

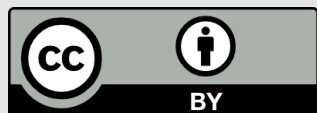
• SÉRIE DE RAPPORTS SUR LES DÉPÔTS OUVERTS

# La fonctionnalité de conservation numérique des systèmes de dépôts canadiens

*par Tomasz Neugebauer, Pierre Lasou, Andrea Kosavic, Tim Walsh*

*(Sous-groupe sur les dépôts de la prochaine génération du Groupe de travail sur les dépôts ouverts de l'ABRC)*

DÉCEMBRE 2019



[La fonctionnalité de conservation numérique des systèmes des dépôts canadiens](#) est publié par les membres du Sous-groupe sur les dépôts de la prochaine génération du [Groupe de travail sur les dépôts ouverts de l'ABRC](#) et est sous une [Licence internationale Creative Commons Attribution 4.0](#).

# Table des matières

Introduction .....	2
Sommaire des recommandations .....	3
<b>Exigences en matière de préservation numérique .....</b>	<b>3</b>
Préservation des bits .....	4
Métadonnées de préservation.....	5
Exportation d’objets archivés.....	5
<b>Systemes de dépôts ouverts du Canada .....</b>	<b>6</b>
DSpace .....	6
EPrints.....	7
Islandora .....	8
Samvera .....	9
Omeka .....	10
ContentDM.....	11
DigitalCommons .....	11
<b>Glossaire des termes de conservation numérique.....</b>	<b>13</b>
<b>Annexe 1. Sommaire des fonctionnalités de conservation numérique des systèmes de dépôt.....</b>	<b>16</b>

# Introduction

La préservation numérique est la série de politiques et d'activités de gestion visant à garantir à long terme l'utilisabilité, l'authenticité, la découvrabilité et l'accès au contenu numérique. La préservation numérique est nécessaire pour faire face à des défis tels que la corruption des fichiers, la défaillance des supports et le changement technologique qui pourraient affecter l'utilisabilité continue du contenu numérique au fil du temps. La fonctionnalité de préservation numérique est l'une des recommandations du document « Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group » [Recommandations techniques et en matière de comportement du groupe de travail sur les dépôts de la prochaine génération de la COAR (Confederation of Open Access Repositories)]<sup>1</sup>.

En mai 2018, le réseau Portage a publié un livre blanc intitulé « La préservation des données de recherche au Canada »<sup>2</sup>. Le livre blanc de Portage offre un aperçu des objectifs, des principes et des organisations de préservation numérique qui visent à bâtir une infrastructure de préservation numérique durable au Canada. Il propose un modèle en réseau où les dépôts fonctionnent comme des nœuds responsables de l'intégration, de l'accès et de la gestion des données, complétés par le développement de nœuds de services de préservation qui offrent une fonction de stockage.

Ce rapport, « La fonctionnalité de conservation numérique des systèmes de dépôts canadiens, » a été préparé dans le cadre des travaux du Groupe de travail sur les dépôts ouverts de l'ABRC, Sous-groupe sur les dépôts de la prochaine génération. Le rapport a été rédigé pour avancer vers l'une des prochaines étapes clés dans la réalisation de la vision décrite dans le livre blanc de Portage : la réalisation d'une compréhension commune des exigences de base de la préservation numérique et des fonctionnalités nécessaires pour atteindre cette vision d'un réseau de préservation numérique durable. Ce rapport se concentre sur les exigences techniques fonctionnelles et, à ce titre, il est destiné aux bibliothécaires de la communication savante et aux administrateurs de dépôts qui évaluent ou cherchent à améliorer les supports de conservation numérique dans leurs dépôts. Nous commençons par énumérer les exigences de préservation numérique minimales qui doivent exister dans les dépôts comme lieux d'intégration et de gestion de contenu. Nous procédons ensuite à une description plus détaillée de ces éléments, en fournissant un cas d'utilisation pour chacun et un résumé du support de ces éléments dans chacun des systèmes de dépôts identifiés comme étant couramment utilisés au Canada. Ce rapport se concentre sur les exigences techniques, mais il est important de souligner

---

<sup>1</sup> Rodrigues, E., & Shearer, K. (2017). Behaviours and technical recommendations of the COAR Next Generation Repositories Working Group. <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>

<sup>2</sup> Qasim, U., Davis, C., Garnett, A., Marks, S., & Moosberger, M. (2018). La préservation des données de recherche au Canada : livre blanc. <https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0371945>

que la création et la mise à jour des politiques et activités de conservation numérique exigent également que les organisations qui gèrent les dépôts définissent et établissent des rôles et des responsabilités à cet effet.

## Sommaire des recommandations

Dans ce rapport, et à la lumière de la vision d'un système de préservation numérique en réseau au Canada, nous tentons d'établir le niveau minimal recommandé de fonctionnalité de préservation numérique dans les systèmes de dépôts (Exigences en matière de préservation numérique). Le renforcement des capacités de préservation numérique nécessite un engagement des organisations grâce à la reconnaissance de nouveaux rôles et au développement de politiques et de technologies de préservation numérique. En nous concentrant spécifiquement sur les aspects technologiques de la préservation numérique dans les systèmes de dépôts, nous recommandons que les bibliothécaires, les gestionnaires et les administrateurs de dépôts continuent à accroître leur compréhension des processus et fonctionnalités de préservation numérique qui sont actuellement offerts dans leurs systèmes. Cette compréhension de la fonctionnalité de préservation numérique dans les systèmes logiciels des dépôts est utile pour identifier et traiter des domaines d'amélioration possibles, tels que: la configuration des plug-ins de préservation numérique, les services ou les scripts en ligne de commande, la promotion du développement de fonctionnalités supplémentaires qui ne sont pas actuellement disponibles, l'assurance que les procédures de stockage et de sauvegarde répondent aux exigences minimales de préservation numérique, ou encore le développement d'une politique de préservation numérique et d'une prise de conscience et d'un engagement organisationnels.

## Exigences en matière de préservation numérique

En tant que point d'intégration et site de gestion des données pour objets numériques ayant une valeur culturelle importante, les dépôts doivent avoir une fonctionnalité de préservation numérique intégrée et doivent interagir avec les services de préservation numérique dédiés actuels et futurs. Les fonctions et les responsabilités liées à la préservation numérique sont bien décrites dans des normes telles que les niveaux de préservation numérique de la National Digital Stewardship Alliance (NDSA)<sup>3</sup> et le modèle de référence pour un système ouvert d'archivage d'information OAIS (ISO 14721)<sup>4</sup>.

---

<sup>3</sup> National Digital Stewardship Alliance – Digital Library Federation. (2019). Levels of digital preservation v2.0. <https://ndsa.org/activities/levels-of-digital-preservation/>

<sup>4</sup> OAIS (2012) CCSDS 650.0-M-2: Reference model for an open archival information system (OAIS). Magenta Book. Numéro 1. Juin 2012 (ISO 14721:2003) <http://public.ccsds.org/publications/archive/650x0m2.pdf>

Dans le cadre d'opération des systèmes de dépôts susceptibles de fonctionner avec des services de préservation numérique externes, une capacité de préservation numérique minimale implique d'assurer que :

- les fichiers ne sont ni perdus ni corrompus (préservation des bits)
- des renseignements détaillés sur les caractéristiques techniques, la chaîne de conservation et la provenance des objets numériques sont collectés pour permettre de futures actions de préservation et d'accès (métadonnées de préservation)
- les objets et métadonnées peuvent être transmis à des systèmes de préservation numériques dédiés, le cas échéant (exportation du paquet d'information archivé [AIP])

Les gestionnaires de dépôts peuvent également vouloir s'assurer que les fichiers puissent continuer à être partagés et accessibles au fil du temps malgré les effets du format des fichiers, de l'obsolescence des logiciels et des matériels. Il serait approprié de trouver des solutions à ce problème, que ce soit par la mise en œuvre de stratégie de préservation comme la migration des formats de fichiers, la normalisation, l'émulation des dépôts ou par l'exportation du contenu vers des fournisseurs de services de préservation numérique.

## Préservation des bits

**Cas d'utilisation:** *Un objet numérique téléchargé dans un dépôt est corrompu lors d'une mise à niveau des disques sur lesquels le dépôt réside. La somme de contrôle de l'objet a été stockée lors de l'intégration et une vérification périodique de la fixité a permis à l'administrateur du dépôt d'identifier le problème de l'objet. L'administrateur peut alors récupérer le fichier non corrompu à partir de l'une des copies de sauvegarde avant que la sauvegarde ne soit écrasée.*

La préservation des bits vise à garantir que les objets numériques restent authentiques au fil du temps. Les menaces à long terme qui pèsent sur les objets numériques sont bien documentées et comprennent une défaillance des supports, une défaillance des matériels, une défaillance des logiciels, une catastrophe naturelle, des attaques externes et internes et une défaillance économique ou organisationnelle.<sup>5</sup>

Les approches techniques de la préservation des bits atténuent ces risques en archivant plusieurs copies indépendantes de données, en utilisant des technologies d'archivage adaptables qui tirent parti de techniques telles que le codage d'effacement pour détecter et réparer les modifications apportées aux objets au niveau du bit ou du bloc, et en exécutant des contrôles de fixité (vérifier périodiquement que les sommes de contrôle des objets numériques archivées correspondent aux valeurs enregistrées précédemment) pour vérifier que les fichiers n'ont pas été corrompus ou modifiés. Les activités de vérification de la fixité doivent être archivées dans des journaux et enregistrées dans des métadonnées de

---

<sup>5</sup> Rosenthal, D. S. H., Robertson, T., Lipkiss, T., Reich, V., Morabito, S. (2005). Requirements for digital preservation systems: A bottom-up approach. *D-Lib Magazine*, 11(11).  
<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>

préservation pour les objets numériques. Les approches d'archivage doivent équilibrer la résilience avec la durabilité organisationnelle et budgétaire afin de minimiser le risque que les activités de préservation numérique deviennent trop coûteuses ou autrement irréalisables sur le plan organisationnel.

Au minimum, les objets numériques de toute valeur significative doivent toujours être stockés et sauvegardés selon la règle 3-2-1 (trois copies de données, sur deux types de supports de stockage, avec au moins une sauvegarde hors site). « Les niveaux de préservation numérique » du NDSA offre une mesure accessible pour améliorer progressivement le stockage dans une perspective de préservation numérique<sup>6</sup>.

## Métadonnées de préservation

**Cas d'utilisation:** *Un déposant de thèse doit télécharger un fichier PDF / A-1 d'archivage, mais télécharge plutôt un document PDF avec plusieurs problèmes de formatage, y compris l'utilisation de sources de contenu externes et de certaines polices personnalisées uniquement disponibles sur l'ordinateur du déposant. Le document PDF s'affiche sur l'ordinateur du déposant, mais il inclut du texte illisible lorsqu'il est ouvert sur d'autres appareils. Le dépôt tente de valider le format de fichier PDF / A-1 de l'objet lors de l'intégration et signale au déposant ou à l'administrateur du dépôt que l'objet téléchargé ne se conforme pas au format requis. Le déposant est alors tenu de corriger les problèmes de formatage avant que l'objet de la thèse ne soit accepté.*

Généralement exprimées à l'aide du dictionnaire de données PREMIS (Métadonnées de préservation nécessaires pour soutenir la préservation à long terme d'objets numériques) et souvent dans le contexte d'un fichier METS, les métadonnées de préservation « [traduction] documentent les processus techniques associés à la préservation, spécifient les renseignements relatifs à la gestion des droits, établissent l'authenticité du contenu numérique et enregistrent la chaîne de possession et la provenance d'un objet numérique. »<sup>7</sup>

En plus de la mise en œuvre de PREMIS, les systèmes de dépôt peuvent supporter les métadonnées de préservation en décrivant les caractéristiques techniques des objets numériques. Cette description comprend l'identification et la validation des formats de fichiers et l'extraction des caractéristiques d'objets numériques (métadonnées techniques sur les fichiers, par exemple les codecs et les débits binaires).

## Exportation d'objets archivés

**Cas d'utilisation:** *La plateforme logicielle actuelle du dépôt n'offre pas un niveau satisfaisant de fonctionnalité de préservation numérique et un nouveau nœud de service de préservation numérique consortial devient disponible. La fonctionnalité d'exportation AIP permet au gestionnaire de dépôt d'exporter tous les objets et métadonnées du dépôt vers le nouveau service de préservation numérique. Si*

---

<sup>6</sup> National Digital Stewardship Alliance - Digital Library Federation. (2019). Levels of digital preservation v2.0. <https://ndsa.org/activities/levels-of-digital-preservation/>

<sup>7</sup> National Digital Stewardship Alliance - Digital Library Federation. (n.d.) Glossary. <https://ndsa.org/glossary/>

*nécessaire, l'exportation AIP permet également au gestionnaire de dépôt de migrer tout le contenu vers une nouvelle plateforme d'accès qui inclut les exigences minimales de préservation.*

La fonctionnalité d'exportation AIP est requise pour l'interopérabilité avec les systèmes de préservation numérique dédiés. Il est possible d'étendre la fonctionnalité de base des systèmes de dépôt pour inclure la suite complète de supports de préservation numérique, mais la préservation numérique est un processus continu qui nécessite une évaluation continue des risques, y compris la plate-forme sur laquelle elle réside. Par conséquent, la fonctionnalité d'exportation des objets numériques avec les métadonnées est requise dans les systèmes de dépôt. Cette fonctionnalité devrait tirer parti des normes telles que METS et PREMIS dans la mesure du possible.

## Systèmes de dépôts ouverts du Canada

Sur la base de l'enquête de 2018 du Groupe de travail sur les dépôts ouverts canadiens, « Inventaire des dépôts au Canada »<sup>8</sup>, voici une liste des systèmes sur lesquels les dépôts ouverts canadiens opèrent principalement : DSpace, EPrints, Islandora, Samvera, Omeka, ContentDM, DigitalCommons. Pour chacun de ces systèmes, le tableau suivant contient un résumé des fonctionnalités de préservation numérique suivi d'une brève description du support de la préservation numérique.

### DSpace

#### Métadonnées de préservation

Les fonctionnalités intégrées pour une caractérisation des fichiers ne sont pas incluses dans DSpace. Le profileur de format binaire<sup>9</sup>, qui fait partie du module des tâches liées à la conservation de données, effectue l'identification de format par extension de fichier uniquement (et non par signature de format) pour les objets numériques lors de l'intégration.

« [traduction] DSpace a défini un ensemble d'éléments de métadonnées techniques pour répondre à leurs besoins en matière de préservation et de gestion du cycle de vie du contenu. Cet ensemble d'éléments de métadonnées techniques est mieux exprimé en utilisant le schéma de métadonnées de préservation PREMIS : objet. L'utilisation de DSpace du dictionnaire de données PREMIS pour représenter les

---

<sup>8</sup> Cet ensemble de données donne un aperçu des systèmes de dépôts utilisés au Canada (nom du dépôt, établissement, URL, serveur OAI, logiciel et versions). Ces infos furent rassemblées en octobre 2018.

<https://drive.google.com/open?id=1WiY9mRL6HNQ1hOf2LeedK33J8npdiLneCkWp-raOrKQ>

<sup>9</sup> DSpace 6.x documentation, curation system: Bitstream format profiler.

<https://wiki.duraspace.org/display/DSDOC6x/Curation+System#CurationSystem-BitstreamFormatProfiler>

éléments de métadonnées techniques nécessaires ne constitue pas une mise en œuvre complète du modèle de données PREMIS. »<sup>10</sup>

### **Préservation des bits**

DSpace génère une somme de contrôle MD5 à l'intégration pour chaque objet numérique. Depuis la version 1.7, DSpace inclut un programme de vérification de données de contrôle Checksum Checker<sup>11</sup> qui peut être exécuté à partir de la ligne de commande et configuré pour s'exécuter périodiquement en tant que tâche cron pour vérifier les sommes de contrôle pour tous les objets stockés.

C'est aux administrateurs du dépôt DSpace de s'assurer que le stockage et les sauvegardes répondent aux exigences minimales de préservation numérique.

### **Exportation d'objets archivés**

Depuis la version 1.7, DSpace dispose d'une fonctionnalité d'exportation AIP<sup>12</sup> intégrée. Ce format d'exportation peut être utilisé pour l'importation dans des systèmes de préservation numérique dédiés (p. ex. Archivematica).

## **EPrints**

La plupart des fonctionnalités de préservation numérique disponibles pour la communauté EPrints sont le résultat du projet Preserv de 2005-2009<sup>13</sup>. La trousse d'outils de préservation numérique « the Digital Preservation Toolkit »<sup>14</sup> qui a été publiée en 2011 est obsolète et incompatible avec la dernière version des dépendances (par exemple DROID).

### **Métadonnées de préservation**

EPrints effectue l'identification du format de fichier pour tous les fichiers à l'aide du logiciel utilitaire de fichier intégré Unix et effectue une extraction des fonctionnalités des formats audiovisuels à l'aide de ffprobe / ffmpeg.

EPrints inclut un plugin d'exportation METS, mais les fichiers METS contiennent des métadonnées de préservation numérique insuffisantes. Il comprend également un module « historique » qui suit les modifications des métadonnées pour chaque élément, mais il ne met pas en œuvre PREMIS.

---

<sup>10</sup> DSpace development projects, DSpace METS SIP profile. <https://wiki.duraspace.org/display/DSPACE/DSPACE+METS+SIP+profile>

<sup>11</sup> DSpace 5.x documentation, system administration, validating checksums of bitstreams. <https://wiki.duraspace.org/display/DSDOC5x/Validating+Checksums+of+Bitstreams>

<sup>12</sup> DSpace 6.x documentation, DSpace AIP format. <https://wiki.duraspace.org/dsdoc6x/system-administration/aip-backup-and-restore/dspace-aip-format#DSpaceAIPFormat-MakeupandDefinitionofAIPs>

<sup>13</sup> Preserv: Repository preservation and interoperability. <http://preserv.eprints.org/>

<sup>14</sup> Tarrant, D. (2012). Preservation toolkit. <http://bazaar.eprints.org/142/>



## Préservation des bits

EPrints génère une somme de contrôle pour les fichiers lors de l'intégration, mais les tests révèlent que tous les objets ne sont pas affectés à une somme de contrôle. Il n'y a pas de vérification périodique de la fixité intégrée dans EPrints.

C'est aux administrateurs de dépôts EPrints de s'assurer que le stockage et les sauvegardes répondent aux exigences minimales de préservation numérique.

## Exportation d'objets archivés

La communauté EPrints travaille sur un plugin d'exportation pour exploiter le logiciel de préservation numérique Archivematica. L'intégration proposée consiste à exporter à partir d'EPrints toutes les archives Eprints et à les regrouper avec les fichiers numériques pour les conserver dans Archivematica<sup>15</sup>.

## Islandora

Actuellement, Islandora 7 comprend Fedora 3, mais la version de la prochaine génération d'Islandora actuellement en développement (Islandora 8) la remplacera par Fedora 4 et 5. Une grande partie des capacités de préservation numérique d'Islandora sont intégrées via des modules utilitaires optionnels et des packs de solutions.<sup>16</sup> C'est aux gestionnaires de dépôt d'intégrer, d'activer et de configurer ces modules pour supporter les processus de préservation numérique dans Islandora.

## Métadonnées de préservation

Avec le module utilitaire Islandora FITS, Islandora caractérisera et extraira les fonctionnalités des objets numériques intégrés à l'aide de l'outil d'information sur les fichiers (FITS) à code source libre.<sup>17</sup>

## Préservation des bits

Islandora supporte la génération de sommes de contrôle et la vérification continue de la fixité via les modules utilitaires Islandora Checksum et Islandora Checksum Checker. Les résultats des contrôles de fixité dans le temps sont enregistrés dans PREMIS à l'aide du module utilitaire Islandora PREMIS. C'est aux administrateurs du dépôt Islandora de s'assurer que le stockage et les sauvegardes répondent aux exigences minimales de préservation numérique.

## Exportation d'objets archivés

Les objets peuvent être exportés depuis Islandora dans le format BagIt compatible avec la préservation à l'aide du module utilitaire Islandora BagIt.

---

<sup>15</sup> Digital preservation through EPrints-Archivematica integration - an EPrints export plugin to Archivematica. <https://github.com/eprintsug/EPrintsArchivematica>

<sup>16</sup> Islandora documentation. <https://wiki.duraspace.org/display/ISLANDORA/Start>

<sup>17</sup> File Information Tool Set (FITS). <https://projects.iq.harvard.edu/fits/home>

Si un site installe également Archivemata, il est possible de pousser des objets d'Islandora vers Archivemata via le module Archidora<sup>18</sup>. Dans ce flux de travail, les objets Islandora deviennent un paquet d'information à verser (SIP) qui est intégré dans Archivemata et ensuite transformé en paquet d'information archivé (AIP) et en paquet d'information diffusé (DIP) selon la configuration dans Archivemata.

## Samvera

Samvera est un logiciel à code source libre basé sur Fedora 4, Ruby on Rails, Solr et Blacklight. Le cadre Samvera « [traduction] consiste en un certain nombre de gemmes Ruby qui peuvent être combinées, configurées et adaptées pour répondre à une grande variété de besoins ». <sup>19</sup> Une façon populaire de mettre en œuvre Samvera consiste à utiliser des « ensembles de solutions » tels que les systèmes Avalon Media System, Hyrax et Hyku (le produit issu du projet « Hydra-In-A-Box »). En tant que tels, les dépôts construits avec le logiciel Samvera peuvent différer considérablement en termes de fonctionnalité de préservation numérique compatible. Cependant, tous les dépôts Samvera exploitent les fonctionnalités de préservation numérique de Fedora, notamment les sommes de contrôle intégrées, le contrôle de version au niveau de l'article avec restauration et un historique d'audit complet.

### Métadonnées de préservation

Le module `hydra-file_characterization`<sup>20</sup>, un élément central de la communauté Samvera, caractérise les objets numériques intégrés à l'aide de FITS. Hyrax inclut cet élément par défaut.

Certaines métadonnées d'objet (par exemple, la valeur de la somme de contrôle / le résumé du message) sont exprimées à l'aide d'éléments du dictionnaire de données de PREMIS, mais les dépôts Samvera n'incluent pas une mise en œuvre complète de PREMIS.

### Préservation des bits

La vérification de la fixité (recalcul et comparaison des sommes de contrôle) est disponible via l'API Fedora. Cependant, cela n'est pas accessible via l'interface d'administration Web de Hyrax. Cela devrait être exécuté à partir de la ligne de commande par un développeur ou un administrateur du système. De plus amples renseignements sont disponibles dans la documentation de Fedora 5.<sup>21</sup>

---

<sup>18</sup> Islandora Documentation: Archidora.

<https://wiki.duraspace.org/display/ISLANDORA/Archidora>

<sup>19</sup> Samvera. <https://samvera.org/samvera-open-source-repository-framework/>

<sup>20</sup> Samvera file characterization (extracted from Sufia). [https://github.com/samvera/hydra-file\\_characterization](https://github.com/samvera/hydra-file_characterization)

<sup>21</sup> Fedora 5.x documentation: Fixity checking.

<https://wiki.duraspace.org/display/FEDORA5x/Fixity+Checking>

Depuis 2018, Hyrax n'inclut pas d'interface utilisateur d'administration pour faire un audit des contrôles de fixité intégrés de Fedora, ou pour restaurer des objets à partir de sauvegardes si une corruption de fichier ou une dégradation de bits est détectée.<sup>22</sup>

### **Exportation d'objets archivés**

L'Université de York (Royaume-Uni) et l'Université de Hull travaillent sur l'intégration de Samvera / Archivematica, mais aucun plugin ouvert n'est disponible.

C'est aux administrateurs de dépôts Samvera de s'assurer que le stockage et les sauvegardes répondent aux exigences minimales de préservation numérique.

## **Omeka**

Comme plateforme de présentation et d'expositions en ligne, Omeka n'offre pas de support natif pour la gestion des actifs numériques tout au long du cycle de vie et pour la préservation numérique. Les administrateurs des systèmes Omeka doivent s'assurer que le contenu numérique nécessitant une préservation est géré à l'aide de systèmes et de processus distincts.

### **Métadonnées de préservation**

Le code source Omeka comprend le Zend Framework, qui contient des bibliothèques PHP pour identifier et stocker les types de MIME et les sommes de contrôle (en utilisant les algorithmes CRC32, MD5 et SHA-1) et pour effectuer une caractérisation de fichiers de base. Le code source Omeka comprend également getID3, un script PHP qui effectue un certain degré de caractérisation de fichiers de formats multimédias. On ne sait pas dans quelle mesure ces logiciels utilitaires sont utilisés pour caractériser régulièrement les fichiers téléchargés sur Omeka.

Omeka ne met pas en œuvre PREMIS.

### **Préservation des bits**

Omeka semble enregistrer des sommes de contrôle pour les fichiers téléchargés, mais ne contient pas de fonctionnalité pour effectuer des contrôles de fixité de routine au fil du temps.

C'est aux administrateurs du dépôt Omeka de s'assurer que le stockage et les sauvegardes répondent aux exigences minimales de préservation numérique.

### **Exportation d'objets archivés**

Les intégrations avec d'autres systèmes de dépôt se sont concentrées sur l'importation de contenu vers Omeka pour présentation. Il existe des modules

---

<sup>22</sup> Rochkind, J. (2017). Exploring and planning with Sufia/Hyrax/Fedora fixity validation. <https://bibwild.wordpress.com/2017/05/01/exploring-and-planning-with-sufiahyraxfedora-fixity-validation/>

d'importation pour Fedora et DSpace. La fonctionnalité d'exportation est beaucoup plus limitée.

## ContentDM

Étant donné que ContentDM est un logiciel propriétaire d'OCLC, les informations sur la fonctionnalité de préservation numérique sont tirées de supports marketing<sup>23</sup> et ne peuvent pas être vérifiées à partir du code source.

### Métadonnées de préservation

ContentDM effectue une vérification antivirus ainsi qu'une certaine identification du format de fichier (y compris la vérification que le format de fichier correspond à l'extension du fichier) pour les fichiers téléchargés.

ContentDM ne met pas en œuvre PREMIS.

### Préservation des bits

ContentDM génère et vérifie les sommes de contrôle des fichiers téléchargés et met à la disposition des administrateurs des « rapports d'intégrité » qui incluent un contrôle périodique de la fixité.

Les fichiers téléchargés vers ContentDM sont archivés dans des centres de données certifiés ISO 27001. Les détails de stockage tels que le nombre et l'emplacement géographique des copies sont inconnus.

### Exportation d'objets archivés

La fonctionnalité d'exportation est limitée aux métadonnées descriptives de Dublin Core dans divers formats tels que XML et format délimité par tabulation, mais les métadonnées de préservation au niveau de l'objet telles que les sommes de contrôle ne semblent pas être incluses. Un établissement décrit<sup>24</sup> avoir demandé au service à la clientèle d'OCLC et téléchargé un fichier d'exportation tar avec les objets.

## DigitalCommons

Étant donné que DigitalCommons<sup>25</sup> est un logiciel propriétaire de Bepress (RELX), les renseignements sur la fonctionnalité de préservation numérique sont tirés de documents marketing et ne peuvent pas être vérifiées à partir du code source. Les documents marketing précisent également qu'une partie de la fonctionnalité de

---

<sup>23</sup> CONTENTdm. Digital preservation: Keep your digital special collections secure. OCLC. [https://www.oclc.org/content/dam/oclc/services/brochures/215630-WWAE\\_Cdm-DigPres-Feature-Flier.pdf](https://www.oclc.org/content/dam/oclc/services/brochures/215630-WWAE_Cdm-DigPres-Feature-Flier.pdf)

<sup>24</sup> Bullen, A. (2016). A Doomsday scenario: Exporting CONTENTdm records to XTF. *D-Lib Magazine*, 22(11). <http://www.dlib.org/dlib/november16/bullen/11bullen.html>

<sup>25</sup> Bepress Digital Commons. <https://www.bepress.com/products/digital-commons/>

préservation numérique est en fait vendue en tant que service d'archives Bepress<sup>26</sup> distinct mis à disposition en 2016.<sup>27</sup>

### **Métadonnées de préservation**

DigitalCommons ne met pas en œuvre PREMIS. DigitalCommons archive le type MIME des fichiers téléchargés. Le logiciel n'est pas open source, il n'est donc pas possible de vérifier l'outil utilisé pour l'identification du type MIME.

DigitalCommons archive les objets dans leur format d'origine. L'option « d'intégrer des capacités » d'émulation et de migration est mentionnée, mais le stade actuel de développement de ces fonctionnalités dans DigitalCommons est inconnu. Bepress affirme également être « [traduction] déterminé à rendre les PDF accessibles sur le Web de manière permanente », promettant de migrer les objets PDF au cas où « Adobe change Acrobat à un point tel que les anciens PDF ne sont plus lisibles »<sup>28</sup>.

### **Préservation des bits**

Les « fichiers de production » téléchargés sur DigitalCommons sont stockés dans un entreposage redondant et sauvegardés via des « instantanés horaires et quotidiens » et des sauvegardes périodiques hors site. Une deuxième copie de la sauvegarde hors site est envoyée au stockage hébergé Amazon S3. Une autre sauvegarde est envoyée à Amazon Glacier, qui « effectue des vérifications régulières et systématiques de l'intégrité des données et est conçue pour s'autoréparer automatiquement. »<sup>29</sup> Cependant, l'abonné ne peut avoir accès à une sauvegarde de préservation que s'il est également abonné au service Bepress Archive.

Bepress affirme être « compatible avec LOCKSS »<sup>30</sup>, encourageant les clients à rejoindre un réseau pour la préservation.

### **Exportation d'objets archivés**

Bepress fait référence à l'initiative Open Archive pour la collecte des métadonnées (OAI-PMH) comme la méthode pour transférer les œuvres savantes hors de la plate-

---

<sup>26</sup> Getting started with Bepress Archive.

[https://www.bepress.com/reference\\_guide\\_dc/getting-started-bepress-archive/](https://www.bepress.com/reference_guide_dc/getting-started-bepress-archive/)

<sup>27</sup> Palmer, L. A. (2017). Storage made simple: Preserving digital objects with bepress Archive and Amazon S3. Northeast Institutional Repository Day. <https://doi.org/10.13028/trd9-mr81>. Relevé à : <https://escholarship.umassmed.edu/neirug/2017/program/11>.

<sup>28</sup> Format migration and emulation. Safeguarding your content with Digital Commons. [https://www.bepress.com/reference\\_guide\\_dc/safeguarding-content-digital-commons/#format-migration-and-emulation](https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/#format-migration-and-emulation)

<sup>29</sup> A Robust infrastructure and dedicated staff to manage it. Safeguarding your content with Digital Commons. [https://www.bepress.com/reference\\_guide\\_dc/safeguarding-content-digital-commons/#a-robust-infrastructure-and-dedicated-staff-to-manage-it](https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/#a-robust-infrastructure-and-dedicated-staff-to-manage-it)

<sup>30</sup> Safeguarding your content with Digital Commons. [https://www.bepress.com/reference\\_guide\\_dc/safeguarding-content-digital-commons/](https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/)

forme<sup>31</sup> tandis qu'un abonné de DigitalCommons a développé un logiciel pour résoudre la « difficulté d'exporter des fichiers d'objets par lots »,<sup>32</sup>

Le service Bepress Archive en combinaison avec un abonnement Amazon S3 fournit un accès institutionnel aux métadonnées Dublin Core ainsi qu'aux versions les plus récentes d'objets numériques téléchargés dans une hiérarchie de fichiers / dossiers. Ce service comprend les sommes de contrôle et la vérification de la fixité<sup>33</sup>, mais ce n'est pas clair si les métadonnées de la somme de contrôle peuvent être exportées. Un fichier METS n'est pas disponible.

## Glossaire des termes de conservation numérique

- **Caractérisation:** « [traduction] La caractérisation comprend quatre éléments: identifier le format de l'objet; valider la conformité de l'objet aux normes techniques de son format; extraire les métadonnées techniques de l'objet; et évaluer si l'objet doit être accepté dans un dépôt, sur la base des politiques définies par le conservateur. »<sup>34</sup>
- **Identification:** « [traduction] L'identification est le processus de détermination du format présomptif d'un objet numérique sur la base d'indices extrinsèques suggestifs et de signatures intrinsèques, à la fois internes (par exemple, nombre magique) et externes (par exemple, extension de fichier). »<sup>35</sup>
- **Validation:** « [traduction] La validation est le processus de détermination du niveau de conformité d'un objet numérique aux règles normatives syntaxiques et sémantiques définies par la spécification faisant autorité du format de l'objet. »<sup>36</sup>
- **Extraction de caractéristiques:** « [traduction] L'extraction de caractéristiques est le processus qui consiste à présenter les propriétés intrinsèques d'un objet numérique important pour la planification et les activités de préservation. »<sup>37</sup>

---

<sup>31</sup> Bepress Hosted & Cloud-based. <https://www.bepress.com/hosted-cloud-based/>

<sup>32</sup> Migration Guide: Digital Commons. <https://github.com/MarcusBarnes/mik/wiki/Migration-Guide:-Digital-Commons>

<sup>33</sup> Safeguarding Your Content with Digital Commons. [https://www.bepress.com/reference\\_guide\\_dc/safeguarding-content-digital-commons/](https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/)

<sup>34</sup> Digital Preservation Coalition. Glossary. Digital Preservation Handbook. <https://www.dpconline.org/handbook/glossary>

<sup>35</sup> Documentation JHOVE2. [https://bitbucket.org/jhove2/main/wiki/JHOVE2\\_Frequently\\_Asked\\_Questions\\_\(FAQ\)](https://bitbucket.org/jhove2/main/wiki/JHOVE2_Frequently_Asked_Questions_(FAQ))

<sup>36</sup> Ibid

<sup>37</sup> Ibid

- **Somme de contrôle:** « [traduction] Une valeur numérique calculée par algorithme pour un fichier ou un ensemble de fichiers utilisée pour valider l'état et le contenu du fichier afin de détecter les erreurs accidentelles qui peuvent avoir été introduites lors de sa transmission ou de son stockage. L'intégrité des données peut être vérifiée à tout moment ultérieur en recalculant la somme de contrôle et en la comparant à celle stockée. »<sup>38</sup>
- **Vérification de la fixité:** « [traduction] Un mécanisme pour vérifier qu'un objet numérique n'a pas été modifié de manière non documentée. Les sommes de contrôle, les résumés de messages et les signatures numériques sont des exemples d'outils pour effectuer des contrôles de fixité. Les informations de fixité, les informations créées par ces contrôles de fixité, fournissent la preuve de l'intégrité et de l'authenticité des objets numériques et sont essentielles pour permettre la confiance. »<sup>39</sup>
- **AIP:** « [traduction] Objets archivés (*AIP*): Un paquet d'information, composé d'informations sur le contenu et d'informations liées à la description des objets de préservation (PDI), qui est conservé dans un système OAIS [Open Archival Information System]. »<sup>40</sup>
- **METS:** « [traduction] Le schéma METS est une norme pour l'encodage des métadonnées descriptives, administratives et structurelles concernant les objets dans une bibliothèque numérique. »<sup>41</sup>
- **Métadonnées de préservation:** « [traduction] Les informations contextuelles nécessaires pour mener à bien, documenter et évaluer les processus qui soutiennent la préservation et l'accessibilité à long terme du contenu numérique. Les métadonnées de préservation documentent les processus techniques associés à la préservation, spécifient les informations de gestion des droits, établissent l'authenticité du contenu numérique et enregistrent la chaîne de possession et la provenance d'un objet numérique. »<sup>42</sup>
- **PREMIS:** « [traduction] Le dictionnaire de données PREMIS pour les métadonnées de préservation est la norme internationale pour les métadonnées afin de soutenir la préservation des objets numériques et d'assurer leur utilisation à long terme. Développé par une équipe internationale d'experts, PREMIS est mis en œuvre dans des projets de préservation numérique à travers

---

<sup>38</sup> National Digital Stewardship Alliance - Digital Library Federation. Glossary.  
<https://ndsa.org/glossary/>

<sup>39</sup> Ibid

<sup>40</sup> OAIS (2012) CCSDS 650.0-M-2: Reference model for an open archival information system (OAIS). Magenta Book. Numéro 1. Juin 2012 (ISO 14721:2003)  
<http://public.ccsds.org/publications/archive/650x0m2.pdf>

<sup>41</sup> Metadata Encoding and Transmission Standard (METS). Library of Congress.  
<http://www.loc.gov/standards/mets/>

<sup>42</sup> National Digital Stewardship Alliance - Digital Library Federation. Glossary.  
<https://ndsa.org/glossary/>

le monde, et le support de PREMIS est intégré à un certain nombre d'outils et de systèmes de préservation numérique commerciaux et open source. »<sup>43</sup>

- **Archivage:** « [traduction] La catégorie de stockage numérique qui fournit les services et fonctions pour le stockage, la maintenance et la récupération à long terme des objets numériques. »<sup>44</sup> L'archivage nécessite plusieurs copies non colocalisées des données stockées et doit prendre en compte les risques tels que la défaillance des supports, les catastrophes naturelles, l'obsolescence et le blocage des fournisseurs. Pour plus de détails, voir les niveaux de conservation numérique NDSA.<sup>45</sup>

---

<sup>43</sup> PREMIS Data Dictionary for Preservation Metadata (PREMIS). Library of Congress. <https://www.loc.gov/standards/premis/>

<sup>44</sup> National Digital Stewardship Alliance - Digital Library Federation. Glossary. <https://ndsa.org/glossary/>

<sup>45</sup> National Digital Stewardship Alliance - Digital Library Federation. (2019). Levels of digital preservation v2.0. <https://ndsa.org/activities/levels-of-digital-preservation/>



## Annexe 1. Sommaire des fonctionnalités de conservation numérique des systèmes de dépôt

Logiciel de dépôt	Métadonnées de préservation				Préservation des bits		Exportation AIP (objets + METS)
	Identification	Validation	Extraction de caractéristiques	PREMIS	Somme de contrôle d'intégration	Vérification de fixité	
DSpace	oui (extension de fichier seulement)	non	non	Partiel. Utilise certains éléments de PREMIS mais pas de mise en œuvre d'événements	oui	Via script de ligne de commande	oui
EPrints	oui (utilitaire de fichier)	non	oui (certains formats)	non	oui	non	Partiel (Exportation de METS mais pas d'AIP)
Islandora *via Modules utilitaires ** via Archidora	oui*	?	oui*	oui*	oui*	oui*	oui**
Samvera	oui	?	oui (certains formats)	Partiel. Utilise certains éléments de PREMIS mais pas de mise en œuvre d'événements	oui	Via script de ligne de commande	non?
Omeka	oui	Par extension de fichier et type de	oui (certains formats)	non	oui	non	non

		mime seulement					
ContentDM	Oui	?	?	non	oui	oui (détails inconnus)	non
DigitalComm ons  * (dans Amazon Glacier et S3) mais le client n'a accès à une copie de l'objet archivé que lorsqu'il achète un service d'archivage supplémentaire	oui	?	?	non	oui*	oui *	non