôts soient désignés comme fiables par le biais de la certification CoreTrustSeal et par le biais de l'audit et de la certification des dépôts numériques fiables (ISO 16363). Or, il ne s'agit pas encore de certifications largement répandues et ces cadres n'exposent pas explicitement la manière de procéder à une réévaluation⁹. Le peu d'attention mis sur l'évaluation en tant qu'outil d'acquisition ou de retrait présente un risque de perte potentiellement considérable, car les futurs conservateurs chargés de la gestion continue de formats de données vieillissants s'efforcent de comprendre comment communiquer au mieux les anciennes données à de nouveaux publics.

- Qu'est-ce qui était important dans le format dans lequel les données ont été partagées à l'origine et est-ce que cela a encore de la valeur pour les intervenants actuels ?
- Qu'est-ce qui serait considéré comme une « perte acceptable » si les fichiers devaient être transformés pour être utilisables ?
- De quelle manière le domaine de recherche a-t-il évolué ?
- La communauté désignée par le dépôt a-t-elle diminué ou augmenté ? Les besoins de cette communauté ont-ils changé ?

Lorsque la curation de données prend également en charge la préservation à long terme, les pratiques d'évaluation et de réévaluation doivent être reconsidérées comme essentielles au maintien de la valeur des données préservées.

4. Processus d'évaluation

Qu'est-ce qu'on évalue ?

Le processus d'évaluation personnelle des chercheurs est souvent adapté au besoin particulier de communiquer les résultats de leurs recherches, ce qui signifie que l'évaluation et la sélection peuvent se faire à un niveau très précis. Idéalement, ce processus d'évaluation s'appuie sur les objectifs définis dans le plan de gestion des données des chercheurs et, le cas échéant, sur

contexte ou des composants fonctionnels essentiels des transformations de fichiers. La définition de ce qui constitue une perte acceptable afin d'assurer une accessibilité permanente est une préoccupation constante parmi la communauté de la préservation au sens large.

⁸ Lorsque la réévaluation apparaît dans les directives de curation, elle semble seulement être déclenchée par des problèmes liés aux données stockées, notamment l'échec des procédures de validation, ou bien elle est simplement indiquée comme une possibilité, si nécessaire, sans information particulière sur le moment où elle devrait être effectuée. Voir, par exemple *The DCC Curation Lifecycle Model*.

⁹ Le guide *CoreTrustSeal Extended Guidance 2020-2022* conseille à l'exigence 8, « Évaluation », que la réévaluation peut seulement être déclenchée par des changements dans les critères de sélection des dépôts. Voir *CoreTrustSeal Standards and Certification Board*, "CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022," Zenodo, consulté le 28 janvier 2022, <https://doi.org/10.5281/zenodo.3632533>; International Organization for Standardization, *Space data and information transfer systems — Audit and certification of trustworthy digital repositories*, ISO 16363:2012 (Geneva, Switzerland: ISO, dernière révision 2017), <https://www.iso.org/standard/56510.html>.



les conditions d'approbation obtenues auprès du comité d'éthique de la recherche. On peut supprimer ou créer de nouveaux fichiers afin de limiter ou d'affiner les points de données qui seront finalement partagés.

Pour les dépôts, l'évaluation se fait au niveau des ensembles des données. Chaque soumission doit être examinée et évaluée globalement pour vérifier qu'elle soit conforme aux politiques sur la collection ou autre politique des dépôts. En revanche, la sélection se fait au niveau des fichiers. Les curateurs peuvent signaler les fichiers problématiques et suggérer la suppression ou le remplacement du contenu problématique ou des formats de fichiers particuliers, ou encore demander des précisions contextuelles. En fin de compte, la décision d'accepter des ensembles de données dans des dépôts est toujours prise en fonction des ensembles de données globalement au lieu de leurs éléments individuels.

Qui fait de l'évaluation ?

Les créateurs de données participent à l'évaluation en prenant des décisions sur les données à partager. Ils peuvent travailler avec des curateurs au cours de ce processus ou entreprendre ce travail de leur propre chef.

Les curateurs de dépôt sont responsables en dernier ressort de l'évaluation des ensembles de données en vue de leur publication et de l'évaluation de leur pertinence pour le dépôt en collaboration avec les créateurs de données, les spécialistes de la discipline, les préservateurs ou d'autres intervenants, le cas échéant. Ces intervenants sont des sources inestimables de connaissances sur le coût de production des données, la singularité des données, l'importance des formats de fichiers et la valeur des données dans le domaine de recherche original.

À quel moment l'évaluation et la sélection ont-elles lieu ?

Les données peuvent être évaluées à n'importe quel moment du processus de recherche, mais l'évaluation et la sélection interviennent généralement lorsque les chercheurs entrent dans la phase de diffusion du cycle de vie de la recherche et que les curateurs de dépôt évaluent les données en vue de leur acceptation et de leur publication.

La réévaluation est la responsabilité du dépôt gestionnaire. La date de réévaluation peut être fixée lors de l'évaluation initiale, en cas d'incertitude quant à la valeur à long terme des données ou de problèmes de préservation liés aux formats de fichiers partagés, ou à des intervalles déterminés par les modalités du dépôt (par exemple, 10 ans après la consignation pour tous les matériels). La réévaluation peut avoir lieu plusieurs fois au cours de la vie d'un ensemble de données.

De quelle manière se déroule l'évaluation ?

Les déposants examinent d'abord les informations et les fichiers de données qu'ils souhaitent partager. Ils examinent aussi s'ils ont l'autorisation de les partager sur le plan juridique et éthique. Si nécessaire, ils peuvent demander des autorisations supplémentaires, ou modifier ou éliminer certaines données de leur dépôt pour se conformer à ces obligations. Ils peuvent choisir de partager des données brutes, analysées ou traitées, ou une combinaison des trois, en fonction de l'utilité de chacune de ces données pour aider les autres à comprendre et à reproduire leur travail.



Les curateurs évaluent le contenu et le contexte des ensembles de données pour déterminer si les données sont effectivement compréhensibles et utilisables par une personne autre que le déposant des données. Dans ce processus, les curateurs examinent également si les soumissions sont conformes aux politiques en matière de collection du dépôt et examinent la valeur potentielle à long terme du stockage et du partage des ensembles de données pour une utilisation future. Les données sont-elles présentées d'une manière appropriée et utile pour d'autres contextes de recherche ?

Tandis que le processus de sélection des déposants peut être particulier pour chaque chercheur, l'évaluation des ensembles de données par les curateurs doit être transparente et responsable. Un groupe de critères formel fait en sorte que l'évaluation n'est pas sujette à des préjugés individuels. Des listes de contrôle d'évaluation et d'autres documents d'orientation peuvent également garantir que les décisions des curateurs soient enregistrées, de sorte que des réévaluations n'aient pas lieu à l'avenir sans accès aux informations disponibles au moment de la soumission.

5. Critères d'évaluation

Le tableau suivant présente plusieurs questions générales pouvant servir à déterminer les besoins immédiats et permanents en matière de préservation, ainsi que leurs implications pour les dépôts gestionnaires pendant l'évaluation et la réévaluation¹⁰.

Critères	Questions clés	Implication de la préservation
Intention de préservation	Est-ce que le propriétaire de l'ensemble de données a l'intention ou l'obligation de préserver les matériaux ? Un bailleur de fonds ou un autre intervenant impose-t-il une obligation ? Quels aspects des matériaux le déposant	Les chercheurs peuvent faire appel à votre dépôt pour différents besoins ; il est important de comprendre leurs exigences afin de conseiller les chercheurs sur les options de préservation de leur travail. Les chercheurs peuvent également indiquer quelles caractéristiques d'un ensemble de données ils considèrent comme importantes à préserver, notamment le contenu ainsi que l'apparence, l'interactivité ou tout autre aspect. L'intention de préservation des

¹⁰ Les auteurs souhaitent reconnaître que ces critères ont été compilés à partir de documents d'orientation sur l'évaluation existants, notamment "Approaching Appraisal" (2019) de Tallman et Work et "How to Appraise & Select Research Data for Curation" (2010) de Whyte et Wilson, dans le but de clarifier les implications particulières de la préservation dans leur utilisation. Veuillez consulter la section "Sources et citations" à la fin du guide pour obtenir d'autres ressources sur l'évaluation.



	considère-t-il comme importants à préserver ?	<p>chercheurs doit être documentée et prise en compte lors de l'évaluation initiale, ainsi que lors de réévaluation ultérieure.</p> <p>Conseil : lorsque les recherches concernent des peuples ou des terres autochtones, assurez-vous que les créateurs des données ont obtenu l'autorisation de partager et de préserver l'ensemble de données, en particulier si les activités de préservation peuvent nécessiter de transformer ou d'augmenter les données d'une manière ou d'une autre. La consultation des parties prenantes peut être judicieuse afin de déterminer si votre dépôt convient à l'intention de préservation des déposants.</p>
Pertinence par rapport à la mission	Les matériaux correspondent-ils au mandat d'acquisition, à la politique de collection, au domaine de spécialité ou à d'autres priorités de votre établissement ou dépôt ?	<p>Les établissements de la mémoire et les dépôts qui ont l'intention de conserver les données de recherche à long terme doivent avoir une mission et un mandat pour acquérir et préserver ces collections. Cette mission peut être définie plus précisément par des politiques, telles que les politiques sur les collections, les directives de soumission, les modalités d'utilisation et les politiques de préservation. En l'absence d'une telle mission ou d'un tel mandat, l'établissement peut ne pas être prêt à consacrer les ressources financières nécessaires à la préservation à long terme et l'évaluation peut ne pas être menée de bonne foi.</p> <p>Conseil : Si des soumissions ne correspondent pas à votre politique sur les collections, pouvez-vous conseiller aux déposants d'autres dépôts plus convenables pour leurs données ?</p>
Valeur	Les données fournissent-elles des preuves valables de l'activité de recherche et présentent-elles une valeur sociale, scientifique ou historique potentielle à long terme ?	La valeur future est très difficile à prévoir ; par conséquent, les archivistes tiennent également compte du paysage actuel dans leurs décisions. Quelle est la valeur de ces données dans la communication des connaissances et des perspectives actuelles ? Existe-t-il des indicateurs



		<p>externes (citations, subventions, prix, statistiques d'accès) ou des normes disciplinaires (périodes de conservation) qui peuvent appuyer les déterminations de la valeur ? Portez une attention particulière aux données aidant à représenter la recherche sur des personnes sous-documentées ou marginalisées ou qui sont créées par celles-ci, ou encore qui documentent les changements dans un domaine de pratique.</p> <p>Conseil : Demandez-vous qui est susceptible d'utiliser ces données, pourquoi, et sous quel format elles seraient utiles. Le dépôt répond-il à ces besoins ? Envisagez une réévaluation quelques années plus tard si vous n'êtes pas sûr - l'intérêt de la communauté et l'utilisation à court terme peuvent aider à répondre à cette question.</p>
Unicité	<p>Les données sont-elles uniques ou sont-elles publiées ou disponibles ailleurs ?</p> <p>L'ensemble de données est-il accompagné d'une copie du code ou du modèle scientifique permettant de le reproduire ?</p>	<p>Si les données ne sont pas uniques et ont été acquises et préservées par un autre dépôt, la décision de ne pas les conserver devrait être simple. La redondance n'est pas forcément une mauvaise chose — parfois, les archives conservent des copies de documents qui existent également dans d'autres archives pour en faciliter l'accès. De même, si les données ont été produites par un modèle, l'impact environnemental de la régénération de ces données peut être supérieur au coût de leur préservation.</p> <p>Conseil : Si l'ensemble de données est utile à votre communauté d'utilisateurs et qu'il est important pour d'autres raisons indiquées dans ce tableau, vous pouvez le préserver malgré la duplication. Le cas échéant, envisagez une réévaluation dans 5 à 10 ans.</p>
Coût/justification économique	<p>Existe-t-il des ressources adéquates pour :</p> <ul style="list-style-type: none"> • Temps du personnel pour valider ou améliorer les métadonnées 	<p>Les coûts de l'équipement et du stockage sont susceptibles de diminuer à l'avenir, mais il n'en va pas de même pour les ressources humaines. La préservation à long terme implique un niveau de soins différent de celui de la simple conservation</p>



	<p>descriptives et administratives, la contextualisation ou la qualité des données, si nécessaire ?</p> <ul style="list-style-type: none"> • Des ressources pour soutenir les mesures de préservation, notamment la conversion de format ou l'émulation de logiciels ? • Les coûts de stockage et de maintenance dans le temps ? 	<p>des données. Combien d'efforts faudra-t-il déployer pour que cet ensemble de données reste utile à l'avenir ? Votre établissement ou votre dépôt est-il prêt à assumer les coûts humains, technologiques et environnementaux permanents nécessaires pour préserver les données ?</p> <p>Conseil : Les politiques des établissements en matière de curation et de préservation sont essentielles pour guider cette prise de décision. La justification des ressources allouées à la validation des fichiers et à la migration vers de nouveaux formats est difficile si vous n'avez pas formulé votre engagement en matière de préservation.</p>
<p>Droits et restrictions/ Potentiel de redistribution</p>	<p>Les données peuvent-elles être rendues accessibles à la communauté désignée du dépôt ou existe-t-il des restrictions qui en empêcheraient l'accès ?</p> <p>Les données contiennent-elles des informations sensibles ou personnelles qui empêcheraient de les rendre accessibles au public ?</p>	<p>L'accessibilité est l'une des questions les plus importantes dans la curation des données de recherche et elle est également importante pour la préservation à long terme. Sans la possibilité de permettre l'accès aux données à un moment précis, l'effort requis pour la préservation n'est pas justifiable.</p> <p>Conseil : Les données consignées dans un dépôt sont là pour être partagées. Les autorisations adéquates (notamment le consentement des participants et les approbations éthiques) ont-elles été obtenues ? Ces autorisations portent-elles spécifiquement sur la préservation à long terme ? Existe-t-il une raison juridique ou éthique pour laquelle vous ne souhaiteriez pas que votre dépôt continue à donner accès à ces données à l'avenir ?</p>
<p>Préservation du contenu et du contexte/ Documentation intégrale</p>	<p>Le contexte de la création des données est-il adéquatement documenté ? Les données sont-elles dans des formats qui</p>	<p>Les ensembles de données qui ne sont pas bien documentés aujourd'hui seront encore moins utiles aux futurs chercheurs à mesure que les méthodologies et les pratiques dans le domaine de la recherche évoluent. Si un chercheur a mal documenté son travail ou</p>



	<p>permettent la préservation et l'accès ?</p>	<p>utilisé des formats de fichiers propriétaires qui ne peuvent pas être facilement convertis en formats de fichiers ouverts, la question n'est peut-être pas de savoir si vous voulez ou non préserver l'ensemble de données, mais simplement de savoir si vous êtes en mesure de le faire.</p> <p>Conseil : Chaque nouveau format de fichier reçu représente un nouveau défi pour l'accessibilité à long terme. Le temps, le coût et les efforts nécessaires pour développer un processus permettant de faire des copies d'accès des fichiers de cet ensemble de données sont-ils justifiables ? (Les ressources DCN Data Curation Primers peuvent être des outils utiles pour guider l'extraction de données en formats propriétaires vers des formats accessibles. Lorsque les instructions ne sont pas disponibles pour un format donné, on peut entreprendre une évaluation plus formelle de la préservation. Un exemple de modèle conceptuel approfondi pour les évaluations de format est disponible auprès des The Danish National Archives.)</p> <p>Conseil : La publication d'une liste de formats préférés démontre un engagement envers la pérennité de ces formats, mais signale également aux déposants que d'autres formats pourraient ne pas bénéficier du même niveau de soutien. Si vous n'avez pas une liste publiée de formats préférés, le tableau Sustainability of Digital Formats ainsi que les recommandations de la Stratégie de numérisation du patrimoine documentaire sont de bonnes ressources pour suivre les effets du temps et de l'évolution technologique sur la préservation des formats de fichiers. Songez à réévaluer les formats de fichiers problématiques tous les 5 à 10 ans pour déterminer s'ils sont toujours utilisables ou recherchez l'aide d'un expert externe pour mieux gérer ces formats.</p>
--	--	---



6. Résultats de l'évaluation

L'évaluation des données de recherche commence lorsque les créateurs de données envisagent les données, les formats de fichier et la documentation à consigner et à partager dans un dépôt. Elle se termine généralement lorsque le dépôt accepte ou refuse de gérer l'ensemble de données pour le compte du créateur de données. Cependant, si le dépôt sélectionné s'engage à assurer la préservation à long terme, le processus d'évaluation remplit une fonction importante dépassant le seuil d'acceptation. De nombreuses décisions de réévaluation peuvent être prises au cours de la vie d'un ensemble de données ; dans chaque cas, les curateurs et les préservateurs doivent se demander s'ils peuvent ou doivent continuer à gérer les données, sous quelle forme, et pendant combien de temps. Chaque itération du processus d'évaluation d'un dépôt produira des recommandations qui peuvent être mises en œuvre et documentées pour soutenir la prochaine décision d'évaluation pouvant inclure les éléments suivants :

Lors de l'évaluation initiale :

- Demander aux chercheurs de fournir plus de documentation pour décrire l'ensemble de données ;
- Discuter avec les chercheurs de la valeur à long terme perçue des ensembles de données et de leur intention de préservation ;
- Convertir des fichiers en format propriétaire en un format de données ouvert pour l'accès et la préservation ;
- Documenter les décisions de curation dans les métadonnées qui seront préservées avec les ensembles de données ;
- Signaler les ensembles de données dont la réévaluation à court terme est prioritaire ;
- Identifier les fichiers particuliers qui pourraient être retirés d'un ensemble de données en raison de restrictions de licence, de contenu sensible ou d'autres considérations de sélection.

Lors de la réévaluation :

- Consulter de nouveau les intervenants de la communauté désignée du dépôt ;
- Transformer des fichiers en nouveaux formats d'accès ;
- Ajouter des métadonnées d'administration ou de préservation nécessaires à la gestion des ensembles de données ;
- Ajouter de la documentation pour les utilisateurs ou réviser celle-ci afin d'expliquer les changements de format de fichier ou les logiciels nécessaires pour accéder aux fichiers ;
- Transférer la garde des données à d'autres parties prenantes, ce qui peut inclure la remise des données aux propriétaires reconnus ou le transfert de celles-ci à un autre dépôt ayant l'expertise nécessaire pour gérer les formats de données ;
- Retirer ou supprimer les données qui ne sont plus considérées comme ayant une valeur à long terme.



Sources et citations

Beagrie, Neil. *What to Keep: A Jisc Research Data Study*. Bristol : Jisc, 2019. Dernière révision — février 2019. <https://repository.jisc.ac.uk/id/eprint/7262>.

Bieman, Ern and William Vinh-Doyle. “Recommandations relatives aux formats de fichier pour la préservation numérique — Stratégie de numérisation du patrimoine documentaire.” Ottawa : Réseau canadien d'information sur le patrimoine, ministère du Patrimoine canadien. Dernière révision — 5 mars 2021. <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/recommandations-formats-fichier-preservation-numerique.html>.

Caswell, Michelle. “The Archive” Is Not an Archives: On Acknowledging the Intellectual Contributions of Archival Studies.” *Reconstruction: Studies in Contemporary Culture* 16, no. 1 (March 2016): [n.p.]. <http://reconstruction.digitalodu.com/Issues/161/Caswell.shtml>.

Cooper, Alexandra, Michael Steeleworthy, Ève Paquette-Bigras, Erin Clary, Erin MacPherson, Louise Gillis, Lee Wilson et Jason Brodeur. « Guide pour la curation des données ». Zenodo. 19 octobre 2021. <https://doi.org/10.5281/zenodo.5579820>.

Data Curation Network. *Data Curation Primers*. [Collection]. Minneapolis: University of Minnesota Digital Conservancy. Consulté le 27 janvier 2022. <https://hdl.handle.net/11299/202810>.

“The DCC Curation Lifecycle Model.” Digital Curation Centre. Consulté le 26 janvier 2022. <https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>.

“The DCN Curation Workflow: The CURATED Steps.” Data Curation Network. Consulté le 26 janvier 2022. <https://datacurationnetwork.org/outputs/workflows/>.

CoreTrustSeal Standards and Certification Board. “CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022.” Zenodo. Consulté le 28 janvier 2022. <https://doi.org/10.5281/zenodo.3632533>.

“Glossary—Digital Preservation.” Digital Preservation Coalition. Consulté le 28 janvier 2022. <https://www.dpconline.org/handbook/glossary#D>.

Green, Ann, Stuart Macdonald, and Robin Rice. *Policy-making for Research Data in Repositories: A Guide, v. 1.2*. Edinburgh: DISC-UK DataShare and EDINA and University Data Library, Université d'Édimbourg, Mai 2009. <https://www.coar-repositories.org/files/guide.pdf>.



Harvey, Ross. “Appraisal and Selection,” in *DCC Digital Curation Manual*, edited by Seamus Ross and Michael Day. Digital Curation Centre. Consulté le 28 janvier 2022. <https://www.dcc.ac.uk/sites/default/files/appraisal-and-selection%5b1%5d.pdf>.

International Organization for Standardization. *Space data and information transfer systems — Audit and certification of trustworthy digital repositories*. ISO 16363:2012. Genève, Suisse : ISO. Dernière révision — 2017, <https://www.iso.org/standard/56510.html>.

Kjærskov, Frederik Holmelund, Asbjørn Skødt, Ann-Kristin Egeland, René Mittå, Jan Dalsten Sørensen. “Concept Model for Development of Preservation Plans, version 1.0.” Copenhagen: Danish National Archives, 2021. Consulté le 27 janvier 2022. <https://github.com/the-danish-national-archives/concept-model>.

“Policy Tools and Guidance.” Digital Curation Centre. Consulté le 28 janvier 2022. <https://www.dcc.ac.uk/guidance/policy/policy-tools-and-guidance>.

“Sustainability of Digital Formats: Planning for Library of Congress Collections.” Library of Congress. Consulté le 27 janvier 2022. <https://www.loc.gov/preservation/digital/formats/>.

Tallman, Nathan and Lauren Work. “308.1 Approaching Appraisal: Framing Criteria for Selecting Digital Content for Preservation.” OSF. 20 juin 2019. <https://doi.org/10.17605/OSF.IO/8Y6DC>.

Tjalsma, Heiko and Jeroen Rombouts. “Selection of Research Data: Guidelines for appraising and selecting research data. A Report by DANS and 3TU.Datacentrum.” Utrecht : SURFfoundation, 1^{er} juillet 2010. <http://resolver.tudelft.nl/uuid:dbab8a19-542a-4c4d-96b4-df8cc39333db>.

Whyte, Angus and Andrew Wilson. “How to Appraise and Select Research Data for Curation.” DCC How-to Guides. Edinburgh: Digital Curation Centre, 2010. Consulté le 27 janvier 2022. <https://www.dcc.ac.uk/sites/default/files/documents/How%20to%20Appraise%20and%20Select%20Research%20Data.pdf>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3, no. 1 (2016): 1-9. <https://doi.org/10.1038/sdata.2016.18>.



Remerciements

Les auteurs remercient le groupe d'experts sur la préservation et le groupe d'experts sur la curation de l'Alliance de recherche numérique du Canada pour avoir révisé les versions provisoires de ce document et fourni de précieux commentaires. Ils remercient notamment : Michael Steeleworthy, Jay Brodeur, Erin Clary et Qian Zhang. Ils remercient également Hannah Hadley, responsable du dépôt Open Scholarship Repository des bibliothèques de Princeton pour sa précieuse contribution.